An Introduction to Marketing Research

Scott M. Smith
James Passey Professor of Marketing
Founder, Qualtrics

Gerald S. Albaum
Emeritus Professor of Marketing
University of New Mexico

Table of Contents

Preface

This book draws its "parentage" from the classic *Research for Marketing Decisions* by Paul E. Green, Donald S. Tull, and Gerald Albaum. But, it is not a revision of that book. Rather, it might best be viewed as a "child" which is targeted to a different audience—primarily senior-level undergraduate and MBA students who are users of Qualtrics.com.

We believe this book is "novel" in at least three major respects. First, with respect to *method*, the unifying concept of this book is that marketing research is a cost-incurring activity whose output is information of potential value to managers in making decisions. Second, with respect to *technique*, this book again departs from tradition in terms of an applied approach to the relatively large coverage of more sophisticated, yet relatively easily implemented, research techniques. The entire book focuses on implementation of **online marketing research.** Question types and examples are implemented using internet survey provider, Qualtrics.com, so that students can design, plan and implement an online survey of their own at no charge.

Finally, with respect to *analysis*, the book is expansive in its coverage, including relative emphasis on modern analytical tools such as multivariate analysis. In terms of number of chapters, 30% of the book is devoted to analysis, but, the discussion is at a level that senior-level undergraduates can understand, and the techniques are explained within the context of computer-based analysis.

This book is concerned with providing an ***introduction*** to marketing research. This means that all the basic elements of method, techniques, and analysis are covered, including those at a more sophisticated level. But, the book is <u>NOT</u> a book of only essentials. The methodological scope regarding research design, data collection techniques, and measurement is broad. For example, two chapters are devoted to the critical area of measurement and scaling. The book presents its material from primarily a pragmatic and user-oriented (rather than theoretical research technician) perspective. User-orientation is based on the premise that users need to know method in order to evaluate research presented to them.

Because the book is available online, it can be used in a modular fashion at no cost to the student. For example, if chapters on experimental design or multivariate statistics are beyond the scope of the instructor's focus, then they can simply be ignored. Similarly, if the course focuses on survey research, chapters 9 and 10 could be the focal point, supplemented with chapters 1,2,4,5,6,7 plus analysis chapters as appropriate. Note that because of the dynamic nature of electronic publication, chapters may be edited, and additional chapters may be added from time to time. There is a Glossary of Terms and an appendix that includes some widely-used statistical tables for analysis. These tables will be useful for analyzing appropriate cases.

Many people helped shape the content and style of this book, but most importantly Professors Paul E. Green and the late Donald S. Tull have had a profound influence on the authors' thinking about research and their book with one of the present authors provided a platform from which the present book was launched.

SCOTT M. SMITH
GERALD S. ALBAUM

Chapter 1
AN INTRODUCTION TO MARKETING RESEARCH

Marketing is a restless, changing, and dynamic business activity. The role of marketing itself has changed dramatically due to various crises—material and energy shortages, inflation, economic recessions, high unemployment, dying industries, dying companies, terrorism and war, and effects due to rapid technological changes in certain industries. Such changes, including the Internet, have forced today's marketing executive to becoming more market driven in their strategic decision-making, requiring a formalized means of acquiring accurate and timely information about customers, products and the marketplace and the overall environment. The means to help them do this is marketing research.

WHAT IS RESEARCH?

Research is a systematic and objective investigation of a subject or problem in order to discover relevant information or principles. It can be considered to be either primarily fundamental or applied in nature. *Fundamental research*, frequently called *basic* or *pure research*, seeks to extend the boundaries of knowledge in a given area with no necessary immediate application to existing problems, for example, the development of a research method that would be able to predict what people will be like *x* years in the future. In contrast, *applied research*, also known as *decisional research*, attempts to use existing knowledge to aid in the solution of some given problem or set of problems.

Marketing research assists in the overall management of the marketing function. A marketing manager must prioritize the more important and pressing problems selected for solution, reach the best possible solution based on the information available, implement the solution, modify the solution when additional information so dictates, and establish policy to act as a ready-made solution for any recurrence of the problem.

Marketing research often focuses on understanding the "Customer" (purchasers, consumers, influencers), the "Company" (product design, promotion, pricing, placement, service, sales), and can also be expanded toward the environment to include "Competitors" (and how their market offerings interact in the market environment).

Figure 1.1 Marketing Environment (Source: Modified from Perrault and McCarthy,)

Within this "Company-Customer-Competition" environment, many types of marketing research can be conducted, much of which is focused on using surveys for

- Monitoring customers and markets
- Measuring awareness, attitudes, and image
- Tracking product usage behavior
- Diagnosing immediate business problems
- Supporting strategy development

More specific examples are found in the Qualtrics.com Survey University. This provider of professional survey software identifies twenty different kinds of surveys that are of use to marketing researchers. Each focuses on a different aspect of the "Company" and it's interaction with the "Customer" and "Competition" in the market environment:

Exhibit 1.1 Twenty Different Types of Marketing Surveys

1 - Market Description Surveys
To determine the size and relative market share of the market. Such studies provide key information about market growth, competitive positioning and tracking share of market.

2 - Market Profiling-Segmentation Surveys
To identify who the customers are, who they are not, and why they are or are not your customers. This is often a descriptive market segmentation and market share analysis

3 - Stage in the Purchase Process / Tracking Surveys
Where is the customer in the adoption process? This information shows market Awareness – Knowledge – Intention – Trial – Purchase – Repurchase of the product.

4 - Customer Intention - Purchase Analysis Surveys
Directed at understanding the current customer. What motivates the customer to move from interest in the product to actual purchase? This is a key to understanding customer conversion, commitment and loyalty.

5 - Customer Attitudes and Expectations Surveys
Does the product meet customer expectations? What attitudes have customers formed about the product and/or company. Used to direct advertising and improve customer conversion, commitment and loyalty.

6 - Customer Trust - Loyalty – Retention Analysis Surveys
Especially for high priced consumer goods with long decision and purchase processes (time from need recognition to purchase), and depth of consumer attitudes formed about the product and/or company.

7 - New Product Concept Analysis Surveys
Concept test studies are appropriate in the initial screening of new product concepts. Likes and dislikes about the concept and evaluation of acceptability and likelihood of purchase are especially useful measures.

8 - New Product Acceptance and Demand Surveys (Conjoint Analysis)
Primarily for estimating demand for new products that can be described or have been developed in drawing or concept, but have not yet been developed physically. Develops develop market share estimates of market potential for the alternative potential products.

9 - Habits and Uses Surveys
Directed at understanding usage situations, including how, when and where the product is used. Habits and uses studies sometimes include a real or virtual pantry audit.

10 - Product Fulfillment Surveys (Attribute, Features, Promised Benefits)
Evaluation of the product's promised bundle of benefits (both tangible and image). Are expectations created for the product by advertising, packaging and the produce appearance fulfilled by the product?

11 - Product Positioning Surveys (Competitive Market Position)
A "Best Practices" study of "How does the market view us relative to the competition?" Competitive positioning analyses often compare the attributes and benefits that make up the product using multidimensional scaling.

12 - Brand Equity Analysis Surveys
What is psychological value that a brand holds in the market place? Brand equity is a composite of brand awareness, brand quality, brand associations and brand loyalty measures.

13 - Advertising Value Identification and Analysis Surveys
Advertising value analysis focuses on mapping the hierarchical attributes, benefits and values that are associated with and portrayed by an advertisement. Means-end analysis is often part of this type of study.

14 - Advertising Message Effectiveness Surveys (Media and Message)
Message effectiveness testing identifies the impressions, feelings, and effectiveness in moving the respondent to a desired goal (increased awareness, more product information, trial, repeat purchase).

15 - Sales Force Effectiveness Surveys
A combination of measures that focus on the sales activities, performance and effectiveness in producing the desired and measurable effect or goal. Often measured as a 360 degree survey completed by the sales person, the client (evaluating the sales call) and the supervisor responsible for evaluating the sales person.

16 - Sales Lead Generation Surveys
Sales lead generation surveys for (1) assuring timely use and follow-up of sales leads, (2) qualifying sales leads (thereby saving valuable sales force time) and (3) providing more effective tracking of sales leads.

17 - Customer Service Surveys
Akin to customer satisfaction surveys, but focus in detail on the actual customer service that was received, the process involved in receiving that service and the evaluation of the participants in the service process.

18 - Customer Service Representative (CSR) Surveys: Attitudes, Burnout, Turnover and Retention:
CSRs hold attitudes that reflect on their job related activities including (1) the allocation of time; (2) solutions to customer needs; (3) how to improve their job; (4) best practices; (5) How well internal departments help customers. CSRs often exhibit frustration, burnout and high turnover and surveys focus on CSR retention, reducing costs and increasing the quality of customer relationships.

19 - Sales Forecasting and Market Tracking Surveys
Sales forecasting and market tracking studies can include expert opinion (experts estimate the market), judgmental bootstrapping (expert based rules describing how to use available secondary market information), conjoint analysis (estimation of consumer intentions based on product attributes that are important in the decision), and intentions evaluations (consumer self reported intentions of future purchases) are to be made.

20 - Price Setting Surveys and Elasticity of Demand Analysis
Price surveys estimate the elasticity of demand and show optimal price points, including prices too low or too high. Price surveys may estimate the demand for different product or service segments, or different usage situations.

Source: Twenty Different Types of Marketing Surveys: http://www.qualtrics.com/wiki/index.php/Market_Surveys.

Each of the above surveys focuses on a specific area of research and involves the development of conceptual models directed at predicting or explaining a specific type of behavior that is being measured. This level of specificity is desirable for several reasons. Within the research process, this specificity brings:

1. ***Clarification***. Explication usually results in the clarification of relationships and interactions. The need for more rigorous definitions of key variables often becomes apparent.
2. ***Objectivity***. The process of explicating the modeled behavior often discloses rationalizations and unfounded opinions that had not been recognized as such before.
3. ***Communication***. Discussion helps to identify problems and common points of reference when different people hold alternative implicit models of the same problem situation.
4. ***Improvement of models***. Explicit models can be tested in differing situations to see if the results are reproducible. The degree and range of adaptability can thus be extended.

5. ***Guide to research needs***. Formulating models explicitly can better pinpoint information gaps and, thus, aid in determining the nature of research needs.

While varying information is required for the different types of marketing research projects, the key to conducting a successful research project lies with the researcher and the client. They must come to a common understanding of the nature of the exact research problem, and then agree on the information required to answer this problem. This requires identifying the appropriate questions, respondents, methodology, analysis and reporting. All studies must address these same basic issues (see Exhibit 1.2).

EXHIBIT 1.2 Basic Research Issues

As technology advances, marketing researchers are continually looking for ways to adapt new technology to the practice of research. Both hardware and software are involved in such adaptations. However, researchers must never forget that research basics cannot be overlooked. Rather, what must be done is to adapt the new techniques and technologies to these basics. All studies must address the following basic issues (Anderson, Berdie, & Liestman, 1984):

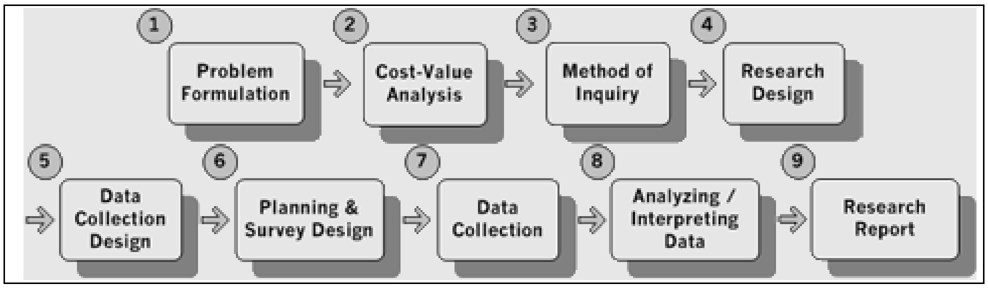
1. ***Ask the right questions.*** This is the essence of project design, and the heart of proper planning. The research planner must remember that every project is unique, and as such must be tailored to the user's needs.
2. ***Ask the right people.*** Sample design should be such that only those people who are of interest to the research user are contacted, and such that those who are contacted are reasonably representative of the group of interest.
3. ***Ask questions the right way.*** It is not enough to be able to ask the right questions; they must be asked in the right way. This is the essence of questionnaire design. The researcher can use all the aids available from the new technologies, but if the wording of the questions is not clear to the respondents, the results will be useless. One basic that is overlooked all too often is pretesting the questionnaire; this is crucial for ensuring that responses are the ones that are needed to address the problem.
4. ***Obtain answers to questions.*** The process of data collection is central to all marketing research. Techniques used should be selected for how each bears on nonresponse and response alike.
5. ***Relate answers to the needs of the research user/client.*** Data seldom speak for themselves. Proper data analysis is needed if a study is to have any value to the user. Here there is a risk of letting advanced techniques become the master of the researcher rather than the opposite. Common sense is a valuable tool for the researcher when considering alternative analysis approaches for any project.
6. ***Communicate effectively and in a way that the client understands.*** Many good projects are ruined in this stage. The information that is reported to the user should be in a form that is understandable to the user so that he or she can tell that it is relevant to the issues at hand.

Having considered these general topic-situation issues in conducting research, let's now turn to the basic process of conducting a research process.

THE BASIC RESEARCH PROCESS

How is marketing research actually conducted? What are the general steps in completing a research project? These questions are answered in the steps of the research process. While the steps are shown as a linear process, some of the steps may be performed simultaneously, such as selecting data collection techniques and sample design. There are other times when "later" decisions influence decisions that are made early in the research planning process. For example, desired analysis techniques often influence the selection of data collection techniques (e.g., measurement) and sample design.

Figure 1.2 The Research Process



It is important to carefully plan the research process and formally recognize the relationship between the stages. The researcher should write a formal plan for the project, including the background information and statement of objectives, which then becomes the master guide for implementing and controlling the research project. Each step in this research process will now be introduced.

STAGE 1: PROBLEM FORMULATION

In a very real sense, problem formulation is the heart of the research process. As such, it represents the single most important step to be performed. From the researcher's point of view, problem formulation means translating the management problem into a research problem.

As previously discussed, in order to formulate an appropriate research problem, the researcher must understand the origin and nature of management's problem and then be able to rephrase it into meaningful terms from an analytical point of view. This involves timely and clear communication between manager and researcher.

The end result of problem formulation is a statement of the management problem that is analytically meaningful and that often points the way to alternative solutions. An accurate problem formulation specifies the types of information needed to help solve the management problem. In short, quality thinking about a problem prior to data collection largely determines the quality of data collection, analysis and problem solving.

Exhibit 1.2 Examples of Management Problems and Related Research Problems

Management Problems	***Research Problems***
Allocate advertising budget to media	Estimate awareness generated by each media type
Decide whether to keep office open Saturday	Evaluate use of services on Saturday and determine on whether customers will shift usage to weekdays
Introduce a new health service	Design a concept test and assess acceptance and use
Change the marketing program the new	Design a test-marketing situation such that the effect of program can be estimated
Increase the sales of a product	Measure a product's current image

Closely related to problem formulation is the development of a working hypothesis, or an assertion about a state of nature. While hypotheses are crucial for basic research because they tell the researcher what to do, the concept of a hypothesis can also be useful in decisional research to direct the development of the research problem statement. In most cases, the marketing researcher will not explicitly state hypotheses for the research. Kerlinger and Lee (2000, Chapter 2) suggest that research problems and hypotheses meet the following criteria:

1. The problem statement expresses a relationship between two or more variables.
2. The problem is stated clearly and unambiguously in question form.
3. The problem statement implies possibilities of empirical testing.

Where properties of good hypotheses include the following:
1. The hypothesis is a statement about the relationship between two or more variables in declarative statement form.
2. The hypothesis carries clear implications for testing the stated relationship (i.e., variables must be measurable or potentially measurable).

How to Formulate the Research Problem

Problem formulation is much easier when specific components of the research problem are defined:

1. Specify the Research Objectives

Objectives guide the researcher in developing good, useful research, and they help the client evaluate the completed project. Objectives range from the very general, such as profit maximization, to the highly specific, such as measuring market interest in a new product. It is rare that the objectives are well explained to the researcher. However, the researcher needs to take the initiative to develop a clear statement of objectives.

Each study should have a very limited and manageable set of objectives that focus on the problem being solved. Two or three well targeted objectives is preferable to many that are ill-conceived. Fewer the objectives make it easier to keep track of progress toward the objectives, to ensure that each is properly addressed, and to determine the best methodology. If there are too many objectives separate studies may be appropriate.

2. The Environment or Context of the Problem

Consider the problem of deciding whether to introduce a new consumer product. The marketing researcher must work closely with the client in transforming the client's problem into a workable research problem.

The researcher's efforts should be oriented toward helping the manager decide whether any investigation is justified based on the potential value of the research findings versus their cost. The researcher must be aware of, and assist in, the identification of objectives, courses of action, and environmental variables, insofar as they affect the design of the research investigation.

If the research is undertaken and if the resulting findings are to be utilized (i.e., have an influence on the user's decision making), the manager and researcher must have a productive and trusting relationship that is based on the researcher's ability to perform and deliver the research as promised.

3. The Nature of the Problem

Every research problem may be evaluated on a scale that ranges from very simple to very complex. The degree of complexity depends on the number of variables that influence the problem. Understanding the nature of the problem helps a researcher ensure that the right problem is being investigated and that a marketing plan can be developed to solve the problem. A thorough preliminary investigation using focus groups of consumers, salespeople, managers, or others close to the problem may produce much needed insight.

4. Alternative Courses of Action

A course of action specifies a behavioral sequence that occurs over time, such as the adoption of a new package design, or the introduction of a new product. Such a program of action becomes a commitment, made in the present, to follow some behavioral pattern in the future.

It is usually desirable to generate as many alternatives as possible during the problem formulation stage and state them in the form of research hypotheses to be examined. A hypothesis often implies a possible course of action with a prediction of the outcome if that course of action is followed.

Once the nature of the problem has been agreed upon, the course of action must be specified. This involves:

1. Determining which variables affect the solution to the problem
2. Determining the degree to which each variable can be controlled
3. Determining the functional relationships between the variables and which variables are critical to the solution of the problem.

The following example shows the results of a failure to follow through with these aspects of the problem situation model.

EXHIBIT 1.3 "New Coke" Versus Original Coke

In the mid-1980s the Coca Cola Company made a decision to introduce a new beverage product (Hartley, 1995, pp. 129–145). The company had evidence that taste was the single most important cause of Coke's decline in the market share in the late 1970s and early 1980s. A new product dubbed "New Coke" was developed that was sweeter than the original-formula Coke.

Almost 200,000 blind product taste tests were conducted in the United States, and more than one-half of the participants favored New Coke over both the original formula and Pepsi. The new product was introduced and the original formula was withdrawn from the market. This turned out to be a big mistake! Eventually, the company reintroduced the original formula as Coke Classic and tried to market the two products. Ultimately, New Coke was withdrawn from the market.

What went wrong? Two things stand out. First, there was a flaw in the market research taste tests that were conducted: They assumed that taste was the deciding factor in consumer purchase behavior. Consumers were not told that only one product would be marketed. Thus, they were not asked whether they would give up the original formula for New Coke. Second, no one realized the symbolic value and emotional involvement people had with the original Coke. The bottom line on this is that relevant variables that would affect the problem solution were not included in the research.

CBS New Coke News Clip: http://www.youtube.com/watch?v=-doEpVWFLsE&NR=1&feature=fvwp
New Coke Commercial: http://www.youtube.com/watch?v=o4YvmN1hvNA
New Coke and Coke Classic Commercial: http://www.youtube.com/watch?v=ky45YGUA3co

5. The Consequences of Alternative Courses of Action

A set of consequences always relate to courses of action and even to the occurrence of events not under the control of the manager. One of the manager's primary jobs is to anticipate and communicate the possible outcomes of various courses of action that may result from following the research.

6. Degrees of Uncertainty

Most marketing problems are characterized by a situation of uncertainty as to which course of action is best. Years of experience may allow the decision-making manager to assign various "likelihoods of occurrence" to the various possible outcomes of specific courses of action.

A carefully formulated problem and statement of research purpose is necessary for competent research. The statement of purpose involves a translation of the decision maker's problem into a research problem and the derivation of a study design from this problem formulation. The research problem provides relevant information concerning recognized (or newly generated) alternative solutions to aid in this choice.

STAGE 2: METHOD OF INQUIRY

Market researchers look to the scientific method as the source of their investigative methods. Even though this method is not the only one used, it is the standard against which other investigative methods are measured. The scientific method makes great use of existing knowledge both as a starting point for investigation and as a check on the results of the investigations (i.e., a test of validity). Its most distinctive characteristic is its total lack of subjectivity. The scientific method has evolved objective and rigid procedures for verifying hypotheses or evaluating evidence. It is analytical in its processes and is investigator-independent. Thus, the scientific method is for the most part logical and objective, and frequently makes extensive use of mathematical reasoning and complicated experiments.

The goal of a scientific methodologist, also called an objectivist, is to run a hypothesis test using publicly stated procedures that are investigator-independent.

- Formulate a problem
- Develop a hypothesis
- Make predictions based on the hypothesis
- Devise a test of the hypothesis
- Conduct the test
- Analyze the results

Even though the terminology used is that associated with basic research, the process described is analogous to that of decision making. Although the steps are the same, there are differences in the way in which the steps are performed and in the underlying assumptions about behavior. For example, the essential difference between the objectivist and the subjectivist is the latter's allowance for use of subjective judgments both when collecting data and when analyzing data (Diesing, 1966).

This objectivist-subjectivist distinction has very practical meaning, particularly when considering the use of outside research suppliers. There are commercial research firms that tend to specialize in one or the other method of inquiry. Objectivist-based research is often called *quantitative research*, whereas subjectivist-based research is often called *qualitative research*.

Exhibit 1.4 The Scientific Method

In structure, if not always in application, the scientific method is simple and consists of the following steps:

1. *Observation.* This is the problem-awareness phase, which involves observing a set of significant factors that relate to the problem situation.
2. *Formulation of hypotheses.* In this stage, a hypothesis (i.e., a generalization about reality that permit prediction) is formed that postulates a connection between seemingly unrelated facts. In a sense, the hypothesis suggests an explanation of what has been observed.
3. *Prediction of the future.* After hypotheses are formulated, their logical implications are deduced. This stage uses the hypotheses to predict what will happen.
4. *Testing the hypotheses.* This is the evidence collection and evaluation stage. From a research project perspective this is the design and implementation of the main study. Conclusions are stated based on the data collected and evaluated.

A simple example will show how the scientific method works. Assume a researcher is performing a marketing research project for a manufacturer of men's shirts:

1. Observation: The researcher notices some competitors' sales are increasing and that many competitors have shifted to a new plastic wrapping.
2. Formulation of hypotheses: The researcher assumes his client's products are of similar quality and that the plastic wrapping is the sole cause of increased competitors' sales.
3. Prediction of the future: The hypothesis predicts that sales will increase if the manufacturer shifts to the new wrapping.
4. Testing the hypotheses: The client produces some shirts in the new packaging and market-tests them.

STAGE 3: RESEARCH METHOD

Whether a particular method of inquiry is appropriate for a research problem depends in large part on the nature of the problem itself and the extent or level of existing knowledge. In addition to selecting a method of inquiry, the research planner must also select a research method.

Two broad methodologies can be used to answer any research question–experimental research and non-experimental research. The major advantage of experimental research lies in the ability to control extraneous variables and manipulate one or more variables by the intervention of the investigator. In non-experimental research, there is no intervention beyond that needed for purposes of measurement.

STAGE 4: RESEARCH DESIGN

Research design is defined as the specific methods and procedures for acquiring the information needed. It is a plan or organizational framework for doing the study and collecting the data. Research designs are unique to a methodology. We discuss research design in depth later in this document and in Chapter 3.

STAGE 5: DATA COLLECTION TECHNIQUES

Research design begins to take on detailed focus as the researcher selects the particular techniques to be used in solving the problem formulated and in carrying out the method selected. A number of techniques available for collecting data can be used. Some techniques are unique to a method of inquiry. For example, many of the qualitative research techniques, such as projective techniques, are used only in subjectivist-type research. In general, data collection uses either communication or observation.

Communication involves asking questions and receiving responses. This process can be done in person, by mail, by telephone, by e-mail, and over the Internet. In most instances this constitutes the broad research technique known as the survey. In contrast to this process, data may be obtained by observing present or past behavior. Regarding past behavior, data collection techniques include looking at secondary data such as company records, reviewing studies published by external sources, and examining physical traces such as erosion and accretion.

In order to collect data from communication or observation there must be a means of recording responses or behavior. Thus, the process of measurement and the development of measurement instrument are closely connected to the decision of which data collection technique(s) should be used. The relationship is two-way. That is, the structure and content of the measurement instrument can depend on the data collection technique, and measurement considerations often influence technique selection.

STAGE 6: SAMPLE DESIGN

Rarely will a marketing research project involve examining the entire population that is relevant to the problem. For the most part, practical considerations (e.g., absolute resources available, cost vs. value, etc.) dictate that one use a sample, or subset of the relevant population. In other instances the use of a sample is derived from consideration of the relevant systematic and variable errors that might arise in a project.

In designing the sample, the researcher must specify three things:

1. Where the sample is to be selected
2. The process of selection
3. The size of the sample

The sample design must be consistent with the relevant population, which is usually specified in the problem-formulation stage of the research process. This allows the data obtained from the sample to be used in making inferences about the larger population.

The process of sample selection may be done by probability or non-probability methods. In probability sampling every element in the population has a known nonzero probability (chance) of being selected for inclusion in a study. In contrast, a non-probability sample is one selected on the basis of the judgment of the investigator, convenience, or by some other means not involving the use of probabilities.

STAGE 7: DATA COLLECTION

Data collection begins after the previous six stages of the research process are complete. Data collection, whether by communication or observation, requires the use of data collection personnel which then raises questions regarding managing these people. Because data collection can be costly, firms often utilize outside limited-service research suppliers, particularly when the extent of in-house research activity does not warrant the cost of having permanent data collection personnel. Also, project design may require specialized data collection, which might best be obtained from an outside supplier.

The working relationship between the data collection agency (a so-called field service) and the research supplier or client is a major factor affecting the quality of fieldwork and data collection.

A study of marketing research firms found that the major barriers to the communication of information from clients to research suppliers to field service firms were insufficient information supplied by the client, the research supplier as an intermediary between client and field service firm, and lack of client interest in data collection (Segal & Newberry, 1983).

The major suggestion for improving communication is for clients to provide more information to both suppliers and field service firms. Another way to overcome communication barriers is for the field service to be consulted on such major issues as scheduling, costs, and purpose of the study. Finally, it was suggested that two-way communication with suppliers be established or strengthened. Although this study was conducted more than 20 years ago, these are enduring problems that exist today.

STAGE 8: ANALYSIS AND INTERPRETATION

Data that are obtained and presented in the same form as originally collected are seldom useful to anyone. Data must be analyzed. The data must be edited, coded, and tabulated before performing formal analyses such as statistical tests. The types of analyses that can be properly performed depend upon the sampling procedures, measurement instruments, and data collection techniques used. Consequently, it is imperative that the techniques of analysis, associated descriptive or prescriptive recommendation types, and presentation formats be selected prior to data collection.

STAGE 9: THE RESEARCH REPORT

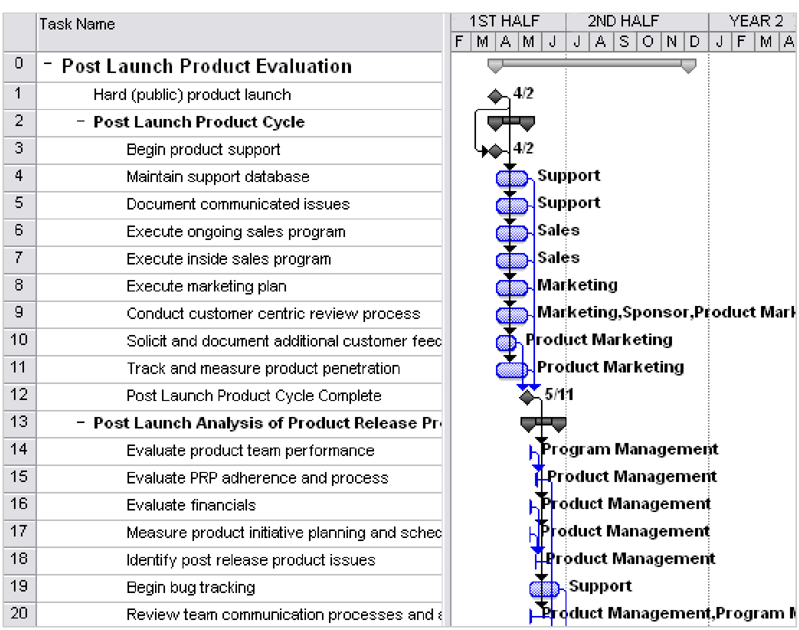
The culmination of the research process is the research report. It includes a clear, accurate, and honest description of everything that has been done and the results, conclusions, and— whenever possible—recommendations for courses of action. Two critical attributes of the report are that it provides all the information readers need using language they understand (completeness) and that it contains selective information chosen by the researcher (conciseness). These attributes are often in conflict with each other.

Two approaches can be taken to ensure that this conflict is not a problem. One approach involves preparing two reports: (1) a technical report that emphasizes the methods used and underlying assumptions, and presents the findings in a detailed manner; and (2) a popular report that minimizes technical details and emphasizes simplicity.

The second approach is concerned with how the report is communicated. Because people vary a great deal in how they are affected by different forms of communication, the ideal reporting process should try to encompass all major forms. Thus, a written report, by itself, may be inadequate and only an invitation to inaction. There are simply a lot of people who, for various reasons, don't respond to the printed word. There are still more that, although they may respond, will often misunderstand the meaning of what is written. For these reasons, it is vitally necessary to get management to sit down with the research manager, or with the researcher and the outside research firm, in a face-to-face reporting situation.

RESOURCE PLANNING FOR YOUR STUDY

When planning for your research, the resources necessary to complete the study should also be identified. Resources include personnel, time and money. Resource plans range from very informal to very formal and may include a list all personnel who will be involved with the project, the exact assignment of each person, the time to be spent, and the pay for each. Additionally you will need to prepare a budget and time schedule for the major activities involved in conducting the study. Microsoft Project or similar software may be helpful in planning and monitoring your research project.



Source: http://office.microsoft.com/en-us/templates/TC012330951033.aspx

THE MAKE OR BUY DECISION

A decision facing all companies that want to use marketing research is who should do the research. Alternatives are to have it done in-house, to utilize outside suppliers, or some combination of the two. In short, sourcing marketing research is a "make or buy" decision. For some companies, this decision is automatic—the in-house organization will do all research unless it is beyond their technical expertise. Other companies with in-house capabilities treat the internal units the same as outside suppliers: they must compete with outside suppliers by preparing proposals and making bids for the business. Almost all research users will at some time require the services of outside research suppliers.

Outside suppliers range from a full-service marketing research agency such as M/A/R/C Research (http://www.marcresearch.com), Burke (http://www.burke.com), and Maritz Marketing Research, Inc. (http://www.maritz.com), to a specialized survey software company like Qualtrics (http://www.qualtrics.com) that provides sophisticated, yet easy to use online data collection and analysis tools, combined with online training, customer support, respondent panels, and analytical services for the client. Clients can negotiate with full-service companies to perform only limited services, for example, research design and data collection only, if that is all the client wants. Thus, there are many variations in the way outside suppliers are used.

When might the use of an outside research supplier be appropriate? There are a number of situations that may call for the use of such firms:

1. The capabilities or technical expertise of in-house researchers are not adequate.
2. You are not able to hire needed personnel.
3. The outside supplier has the needed facilities for doing the research, such as those needed for focus groups or laboratory experiments.
4. A research firm has a demonstrated expertise in a specific industry.
5. There is no unused capacity in the in-house research organization.
6. Lack of objectivity on the part of in-house personnel
7. The outside research supplier can do the research quicker.
8. Some aspect such as data collection may be cheaper when done by an outside supplier (purchasing supermarket-based scanner data may be less costly than collecting it yourself).
9. There is a need for anonymity or confidentiality that may be provided best by an outside research firm.
10. The results of the research may be used in legal proceedings. If so, the outside research firm may have more credibility in the eyes of the court or regulatory or legislative body.

Exhibit 1.5 How to Develop a Successful Research-Consulting Relationship

Some rules of thumb for developing a quality relationship with a research client warrant consideration (proposed by Schmalensee, 2001). For the most part these represent adaptation of more standard techniques and methods to fit a B2B situation. These suggestions are organized around the typical flow of a research project:

1. ***Design research to foster customer relationships.*** This applies to all stages of a project. The research process should be designed to strengthen relationships with business customers.
2. ***Lay the groundwork.*** It is suggested that the researcher allow extra time to talk with the staff, especially those with customer contact. In B2B situations there may be many people who have customer contact, and their views may differ enough that it is beneficial to talk with as many as possible.

3. ***Select and draw the samples.*** There may be a choice of respondents within each business organization, including senior executives, second tier administrators and even customers. In a typical project it is often difficult to decide which type of respondent to contact. One way to overcome this is to interview all major types identified as having relevant information for the problem at hand, although the questions asked each type of respondent will differ.
4. ***Select the research approach and methodology.*** Business respondents tend to be busy people, so it is important to be creative in selecting data collection methodologies. For example a combined telephone and Internet may lead to better information that a telephone or mail questionnaires. Understanding how your target respondent can best be contacted can be helpful in selecting the best method for the majority of the sample.
5. ***Design the questions.*** Keep the questionnaire as "short and sweet" as possible. This, of course, applies to all research projects. Business respondents will be more likely to respond if the questions are interesting and allow them to respond in their own words in a conversational way.
6. ***Record and analyze the data.*** Much of the information collected in B2B research is qualitative, making the analysis crucial.
7. ***Report the results.*** A good way to increase credibility and ensure that results lead to action is to personalize results. This includes use of individual respondent anecdotes and other humanizing details.
8. ***Plan, communicate, and act.*** A good way to increase response rates and build relationships with customers is to share with them what has been learned and what is planned. Communicating with customers allows a company to involve them in implementing whatever action the research suggests. This, again, is part of relationship building.

EXHIBIT 1.6 Basic Research Issues

As technology advances, marketing researchers are continually looking for ways to adapt new technology to the practice of research. Both hardware and software are involved in such adaptations. However, researchers must never forget that research basics cannot be overlooked. Rather, what must be done is to adapt the new techniques and technologies to these basics. All studies must address the following basic issues (Anderson, Berdie, & Liestman, 1984):

1. ***Ask the right questions.*** This is the essence of project design, and the heart of proper planning. The research planner must remember that every project is unique, and as such must be tailored to the user's needs.
2. ***Ask the right people.*** Sample design should be such that only those people who are of interest to the research user are contacted, and such that those who are contacted are reasonably representative of the group of interest.
3. ***Ask questions the right way.*** It is not enough to be able to ask the right questions; they must be asked in the right way. This is the essence of questionnaire design. The researcher can use all the aids available from the new technologies, but if the wording of the questions is not clear to the respondents, the results will be useless. Always pretest the questionnaire to ensure that responses are the ones that are needed to address the problem.
4. ***Obtain answers to questions.*** The process of data collection is central to all marketing research. Techniques used should be selected for how each bears on nonresponse and response alike.
5. ***Relate answers to the needs of the research user/client.*** Data seldom speak for themselves. Proper data analysis is needed if a study is to have any value to the user. Here there is a risk of letting advanced techniques become the master of the researcher rather than the opposite. Common sense is a valuable tool for the researcher when considering alternative analysis approaches for any project.
6. ***Communicate effectively.*** Many good projects are ruined in this stage. The information that is reported to the user should be in a form that is understandable to the user so that he or she can tell that it is relevant to the issues at hand.

Ethical Considerations in Survey Research

There are a number of ethical considerations that arise both in conducting marketing research projects and in marketing related activities in general. Many of these ethical issues are the result of marketing activities that are conducted under the guise of surveys. Exhibit 1.7 summarizes the major practices that are considered unethical, these being deceptive and fraudulent practices, invasion of privacy, and lack of consideration for research subjects and respondents.

EXHIBIT 1.7 Ethical Considerations in Treatment of Subjects and Respondents

Schneider (1977) enumerated three general areas of ethical concern: deceptive practices, invasion of privacy, and lack of consideration. An additional concern too frequent interviewing of the respondent.

Deceptive or fraudulent practices include the following:

- Unrealized promise of anonymity
- Use of disguised questionnaires and interviews
- Faked sponsor identification
- Implication of required response
- Lying about research procedure
- Faked testing in experimental research
- Promise of undelivered compensation
- Sales solicitation disguised as research

Invasions of privacy includes the following examples:

- Observation without informed consent
- Questions concerning people other than the subject
- Projective techniques
- Personal classification data
- Full disclosure and use of "optional" participation

Lack of consideration for subjects or respondents is exhibited in all of the following practices:

- Overuse of public (i.e., unreasonable demands on the time and energy of respondents)
- Research in subject areas with a depressing effect on respondents
- Subjects of no immediate interest to respondents
- Poor interviewers
- Contacts at inconvenient times
- No mention of procedural aspects
- Failure to debrief
- Failure to present subject with option to discard results upon completion

Subjects' rights are an important consideration in the ethical treatment of research participants. Respondents have rights that should not be ignored or violated. Research should not be deceptive or coerced. The researcher is often in a position of authority and as such should assure that the participant does not feel forced to comply, has the ability to choose and make informed choices, is safe from stress, psychological and physical harm, providing information detrimental to their self-interest, and have the right to be informed of the purpose of the research.

Furthermore, promises of anonymity must be kept. Exhibit 1.8 identifies the ethical issues involved in subjects' rights.

The American Marketing Association has provided a statement identifying principles of ethical practice of marketing research (Exhibit 1.9). These broad guidelines provide standards for the protection of the marketer and respondent alike and even extend to researchers and to marketers who are not engaged in research activities.

Exhibit 1.8 Ethical Questions Regarding Subjects' Rights

Subjects' Rights	***Possible Results of Violation of Rights***
A. The right to choose	
a. Awareness of right	*a.* Feelings of forced compliance, biased data
b. Adequate information for an informed choice	b. May violate the client's desire for anonymity, may enable subjects to enact subject role
c. Opportunity to make a choice	*c.* Subjects may avoid environments where this right is violated
B. The right to be safe	
a. Protection of anonymity	*a.* Biased data, refusal to participate in future research
b. Subjects' right to be free from stress	*b.* Biased data, refusal to participate in future research
C. The right to be informed	
a. Debriefing	a. Unrelieved stress, feelings of being used, refusal to participate in future research
b. Dissemination of data	b. Subjects may feel that they gain nothing from and are exploited by participating in research and consequently may distort their response and decline to participate in future research
c. Right to not be deceived	c. Biased data, refusal to participate in future research

SOURCE: "Ethics in Marketing Research: Their Practical Relevance," by Tybout, A.M. & Zaltman, G., in *Journal of Marketing, 11,* p. 359. November, 1974. Published by the American Marketing Association

Exhibit 1.9 Ethics in Marketing Research

The American Marketing Association on Ethics in Marketing Research

The American Marketing Association has established principles of ethical practice of marketing research for the guidance of its members. Marketing management must acknowledge its obligation to protect the public from misrepresentation and exploitation under the guise of research. Similarly, the research practitioner has an obligation to the discipline and to those who provide support for it—an obligation to adhere to basic and commonly accepted standards of scientific investigation as they apply to the domain of marketing research.

FOR RESEARCH USERS, PRACTITIONERS, AND INTERVIEWERS

1. No individual or organization will undertake any activity which is directly or indirectly represented to be marketing research, but which has as its real purpose the attempted sales of merchandise or services to some or all of the respondents interviewed in the course of the research.
2. If respondents have been led to believe, directly or indirectly, that they are participating in a marketing research survey and that their anonymity will be protected, their names shall not be made known to any one outside the research organization or research department, or used for other than research purposes.

FOR RESEARCH PRACTITIONERS

1. There will be no intentional or deliberate misrepresentation of research methods or results. An adequate description of methods employed will be made available upon request to the sponsor of the research. Evidence that fieldwork has been completed according to specifications will, upon request, be made available to buyers of the research.
2. The identity of the survey sponsor and/or the ultimate client for whom a survey is being done will be held in confidence at all times, unless this identity is to be revealed as part of the research design. Research information shall be held in confidence by the research organization or department and not used for personal gain or made available to any outside party unless the client specifically authorizes such release.
3. A research organization shall not undertake marketing studies for competitive clients when such studies would jeopardize the confidential nature of client-agency relationships.

FOR USERS OF MARKETING RESEARCH

1. A user of research shall not knowingly disseminate conclusions from a given research project or service that are inconsistent with or not warranted by the data.
2. To the extent that there is involved in a research project a unique design involving techniques, approaches, or concepts not commonly available to research practitioners, the prospective user of research shall not solicit such a design from one practitioner and deliver it to another for execution without the approval of the design originator.

FOR FIELD INTERVIEWERS

1. Research assignments and materials received, as well as information obtained form respondents, shall be held in confidence by the interviewer and revealed to no one except the research organization conducting the marketing study.
2. No information gained through a marketing research activity shall be used, directly or indirectly, for the personal gain or advantage of the interviewer.
3. Interviews shall be conducted in strict accordance with specifications and instructions received.
4. An interviewer shall not carry out two or more interviewing assignments simultaneously, unless authorized by all contractors or employers concerned. Members of the American Marketing Association will be expected to conduct themselves in accordance with the provisions of this code in all of their marketing research activities.

SUMMARY

This chapter has introduced the research process planning from the perspective of valuing research on the basis of how well it has been done (the management of total error). Planning a research project includes:

- Problem formulation
- Method of inquiry
- Research method
- Research design
- Selection of data collection techniques
- Sample design
- Data collection
- Analysis and interpretation of data
- Research reports

The differences in client requirements give rise to different requirements in research projects. The value versus cost orientation is an outgrowth of these differences.

In this chapter we explained how managers use marketing research to help decision-making. Marketing research provides the description and explanation required in order to predict and evaluate market opportunities. There are several groups who "do" marketing research but the basic reason for their research is the same—to solve a problem. Whether implicit or explicit, the best method for decision making is the problem-solving model. Information should be accurate, current, sufficient, available and, most important, relevant to be meaningful to organizations.

There can be many types of dialogue and challenges between manager and researcher. The dialogue can encompass objectives, courses of action, and environmental variables affecting decision outcomes. And there are ethical issues in marketing research. Although there are many concerns among researchers and others, they can be summarized as deceptive/fraudulent practices, invasion of privacy, and lack of consideration for subjects/respondents. Ethical dilemmas arise because of the relationships that exist between a researcher and stakeholders in the research process. Professional codes of conduct of marketing research were presented, which are indicative of an industry that is trying to "clean up its own act."

REFERENCES

American Marketing Association: http://www.marketingpower.com/AboutAMA/Pages/Statement%20of%20Ethics.aspx?sq=ethics+statement

Anderson, J. F., Berdie, D. R., & Liestman, R. (1984, January 16). Hi-tech techniques OK, but don't forget research basics. *Marketing News, 18* (Sec. 2), 12.

CBS New Coke News Clip: http://www.youtube.com/watch?v=-doEpVWFLsE&NR=1&feature=fvwp

Diesing, P. (1966, March–June). Objectivism vs. subjectivism in the social sciences. *Philosophy of Science, 33,* 124–133.

Kerlinger, F. N., & Lee, H. (2000) *Foundations of behavioral research* (4th ed.). Belmont, CA: Wadsworth Publishing.

Microsoft Project: http://office.microsoft.com/en-us/templates/TC012330951033.aspx

New Coke Commercial: http://www.youtube.com/watch?v=o4YvmN1hvNA

New Coke and Coke Classic Commercial: http://www.youtube.com/watch?v=ky45YGUA3co

Qualtrics.com (2009), 20 Different Types of Marketing Surveys: http://www.qualtrics.com/wiki/index.php/Market_Surveys.

Schmalensee, D. (2001, November 19). One researcher's rules of thumb for B-to-B arena. *Marketing News, 35,* 17–19.

Schneider, K. C. (1977, Spring). Subjects and respondent abuse in marketing research. *MSU Business Topics,* pp. 13–19.

Segal, M. N., & Newberry, C. (1983,Winter). On the field service agency-supplier/client relationships: Problems and perspectives. *Journal of Data Collection, 23,* 58–59.

Tybout, A.M. & Zaltman, G. (1974). Ethics in Marketing Research: Their Practical Relevance, in *Journal of Marketing, 11,* p. 359. November, 1974. Published by the American Marketing Association

This page is intentionally blank

Chapter 2
DEFINING THE RESEARCH DESIGN AND CONTROLLING RESEARCH ERRORS

A research design specifies the methods and procedures for acquiring the information needed to structure and solve the research problem. The overall operational design for a research project stipulates what information is to be collected, from what sources, and by what procedures. A good research design ensures that the information obtained is relevant to the research problem, and that it is collected by objective and economical procedures. A research design might be described as a series of advance decisions that, taken together, form a master plan or model for conducting a research investigation.

Research designs vary depending on the type of study. Generally designs are associated with three types of studies, those that focus on providing exploratory research, descriptive research and causal research. Each will be described in turn.

Exploratory Studies

The major purposes of exploratory studies are for the identification of problems, the precise formulation of problems (including the identification of relevant variables), and the formulation of new alternative courses of action.

An exploratory study is often the first in a series of projects. That is, an exploratory study is often used as an introductory phase of a larger study, and its results are used to bring focus to the larger study and to develop specific techniques that will be used. Thus flexibility is a key to designing and conducting exploratory studies.

We can distinguish three separate tasks that are usually included in exploratory studies and that are typically conducted in the sequence listed:

- A search of secondary information sources
- Interviews with persons knowledgeable about the subject area
- The examination of analogous situations

Search Secondary Sources

Secondary sources of information are the “literature” on the subject. It is the rare research problem for which there is no relevant information to be found by a relatively quick and inexpensive search of the literature. Secondary information sources are not limited to external sources. Searches should also be made of company records.

Interview Knowledgeable Persons

Having searched secondary sources, it is usually desirable to talk with persons who are well informed in the area being investigated, including company executives, experts, consumers and mavens, and users outside the organization

A widely used technique in exploratory research is the focus group. In focus group interviews, a group of knowledgeable people participate in a joint interview that does not use a structured question-and-answer methodology. The group, usually consisting of 8 to 12 people (but may have as few as 5 or as many as 20), is purposely selected to include individuals who have a common background, or similar buying or use experience, as related to the problem being researched. The interviewer or moderator of the focus group session works with the client to develop a general discussion outline that typically includes such topics as usage experience, problems with use, and how decisions are made. The objective is to foster involvement and interaction among the group members during the interview that will lead to spontaneous discussion and the disclosure of attitudes, opinions, and information on present or prospective buying and use behavior.

Focus groups are used primarily to identify and define problems, provide background information, and generate hypotheses. Focus groups typically do not provide solutions for problems. Areas of application include detecting trends in lifestyles, examining new product or service concepts, generating ideas for improving established products or services, developing creative concepts for advertising, and determining effective means of marketing products or services.

If the sole purpose is to create ideas, then individual interviews may be a better alternative than focus groups. Limited research on this issue conducted more than 20 years ago suggests that the number and quality of ideas generated may be greater from such interviews (Fern, 1982).

More specific uses of focus groups include:

1. Identifying and understanding consumer language relating to the product category in question. What terms do they use? What do they mean?

2. Identifying the range of consumer concerns. How much variability is there among consumers' perception of the product, and in the considerations leading them to accept or reject the product?

3. Identifying the complexity of consumer concerns. Do a few simple attitudes govern consumer reaction toward the product, or is the structure complex, involving many contingencies?

4. Identifying specific methodological or logistical problems that are likely to affect either the cost of the subsequent research, or one's ability to generate meaningful, actionable findings.

An example of focus group usage might be to determine the reasons for the decline in a product's overall rating, as reported in a syndicated research report.

Examine Analogous Situations

It is also logical that a researcher will want to examine analogous situations to determine what else can be learned about the nature of the problem and its variables. Analogous situations include case histories and simulations. More discussion of the use of focus groups is given in Chapter 4.

Descriptive Studies

Much research is concerned with describing market characteristics or functions. For example, a market potential study may describe the number, distribution, and socioeconomic characteristics of potential customers of a product. A market-share study finds the share of the market received by both the company and its major competitors. A sales analysis study describes sales by territory, type of account, size or model of product, and the like. In marketing, descriptive studies are also made in the following areas:

- Product research: identification and comparison of functional features and specifications of competitive products
- Promotion research: description of the demographic characteristics of the audience being reached by the current advertising program
- Distribution research: determining the number and location of retailers handling the company's products that are supplied by wholesalers versus those supplied by the company's distribution centers
- Pricing research: identifying competitors' prices by geographic area

These examples of descriptive research cover only a few of the possibilities. Descriptive designs, often called observational designs by some researchers, provide information on groups and phenomena that already exist; no new groups are created (Fink, 2003).

One example of a descriptive study is one conducted by a school-employees credit union in order to gain information useful to provide better service to its members. Management knew very little about the members, other than that they were school employees, family members of employees, or former employees. In addition, the credit union knew very little about member awareness and use of, and attitudes toward individual services available to them. Consequently, investigators undertook a study to answer the following research questions:

1. What are the demographic and socioeconomic characteristics of primary members?
2. How extensively are existing services being used, and what are members' attitudes toward such services?
3. What is the degree of interest in specific new services?

Although associations can be used only to make inferences, and not establish a causal relationship, they are often useful for predictive purposes. It is not always necessary to understand causal relations in order to make accurate predictive statements. Descriptive information often provides a sound basis for the solution of marketing problems, even though it does not explain the nature of the relationship involved. The basic principle involved is to find desirable behavior correlates that are measurable when the predictive statement is made.

Causal Studies

Although descriptive information is often useful for predictive purposes, where possible we would like to know the causes of what we are predicting—the "reasons why." Further, we would like to know the relationships of these causal factors to the effects that we are predicting. If we understand the causes of the effects we want to predict, we invariably improve our ability both to predict and to control these effects.

Bases for Inferring Causal Relationships

There are three types of evidence that can be used for drawing inferences about causal relationships:

1. Associative variation
2. Sequence of events
3. Absence of other possible causal factors

In addition, the cause and effect have to be related. That is, there must be logical implication (or theoretical justification) to imply the specific causal relation.

Associative Variation

Associative variation, or "concomitant variation," as it is often termed, is a measure of the extent to which occurrences of two variables are associated. Two types of associative variation may be distinguished:

1. Association by presence: A measure of the extent to which the presence of one variable is associated with the presence of the other

2. Association by change: A measure of the extent to which a change in the level of one variable is associated with a change in the level of the other.

It has been argued that two other conditions may also exist, particularly for continuous variables: (a) the presence of one variable is associated with a change in the level of other; and (b) a change in the level of one variable is associated with the presence of the other (Feldman, 1975).

Sequence of Events

A second characteristic of a causal relationship is the requirement that the causal factor occur first; the cause must precede the result. In order for salesperson retraining to result in increased sales, the retraining must have taken place prior to the sales increase.

Absence of Other Possible Causal Factors

A final basis for inferring causation is the absence of any other possible causes other than the one(s) being investigated. If it could be demonstrated, for example, that no other factors present could have caused the sales increase in the third quarter, we could then logically conclude that the salesperson training must have been responsible.

Obviously, in an after-the-fact examination of an uncontrolled result such as an increase in detergent sales, it is impossible to clearly rule out all causal factors other than salesperson retraining. One could never be completely sure that there were no competitor-, customer-, or company-initiated causal factors that would account for the sales increase.

Conclusions Concerning Types of Evidence

No single type of evidence, or even the combination of all three types considered, can ever conclusively demonstrate that a causal relationship exists. Other unknown factors may exist. However, we can obtain evidence that makes it highly reasonable to conclude that a particular relationship exists. Exhibit 2.1 shows certain questions that are necessary to answer.

EXHIBIT 2.1 Issues in Determining Causation

Several questions arise when determining whether a variable *X* has causal power over another variable, *Y:*

1. What is the source of causality—does *X* cause *Y,* or does *Y* cause *X?*
2. What is the direction of causality—does *X* positively influence *Y,* or is the relationship negative?
3. Is *X* a necessary and sufficient cause—or necessary, but not sufficient cause—of *Y?* Is *X*'s causation deterministic or probabilistic?
4. Which value of the believed cause exerts a causal influence—its presence or absence?
5. Are the causes and effects the states themselves or changes in the states? Is the relationship static or dynamic?

In the end, the necessary conditions for causality to exist are a physical basis for causality, a cause that temporally precedes the effect (even for associative variation), and a logical reason to imply the specific causal relation being examined. (Monroe and Petroshius, n.d.).

SOURCES OF MARKETING INFORMATION

There are five major sources for obtaining marketing information. In this section we briefly describe each as an introduction to subsequent chapters that describe some of these sources in more depth.

- Secondary sources
- Respondents
- Natural experiments
- Controlled experiments
- Simulation

Secondary Sources of Information

Secondary information is information that has been collected by persons or agencies for purposes other than the solution of the problem at hand.

> If a furniture manufacturer, for example, needs information on the potential market for furniture in the Middle Atlantic States, many government and industry sources of secondary information are available.
>
> The federal government collects and publishes information on the numbers of families, family formation, income, and the number and sales volume of retail stores, all by geographic area. It also publishes special reports on the furniture industry. Many state and local governments collect similar information for their respective areas.
>
> The trade associations in the furniture field collect and publish an extensive amount of information about the industry. Trade journals are also a valuable source of secondary information, as are special studies done by other advertising media.

Private research firms collect specialized marketing information on a continuing basis and sell it to companies. These so-called syndicated services, particularly those for packaged consumer goods, are becoming more sophisticated as they are increasingly becoming based on scanner data. Technology advancements are having a measurable impact on the availability of secondary data.

Information from Respondents

A second major source of information is obtained from respondents. Asking questions and observing behavior are primary means of obtaining information whenever people's actions are being investigated or predicted.

The term respondent literally means "one who responds or answers."
Both verbal and behavioral responses should be considered.

In this book we shall consider both the information obtained from asking people questions, and that provided by observing behavior (or the results of past behavior) to comprise information from respondents.

Information from Natural and Controlled Experiments

As described earlier, three types of evidence provide the bases for drawing inferences about causal relationships. While both natural and controlled experimental designs are capable of providing associative variation and sequence of events, only controlled experiments can provide reasonably conclusive evidence concerning the third type of evidence, the absence of other possible producers.

A natural experiment is one in which the investigator intervenes only to the extent required for measurement. That is, there is no manipulation of an assumed causal variable. The investigator merely looks at what has happened. As such, the natural experiment is a form of ex post facto research. In this type of study, the researcher approaches data collection as if a controlled experimental design were used. The variable of interest has occurred in a natural setting, and the researcher looks for respondents who have been exposed to it and also, if a control group is desired, respondents who have not been exposed.

Measurements can then be made on a dependent variable of interest. For example, if the impact of a television commercial on attitudes were desired, the investigator would conduct a survey of people after the commercial was shown. Those who saw the commercial would constitute the experimental group, and those who did not see it would be a type of control group. Differences in attitudes could be compared as a crude measure of impact. Unfortunately, one can never be sure whether the obtained relationship is causal or non-causal, since the attitudes may be affected by the presence of other variables. For a brief discussion of natural experiments, see Anderson (1971).

In controlled experiments, investigator intervention is required beyond that needed for measurement purposes. Specifically, two kinds of intervention are required:

1. Manipulation of at least one assumed causal variable
2. Random assignment of subjects to experimental and control groups

The researcher conducts the experiment by assigning the subjects to an experimental group where the causal variable is manipulated, or to a control group where no manipulation of the causal variable occurs. The researcher measures the dependent variable in both situations and then tests for differences between the groups. Differences between the groups, if present, are attributed to the manipulation variable.

Field experiments are increasingly being completed using online survey instruments. For example, researchers often use the advanced branching logic, randomization, question block presentation, question timing, and java scripting capabilities of Qualtrics.com to conduct time and cost effective field experiments.

Simulation

The expense and time involved in the personal interviews often associated with field experimentation may preclude it as a source of information for a particular operational situation. In such cases it may be desirable to construct a model of the operational situation and to experiment with it instead of venturing into a real-world situation. The manipulation of such models is called simulation.

Simulation can be defined as a set of techniques for manipulating a model of some real-world process to find numerical solutions that represent the real process being modeled. Models that are environmentally rich (that is, that contain many complex interactions and nonlinear relationships among the variables, probabilistic components, time dependencies, etc.) are usually too difficult to solve by standard analytical methods such as calculus or other mathematical programming techniques. Rather, the analyst views a simulation model as a limited imitation of the process or system under study and attempts to run the system on a computer to see what would happen if a particular policy were put into effect.

Simulations may be used for research, instruction, decision-making, or some combination of these applications. During the past 50 or more years, simulations have been developed for such marketing decision- making applications as marketing systems, marketing-mix elements (new-product, price advertising, and sales-force decisions), and interviewing costs in marketing surveys.

TYPES OF ERRORS AFFECTING RESEARCH DESIGNS

The marketing research process (and research design) involves the management of error. Potential errors can arise at any point from problem formulation through report preparation, and rarely will a research project be error-free. Consequently, the research designer must adopt a strategy for managing and minimizing this error. As we shall see in the next section of this chapter, there are alternative strategies that can be followed.

The objective underlying any research project is to provide information that is as accurate as possible. Maximizing accuracy requires that total study errors be minimized. Total study error has two components—sampling error and non-sampling error—and can be expressed as follows:

Total error = Sampling error + Non-sampling error

Total error is usually measured as total error variance, also known as the mean-squared error (Assael & Keon, 1982):

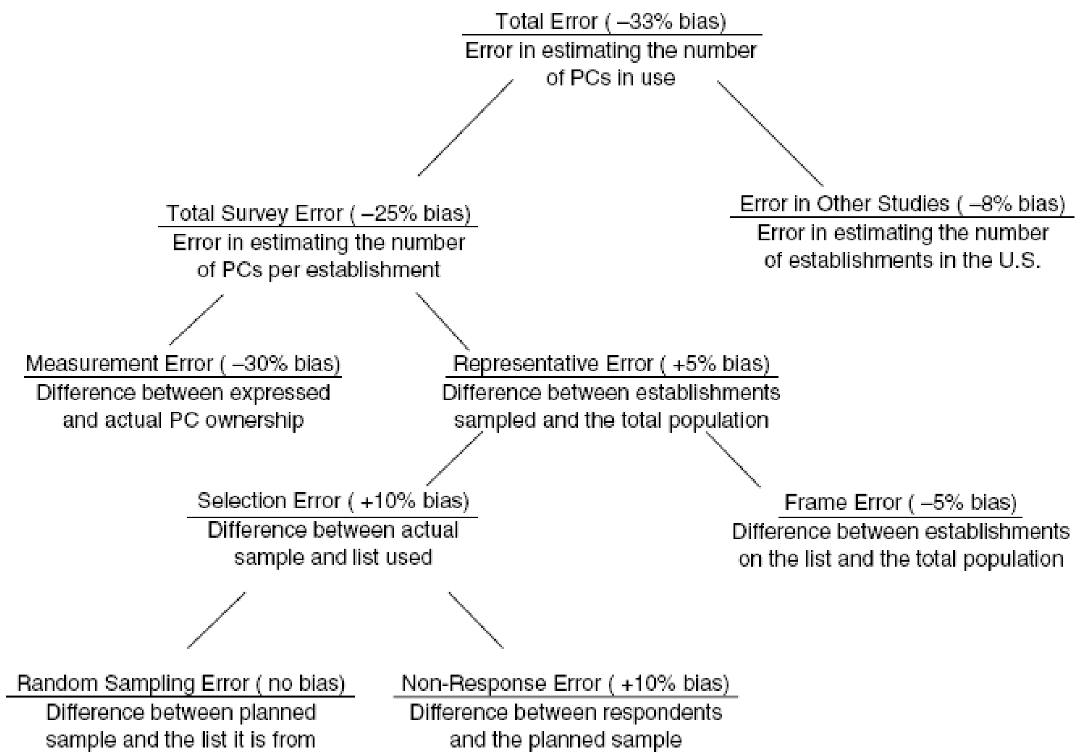
$$(\text{Total error})^2 = (\text{Sampling error})^2 + (\text{Non-sampling error})^2$$

Sampling error refers to the variable error resulting from the chance specification of population from elements according to the sampling plan. Since this introduces random variability into the precision with which a sample statistic is calculated, it is often called random sampling error. Exhibit 2.2 gives an illustration of how total error is assessed.

EXHIBIT 2.2 How Errors Add Up

It is important to know all the sources of error that contribute to inaccuracy, and to assess the impact of each. As an example, consider the figure below, which shows components of error in a study designed to estimate the size of the personal computer market (Lilien, Brown, & Searls, 1991).

When estimating the market, adjustments are made for each source of error. The components are then combined mathematically to create the total error. For purposes of simplicity, total error is shown here as the sum of the component errors. In actuality, total error would be smaller, as it is usually based on the square roots of summed squares of component errors. Assessing the individual components of total error is highly judgmental and subjective, but it is worth the effort.



SOURCE: Reprinted with permission from "How Errors Add Up," by Lilien, G., Brown, R., & Searls, K. in *Marketing News, 33,* January 7, 1991. Published by the American Marketing Association.

Non-sampling error consists of all other errors associated with a research project. Such errors are diverse in nature and are often thought of as resulting in some sort of bias, which implies systematic error. Bias can be defined simply as the difference between the true value of that which is being measured and the researchers' estimate of the true value. However, there can be a random component of non-sampling error. For example, misrecording a response during data collection would represent a random error, whereas using a loaded question would be a systematic error. Non-sampling errors have both nonresponse and response based origins.

To a large extent these major error components are inversely related. Increasing the sample size to reduce sampling error can increase non-sampling error in that, for example, there are more instances where such things as recording errors can occur, and the impact of biased (i.e., nonobjective) questions and other systematic errors will be greater. Thus, this inverse relationship lies at the heart of our concern for total error.

Ideally, efforts should be made to minimize each component. Considering time and cost limitations this can rarely be done. The researcher must make a decision that involves a tradeoff between sampling and non-sampling errors. Unfortunately, very little is known empirically about the relative size of the two error components, although there is some evidence that non-sampling error tends to be the larger of the two. In a study comparing several research designs and data collection methods, Assael and Keon (1982) concluded that non-sampling error far outweighs random sampling error in contributing to total survey error.

As an introduction, Exhibit 2.3 briefly defines eight major types of errors that can influence research results.

EXHIBIT 2.3 Types of Errors in the Research Process

Different types of errors can influence research results:

- **Population specification:** noncorrespondence of the required population to the population selected by the researcher
- **Sampling:** noncorrespondence of the sample selected by probability means and the representative sample sought by the researcher
- **Selection:** noncorrespondence of the sample selected by nonprobability means and the sought representative sample
- **Frame:** noncorrespondence of the sought sample to the required sample
- **Nonresponse:** noncorrespondence of the achieved (or obtained) sample to the selected sample
- **Surrogate information:** noncorrespondence of the information being sought by the researcher and that required to solve the problem
- **Measurement:** noncorrespondence of the information obtained by the measurement process and the information sought by the researcher
- **Experimental:** noncorrespondence of the true (or actual) impact of, and the impact attributed to, the independent variable(s)

Population Specification Error

This type of error occurs when the researcher selects an inappropriate population or universe from which to obtain data.

Examples: Cessna Aircraft conducts an online survey to learn what features should be added to a proposed corporate jet. They consider conducting a survey of purchasing agents from major corporations presently owning such aircraft. However, they believe that that this would be an inappropriate research universe; since pilots are most likely play a key role in the purchase decision.

Similarly, packaged goods manufacturers often conduct surveys of housewives, because they are easier to contact, and it is assumed they decide what is to be purchased and also do the actual purchasing. In this situation there often is population specification error. The husband may purchase a significant share of the packaged goods, and have significant direct and indirect influence over what is bought.

Sampling Error

Sampling error occurs when a probability sampling method is used to select a sample, but the resulting sample is not representative of the population concern.

> **Example**: Suppose that we collected a random sample of 500 people from the general adult population and upon analysis found it to be composed only of people aged 35 to 55. This sample would not be representative of the general adult population. Sampling error is affected by the homogeneity of the population being studied and sampled from and by the size of the sample.

In general, the more homogeneous the population (meaning smaller variance on any given characteristic of interest), the smaller the sampling error; as sample size increases, sampling error decreases. If a census were conducted (i.e., all elements of the population were included) there would be no sampling error.

Selection Error

Selection error is the sampling error for a sample selected by a nonprobability method.

> **Example**: Interviewers conducting a mall intercept study have a natural tendency to select those respondents who are the most accessible and agreeable whenever there is latitude to do so. Such samples often comprise friends and associates who bear some degree of resemblance in characteristics to those of the desired population.

Selection error often reflects people, who are most easily reached, better dressed, and have better kept homes or more pleasant personalities. Samples of these types rarely are representative of the desired population.

Frame Error

A sampling frame is the source for sampling that represents all the members of the population. It is usually a listing of the prospective respondents to be sampled.

> **Example**: Consider the sample frame for a shopper intercept study at a shopping mall. This sample frame includes all shoppers in the mall during the period of data collection. A commonly used frame for consumer research is the telephone directory. This frame introduces error because many elements of the population are not included in the directory (unlisted phone numbers, new arrivals), some elements are listed more than once, and nonpopulation elements are also included (businesses, people who have left the area).

A perfect frame identifies each member of the population once, but only once, and does not include members not in the population of interest.

Nonresponse Error

Nonresponse error can exist when an obtained sample differs from the original selected sample. There are two ways in which nonresponse can occur: (a) noncontact (the inability to contact all members of the sample); and (b) refusal (nonresponse to some or all items on the measurement instrument). Errors arise in virtually every survey from the inability to reach respondents.

Example: In telephone surveys, some respondents are inaccessible because they are not at home (NAH) for the initial call or call-backs. Others have moved or are away from home for the period of the survey. Not-at-home respondents are typically younger with no small children, and have a much higher proportion of working wives than households with someone at home. People who have moved or are away for the survey period have a higher geographic mobility than the average of the population. Thus, most surveys can anticipate errors from non-contact of respondents.

Refusals may be by item or for the entire interview. Income, religion, sex, and politics are topics that may elicit item refusals. Some respondents refuse to participate at all because of time requirements, health issues, past experiences in which an "interviewer" turned out to be a telemarketer, or other reasons. Refusals can also be specific to the method of data collection, as in nonresponse to a mail and email questionnaires or using caller ID to screen and avoid telephone surveys. Nonresponse to mail and email questionnaires sometimes runs as high as 90 percent of the initial mailing, even after several successive mailings.

The amount of effort involved in data collection is another possible way to affect nonresponse error. However, little research has been done to examine the impact of effort.

Example: In a national telephone survey, a so-called five-day "standard" survey was compared to a "rigorous" survey conducted over an eight-week period (Keeter, Miller, Kohut, Groves, & Presser, 2000). Response rates were significantly different; the rigorous survey generated about two-thirds greater response. But the two surveys produced similar results. Most of the statistically significant differences were for demographic items. Very few differences were found on substantive variables.

Nonresponse is also a potential problem in business-to-business and within organization research situations. Although specific respondents are individuals, organizations are not, as they are differentiated and hierarchical. These characteristics may affect organizational response to survey requests.

Tomaskovic-Devey, Leiter, and Thompson (1994) in a study of organizational response, stated the likelihood that an organizational respondent will respond is a function of three characteristics of the respondent:

1. Authority to respond: The degree to which a designated respondent has the formal or informal authority to respond to a survey request
2. Capacity to respond: Organizational practices and the division of labor and information affect the assembly of relevant knowledge to reply adequately
3. Motive to respond: Both individual and organizational motivations to provide information (or not provide information) about the organization.

Surrogate Information Error

In many research situations, it is necessary to obtain information that acts as a surrogate for that which is required. The necessity to accept substitute information arises from either the inability or unwillingness of respondents to provide the information requested.

Decision-oriented behavioral research is always concerned with the prediction of behavior. This limits most marketing research projects to using proxy information, since one cannot observe future behavior. Typically, researchers obtain one or more kinds of surrogate information believed to be useful in predicting behavior.

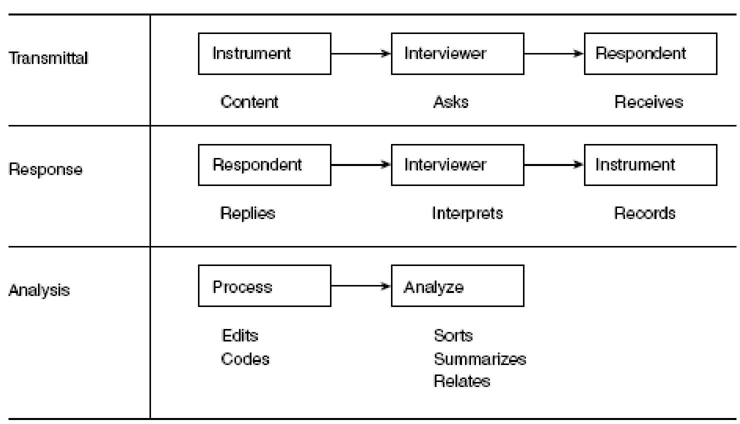
> **Examples**: One may obtain information on past behavior because it is believed that there is sufficient stability in the underlying behavior pattern to give it reasonably high predictive validity. One may ask about intended behavior as a means of prediction. Or one may obtain information about attitudes, level of knowledge, or socioeconomic characteristics of the respondent in the belief that, individually or collectively, they have a high degree of association with future behavior.

Since the type of information required is identified during the problem-formulation stage of the research process, minimizing this error requires an accurate problem definition.

Measurement Error

Measurement error is generated by the measurement process itself, and represents the difference between the information generated and the information wanted by the researcher. Such error can potentially arise at any stage of the measurement process, from the development of an instrument through the analysis of the findings. To illustrate, Figure 2.1 depicts the stages at which errors in eliciting information may arise when interviewing respondents for a survey.

FIGURE 2.1 Potential Sources of Measurement Error in a Survey



In the transmittal stage, errors may be due to the faulty wording of questions or preparation of nonverbal materials, unintentional interviewer modification of the question's wording, or the way in which a respondent interprets the question. In the response phase, errors may occur because the respondent gives incorrect information, the interviewer interprets it incorrectly, or recording errors occur. One aspect of this regards form; form-related errors concern psychological orientation toward responding differently to different item formats and include:

1. Leniency: the tendency to rate something too high or too low
2. Central tendency: reluctance to give extreme scores
3. Proximity: giving similar responses to items that occur close to one another (Yu, Albaum, & Swenson, 2003, p. 217)

In the analysis stage, errors of incorrect editing and coding, descriptive summarization, and inference can contribute substantially to measurement error. Measurement error is particularly troublesome for the researcher, since it can arise from many different sources and take on many different forms.

Experimental Error

When an experiment is conducted, the researcher attempts to measure the impact of one or more manipulated independent variables on some dependent variable of interest, while controlling for the influence of all other (i.e., extraneous) variables. Unfortunately, control over all possible extraneous variables is rarely possible. Consequently, what may be measured is not the effect of the independent variables but the effect of the experimental situation itself.

METHODS FOR DEALING WITH POTENTIAL ERRORS

For any research design, recognizing that potential errors exist is one thing, but doing something about them is another matter. There are two basic approaches for handling potential errors:

1. Minimize errors through precision in the research design
2. Measure or estimate the error or its impact

Minimize Error

Two different approaches can be taken to minimize total error. The first uses the research design to minimize errors that may result from each of the individual error components. Much of the material in Chapters 3 through 9 of this book discusses effective research methods, and as such, involves techniques designed to minimize individual errors. This is consistent with our view that research design innately involves error management. However, this approach is often limited by the budget allotted to a project.

The second approach recognizes that individual error components are not necessarily independent of each other. Thus, attempts to minimize one component may lead to an increase in another. Reducing sampling error by increasing sample size, for example, leads to potentially greater non-sampling error. This means that the research designer must trade off errors when developing a research design that minimizes total error. For a fixed project budget, therefore, it may be prudent for the research designer to choose a smaller sample size (which will increase sampling error) if the cost savings by doing this can develop techniques that will reduce nonresponse and/or improve the measurement process. If the reduction in these nonsampling errors exceeds the increase in sampling error, there will be a reduction in total error.

Estimate or Measure Error

Estimating or measuring individual components and total error is not easy, primarily due to the nature of non-sampling errors. There is a body of accepted sampling theory that allows the researcher to estimate sampling error for a probability sample, but nothing comparable exists for non-sampling errors. Consequently, subjective or judgmental estimates must be made.

For individual error components, many diverse procedures can be used to estimate and measure their impact as illustrated in Table 2.1. These are discussed where appropriate in subsequent chapters.

Table 2.1 Selected Methods for Handling Non-Sampling Errors

Type of Error	*Design to Avoid*	*Measure*	*Estimate*
Surrogate Information	Strive for realism	No method of direct measurement, as event has not yet occurred	Use track record of studies, Use surrogate variables
Measurement **1. Instrument induced**	Pretest, alternative wording, alternative positions, etc.	Experiment by using alternative positioning, etc. in a subsample	Estimate will likely be for no bias but some variable error
2. Interviewing-associated (e.g., bias, recording, cheating in telephone and personal interviews)	Select and train interviewers correctly Use same editor of all interviews by one interviewer Use cheater questions Use computer program to analyze for patterns of responses by interviewer	Re-interview subsample using expert interviewer Analysis of variance Use cheater questions Use computer program to analyze for patterns Use interpenetrating sample	Estimate will be for both bias and variable error
3. Response	Randomize response technique Ask for verification checks Cross-check questions Use mail-back technique	Compare with known data	Have interviewer evaluate respondent Estimate will be for both bias and variable error
4. Editing	Prepare editing manual Train editors Require daily return of data	Use master editor to edit subsample	Estimate will be for limited bias, some variable error
5. Coding (text and manually entered data)	Pre-code variables Use coding manual User computer program to clean data	Use master coder to validate subsample	Some bias and variable error
6. Tabulation	Use verification for data entry	Recheck sample of forms	Variable error
7. Analysis	No remedy except competence	Use more competent analyst	
Frame	Use multiple frames	Take subsample of excluded segments	Use compensating weights Use past data
Selection	Make sample element and sample unit the same Use probability sample	Compare with known population	Use compensating weights
Nonresponse	Use callbacks Call at appropriate time Use trained interviewers	Take subsample of nonrespondents	Use Politz-Simmons method Use wave analysis

As a final note, even though the researcher has designed a project to minimize error, it is almost never completely eliminated. Consequently, the error that exists for every project must be estimated or measured. This is recognized for sampling error when probability samples are used, though non-sampling errors typically are ignored. Although estimating or measuring errors is better than ignoring them, there may be times when ignoring non-sampling error may not be that bad. For example, if non-sampling error is viewed as a multiple of sampling error, ignoring non-sampling errors up to an amount equal to one-half of sampling error reduces a .95 confidence level only to .92 (Tull & Albaum, 1973). However, ignoring a non-sampling error equal in amount to sampling error reduces the .95 level to .83.

CHOOSING A RESEARCH DESIGN

The overview of research designs and sources of error just presented should make it apparent that, given a specified problem, many competing designs can provide relevant information. Each design will have an associated expected value of information and incurred cost.

Suppose, for example, that a researcher is assigned to determine the market share of the ten leading brands of energy drinks. There are many possible ways of measuring market share of energy drink brands, including questioning a sample of respondents, observing purchases at a sample of retail outlets, obtaining sales figures from a sample of wholesalers, obtaining sales figures from a sample of retailers and vending machine operators, obtaining tax data, subscribing to a national consumer panel, subscribing to a national panel of retail stores, and, possibly, obtaining data directly from trade association reports or a recent study by some other investigative agency. Though lengthy, this listing is not exhaustive.

The selection of the best design from the alternatives is no different in principle from choosing among the alternatives in making any decision. The associated expected value and cost of information must be determined for each competing design option. If the design is such that the project will yield information for solving more than one problem, the expected value should be determined for all applicable problems and summed. The design with the highest, positive, net expected payoff of research should be selected.

SUMMARY

In this chapter we dealt with a subject of single most importance to the research project: the research design. We described what a research design is, discussed the classes of designs, and examined major sources of marketing information that various designs employ. Finally, we considered the errors that affect research designs.

Presenting these topics as an introduction and overview, we deal with the topics in more depth in the next several chapters. These chapters deal with major sources of marketing information—respondents and experimentation—and the means of obtaining and analyzing research information.

REFERENCES

Anderson, B. F. (1971). *The psychology experiment: An introduction to the scientific method* (2nd ed.). Belmont, CA: Brooks/Cole.

Assael, H., & Keon, J. (1982, Spring). Non-sampling vs. sampling errors in survey research. *Journal of Marketing, 46,* 114–123.

Feldman, J. (1975). Considerations in the use of causal-correlational technique in applied psychology. *Journal of Applied Psychology, 60,* 663–670.

Fern, E. F. (1982). The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research, 19* (February), 1–13.

Fink, A. (2003). *How to design surveys* (2nd ed.). Thousand Oaks, CA: Sage.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing non-response in a national telephone survey. *Public Opinion Quarterly, 64,* 125–148.

Lilien, G., Brown, R., & Searls, K. (1991, January 7). How errors add up. *Marketing News, 25,* 20–21.

Monroe, K. B., & Petroshius, S. M. (n.d.). Developing causal priorities. Unpublished working paper, College of Business, Virginia Polytechnic Institute and State University.

Tomaskovic-Devey, D., Leiter, J., & Thompson, S. (1994). Organizational survey response. *Administrative Science Quarterly, 39,* 439–457.

Tull, D. S., & Albaum, G. S. (1973). *Survey research: A decisional approach.* New York: Intext Educational Publishers.

Yu, J., Albaum, G., & Swenson, M. (2003). Is a central tendency error inherent in the use of semantic differential scales in different cultures? *International Journal of Market Research, 45* (2), 213–228.

Chapter 3
SECONDARY SOURCES OF INFORMATION

"A few months of field work collecting data can save several hours in the library"

Secondary data (or secondary information) is information that has been collected by persons or agencies for purposes other than the solution of the marketing research problem at hand. These data may have been collected from sources within the researcher's firm or from sources outside the firm. The key point is that the data were collected for some other project, or reason, than the current one.

In contrast, primary data is data collected for the researcher's current research project. Primary data is often collected from a respondent, an individual who provides information either passively through the observation of his or her behavior, or actively through verbal response. Researchers using primary data must be concerned with information obtained by asking questions, by observing behavior or by examining the results of past behavior.

In addition to primary and secondary data, there exists commercial data sold in the form of syndicated services. These data are collected by commercial marketing research firms or industry associations and, as such, have characteristics of both primary and secondary data. Since these data relate to ongoing concerns of a marketer they can be viewed as primary data. However, the commercial agency did not design its service solely to provide information for one company's specific project. Thus, there are elements of secondary data. It should be clear that distinctions between primary and secondary commercial data may be minimal.

In this chapter, we discuss the reasons for obtaining secondary information, types of secondary information, sources of external secondary data, and syndicated services that provide commercial data. Data, in all their forms, are the heart of research. Secondary research can help provide a clearer picture of a problem so that researchers and managers can make the necessary critical decisions.

REASONS FOR OBTAINING SECONDARY INFORMATION

As a general rule, no research project should be conducted without a search of secondary information sources. This search should be conducted early in the problem investigation stage and prior to any organized collection of information from primary sources. There are several reasons for this.

Secondary Information May Solve the Problem

If adequate data are available from secondary sources, primary data collection will not be required. For example, Campbell Soup Co. based a long-running advertising campaign on the theme "soup is good food." This theme emerged from federal government data pertaining to eating habits, nutritional health, and related topics collected over a period of 15 years. Campbells (1959: youtube.com/watch?v=qMJtLP8jMWQ , 1985 youtube.com/watch?v=4PtZYlcyXfI)

Secondary Information Search Costs Substantially Less

A comprehensive search of secondary sources can almost always be made in a fraction of the time and cost required for the collection of primary information. This is particularly true today with online access to research publications and databases. Searching for secondary research helps you avoid duplicating primary research and optimizes research expenditures by acquiring only information that cannot be found elsewhere. Many marketing problems do not warrant expenditures for primary information collection, but are worth the time and cost of secondary information.

Secondary Information Has Important Supplementary Uses

When secondary information cannot solve the research problem, it can provide invaluable supplemental uses:

1. *Defining the problem and formulating hypotheses about its solution.* The analysis of available secondary data will almost always provide a better understanding of the problem and its context, and will frequently suggest solutions not considered previously.

2. *Planning the collection of primary data.* An examination of the methods and techniques employed by other investigators in similar studies may be useful in planning the present one. It may also be of value in establishing classifications that are compatible with past studies so that trends may be more readily analyzed.

3. *Defining the population and selecting the sample.* Past information and samples may help establish classifications for current primary information collection.

The researcher must be careful when using only secondary data. To be useful, secondary data must be available, relevant to the information needs (which includes being timely), accurate, and sufficient to meet data requirements for the problem at hand. Often, little is known about the reliability of secondary research studies. It is important that the researcher know how the secondary data being considered for use were collected, if the data is reliable, and if the right techniques were used.

> **Example**: One company wants to do a segmentation analysis on foreign markets with a particular emphasis on examining demographics (Albaum, Duerr, & Strandskov, 2005, Chap. 5). The company is considering using the official government census of the population. However, they become aware that the data are not available from all markets in equal quantity, aggregation, and detail, and the reliability of data is not the same.
>
> What one gets from a census depends on what was on the census form in the first place—typically a mixture of traditional questions and new items of interest to public policy makers and civil servants at the time. Some countries publish information about noncitizens, and others collect data on religion—both of these topics are ignored in U.S. Censuses. Income is one of the major dimensions of U.S. segmentation research, but many highly developed nations ignore the income question in their censuses.

In short, one cannot always expect to find the same range of data topics that you are interested in. Moreover, the data may not use the same categories when showing relevant distributions of demographic variables, such as age. In short, comparability and equivalence issues arise, and these can hinder your effectiveness in using secondary data.

TYPES OF SECONDARY INFORMATION

Secondary information falls into two categories, data that is available within the company (internal data) and that which must be obtained from outside sources (external data).

Internal Secondary Information

All companies collect information in the everyday course of conducting business. Orders are received and filled, costs are recorded, warranty cards are returned, salespeople's reports are submitted, engineering reports are made—all are collected for other purposes, but may be useful to the researcher (Andreasen, 1988, pp. 77–89). The key, of course is knowing where they are and how to access them. In order to do this efficiently, the firm must have an effective marketing information system.

> **Example**: Spectra Physics Lasers Division (producing laser grocery store scanners) regularly performs customer satisfaction studies. Whereas these studies are primary research to the Retail Systems organization, they are internal secondary information to other divisions that may want to look at them. Also, they can be secondary data to Retail Systems should they be used at a much later date for aiding in decision-making or for purposes other than those originally intended when the studies were done.

External Secondary Information

External secondary information is available in staggering assortments and volumes. It also is applicable to all of the major types of marketing research projects and is mainly concerned with the noncontrollable aspects of the problem:

- Total market size
- Market characteristics
- Competitor products, prices, promotional efforts, and distribution methods

As an example, a consumer goods company is considering whether it should establish a direct selling operation. Direct selling is defined as personal contact between a salesperson and a consumer away from a fixed business location such as a retail store. The Direct Selling Association (DSA) provides secondary information in the form of a regular survey of the industry. Some types of information on industry statistics and sales force demographics are available on a regular basis. Some of the types of industry statistics are the following:

- Estimated U.S. sales
- Estimated U.S. salespeople
- Percent of sales by major product groups
- Location of sales
- Percent of sales by census region
- Sales strategy
- Compensation structure by percent of firms
- Compensation structure by percent of sales dollars
- Compensation structure by percent of salespeople

The following are some of the sales force demographics available:

- Gender
- Age
- Education
- Independent contractor/employee status
- Hours per week dedicated to direct selling
- Average time spent on direct selling tasks
- Main reasons for becoming a direct sales representative
- Percent of salespeople by distributorship type

Age, education, average time spent on direct selling tasks, and main reason for becoming a direct sales representative data are from DSA's National Sales force Survey. (Direct Selling Association, 2003; http://www.dsa.org/pubs/).

SOURCES OF EXTERNAL SECONDARY DATA

The major original sources of external secondary information are:

1. Government (supranational, federal, state, and local)
2. Trade associations and trade press
3. Periodicals and professional journals
4. Institutions (e.g., universities)
5. Commercial services

Government Data Sources

The federal government is by far the largest single source of this type of data. Both governmental and trade sources are so important that the experienced researcher will be thoroughly familiar with them in his or her field of specialization. Periodicals and research publications of universities and research institutes frequently provide valuable information. Commercial services of many types are available that are highly useful for specific research problems.

Market performance studies on consumer products, for example, will normally provide such demographic information as the number of consumers (or consuming units) by age group, income class, gender, and geographic area. Such data are usually available on a reasonably recent basis from censuses conducted by federal, state, local, and, when needed, supranational governments.

Often, a good first source is the *Statistical Abstract of the United States*, available online from the Bureau of the Census http://www.census.gov/compendia/statab/. This reference abstracts data from original reports and gives some useful material on social, political, and economic matters. The source is a good reference to the more detailed data in the original sources.

The State and Metropolitan Area Data Book is a publication of the Bureau of the Census that is available online in PDF format. It provides detailed comparative data on states, metropolitan areas and their component counties, and central cities. It covers information about numerous topics relevant for both B2C and B2B marketing, including population, income, labor force, commercial office space, banking, health care, housing, and so forth. http://www.census.gov/statab/www/smadb.html

The *Census of Population* http://www.census.gov/population/www/ and the *Census of Housing* http://www.census.gov/hhes/www/housing.html taken by the U.S. Department of Commerce every 10 years are the most comprehensive of such censuses. Updates of various census measurements based on smaller yearly surveys are available in *Current Population Reports* http://www.census.gov/main/www/cprs.html and *Current Construction and Housing Reports* http://www.census.gov/prod/www/abs/cons-hou.html. Many other up-to-date estimates are made periodically by governmental and non–governmental agencies.

Data from the U.S. Census Bureau is available online for custom data analysis, on CD-ROM, and in report form as downloadable PDF files. There are, however, private companies that make such data available—for a fee—in more processed form, which, in effect, adds value to the Census Bureau data. The company previously mentioned, GeoLytics (http://geolytics.com/), markets a line of census data products and a variety of custom data retrieval services.

- Demographic reports (and maps)
  - Custom data sets and reports
  - Area segmentation
  - Area-to-area correspondence files
  - Banking and realtor tract level data and maps
- Services
  - Geocoding (GPS) addresses
  - Custom-built databases
  - Normalized data for across census comparisons

Other companies include census data in mapping software that is used for geographic market analysis. This type of software is potentially useful for such applications as retail site analysis, real estate site reports, direct marketing, database creation, and so forth. One supplier is Scan/US, Inc. (http://www.scanus.com), whose software product *Scan/US Streets and Data U.S.A.* includes maps for the entire United States that include all types of demographics.

Private Data Sources

Private organizations are another source of demographic information useful to marketers. To illustrate, SRDS (www.srds.com/frontMatter/ips/lifestyle/index.html) publishes *The Lifestyle Market Analyst.* This annual provides demographic and lifestyle information for 210 Designated Market Areas (DMAs) in the United States. As shown in Table 3.1 and Figure 3.1, this market data can be accessed in graphic and tabular formats for demographic and lifestyle variables:

- *Demographic categories for each DMA:* Start with a specific demographic segment, such as dual-income households, and identify lifestyles and geographic locations.

- *Most popular lifestyles for each DMA:* Specify a lifestyle and then identify what other interests frequently appeal to those consumers and what demographic information corresponds to that profile.

Table 3.1 Demographics and Lifestyles, Seattle-Tacoma, Washington

Seattle-Tacoma, WA

Demographics
Base Index US = 100

Total Adult Population 2,939,296

Occupation	*Population*	*%*	*Index*
Administrative	379,169	12.9	101
Blue Collar	276,294	9.4	93
Clerical	235,144	8.0	93
Homemaker	373,290	12.7	92
Professional/Technical	837,699	28.5	112
Retired	502,619	17.1	92
Sales/Marketing	173,418	5.9	109
Self Employed	88,179	3.0	107
Student	73,482	2.5	104
Education (1990 Census)			
Elementary (0–8 years)	99,748	4.4	42
High School (1–3 years)	226,599	10.0	69
High School (4 years)	621,155	27.4	91
College (1–3 years)	752,641	33.2	133
College (4+ years)	584,481	24.9	123
Race/Ethnicity			
White	2,486,544	84.6	115
Black	114,633	3.9	33
Asian	182,236	6.2	182
Hispanic	111,693	3.8	36
American Indian	41,150	1.4	200
Other	2,939	0.1	100
Total Households	1,539,954		

Age of Head of Household	*Households*	*%*	*Index*
18–24 years old	90,858	5.9	111
25–34 years old	318,773	20.7	108
35–44 years old	378,831	24.6	109
45–54 years old	277,194	18.0	101
55–64 years old	167,856	10.9	67
65–74 years old	164,776	10.7	88
75 years and older	140,137	9.1	87
Median Age	44.5 years		
Sex/Marital Status			
Single Male	318,773	20.7	102
Single Female	346,492	22.5	93
Married	876,240	56.9	102
Children At Home			
At Least One Child	491,249	31.9	102
Child Age Under 2	73,918	4.8	112
Child Age 2–4	135,517	8.8	107
Child Age 5–7	127,817	8.3	102
Child Age 8–10	126,277	8.2	101
Child Age 11–12	92,398	6.0	103
Child Age 13–15	129,357	8.4	99
Child Age 16–18	109,337	7.1	92
Home Ownership			
Owner	964,017	62.6	96
Renter	575,947	37.4	107

Stage in Family Lifecycle	*Households*	*%*	*Index*
Single, 18–34, No Children	197,115	12.8	107
Single, 35–44, No Children	100,098	6.5	108
Single, 45–64, No Children	129,357	8.4	93
Single, 65+ No Children	117,037	7.6	84
Married, 18–34, No Children	84,698	5.5	120
Married, 35–44, No Children	61,599	4.0	121
Married, 45–64, No Children	201,735	13.1	96
Married, 65+ No Children	157,076	10.2	93
Single, Any Child at Home	120,117	7.8	94
Married, Child Age Under 13	223,295	14.5	109
Married, Child Age 13–18	147,837	9.6	98
Household Income			
Under $20,000	309,533	20.1	76
$20,000–$29,999	212,515	13.6	95
$30,000–$39,999	195,575	12.7	98
$40,000–$49,999	194,035	12.6	108
$50,000–$74,999	344,952	22.4	117
$75,000–$99,999	158,616	10.3	126
$100,000 and over	124,737	8.1	114
Median Income	$42,738		
Income Earners			
Married, One Income	385,531	25.1	98
Married, Two Incomes	488,169	31.7	106
Single	683,724	43.1	97
Dual Income Households			
Children Age Under 13 years	130,897	8.5	105
Children Age 13–18 years	103,178	6.7	99
No Children	255,634	16.6	111
Age By Income			
18–34, Income under $30,000	184,796	12.0	103
35–44, Income under $30,000	80,078	5.2	85
45–64, Income under $30,000	92,398	6.0	69
65+ Income under $30,000	164,776	10.7	74
18–34, Income $30,000–$49,999	107,797	7.0	108
35–44, Income $30,000–$49,999	101,638	6.6	108
45–64, Income $30,000–$49,999	104,718	6.8	91
65+ Income $30,000–$49,999	76,468	4.9	109
18–34, Income $50,000–$74,999	76,468	4.9	120
35–44, Income $50,000–$74,999	106,258	6.9	121
45–64, Income $50,000–$74,999	123,197	8.0	110
65+ Income $50,000–$74,999	40,039	2.6	124
18–34, Income $75,000 and over	40,039	2.6	118
35–44, Income $75,000 and over	90,858	5.9	128
45–64, Income $75,000 and over	126,277	8.2	119
65+ Income $75,000 and over	26,179	1.7	113
Credit Card Usage			
Travel/Entertainment	183,256	11.9	88
Bank Card	1,233,511	80.1	104
Gas/Department Store	631,288	34.5	108
No Credit Cards	218,675	14.2	87

Lifestyles

Base Index US = 100

Seattle-Tacoma, WA

The Top Ten Lifestyles Ranked by Index

Snow Skiling Frequently	181	Use an Apple/Macintosh	138
Camping/Hiking	159	Foreign Travel	136
Boating/Sailing	144	Own a Cat	132
Frequent Flyer	140	Real Estate Investments	132
Recreational Vehicles	139	Wines	127

Home Life	*Households*	*%*	*Index*	*Rank*
Avid Book Reading	660,645	42.9	116	5
Bible/Devotional Reading	266,414	17.3	91	159
Flower Gardening	622,145	40.4	119	19
Grandchildren	323,392	21.0	89	184
Home Furnishing/ Decorating	326,472	21.2	95	141
House Plants	489,709	31.8	100	158
Own a Cat	535,907	34.8	132	16
Own a Dog	494,328	32.1	94	183
Shop by Catalog/Mail	445,050	28.9	98	114
Subscribe to Cable Tv	1,030,236	66.9	103	92
Vegetable Gardening	381,911	24.8	109	115

Good Life				
Attend Cultural/ Arts Events	287,973	18.7	119	15
Fashion Clothing	180,176	11.7	85	135
Fine Art/Antiques	192,496	12.5	113	10
Foreign Travel	300,293	19.5	136	12
Frequent Flyer	472,769	30.7	140	8
Gourmet Cooking/ Fine Foods	340,332	22.1	123	10
Own a Vacation Home/ Property	200,195	13.0	124	21
Travel for Business	358,812	23.3	114	17
Travel for Pleasure/ Vacation	640,625	41.6	110	5
Travel in USA	609,826	39.6	108	8
Wines	264,874	17.2	127	11

Investing & Money				
Casino Gambling	201,735	13.1	98	89
Entering Sweepstakes	217,135	14.1	93	177
Moneymaking Opportunities	174,016	11.3	95	138
Real Estate Investments	133,977	8.7	132	11
Stock/Bond Investments	337,252	21.9	116	7

Great Outdoors				
Boating/Sailing	232,535	15.1	144	13
Camping/Hiking	595,966	38.7	159	27
Fishing Frequently	360,352	23.4	95	175
Hunting/Shooting	209,435	13.6	87	177
Motorcycles	129,357	8.4	111	77
Recreational Vehicles	186,336	12.1	139	42
Wildlife/Environmental	281,813	18.3	111	42

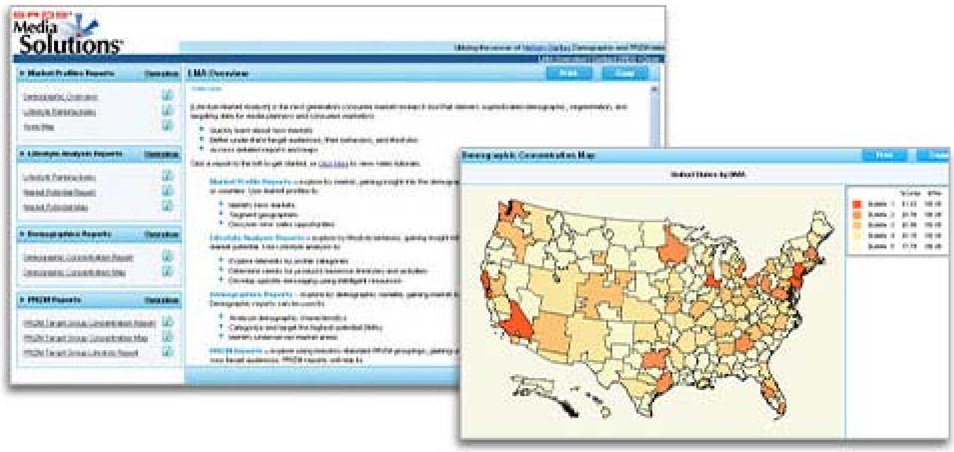
Sports, Fitness & Health	*Households*	*%*	*Index*	*Rank*
Bicycling Frequently	309,533	20.1	111	56
Dieting/Weight Control	309,533	20.1	89	196
Golf	321,852	20.9	105	70
Health/Natural Foods	292,593	19.0	113	23
Improving Your Health	364,971	23.7	100	62
Physical Fitness/Exercise	600,586	39.0	107	23
Running/Jogging	204,815	13.3	114	37
Snow Skilling Frequently	214,055	13.9	181	21
Tennis Frequently	84,698	5.5	93	60
Walking for Health	508,188	33.0	99	133
Watching Sports on TV	569,787	37.0	97	141

Hobbies & Interests				
Automotive Work	234,075	15.2	104	136
Buy Pre-Recorded Videos	304,913	19.8	106	37
Career-Oriented Activities	157,076	10.2	110	30
Coin/Stamp Collecting	103,178	6.7	99	136
Collectibles/Collections	194,035	12.6	101	120
Community/Civic Activities	141,677	9.2	101	117
Crafts	418,870	27.2	100	151
Current Affairs/Politics	277,194	18.0	108	25
Home Workshop	428,570	27.7	107	83
Military Veteran in Household	391,151	25.4	109	79
Needlework/Knitting	247,934	16.1	104	126
Our Nation's Heritage	76,998	5.0	100	116
Self-Improvement	295,673	19.2	104	43
Sewing	271,034	17.6	105	129
Supports Health Charities	283,353	18.4	96	95

High Tech Activites				
Electronics	186,336	12.1	104	50
Home Video Games	192,496	12.5	102	110
Listen to Records/ Tapes/CDs	828,501	53.8	106	16
Own a CD Player	1,036,396	67.3	113	11
Photography	317,233	20.6	116	13
Science Fiction	172,476	11.2	124	8
Science/New Technology	177,096	11.5	126	13
Use a Personal Computer	814,641	52.9	124	13
Use an Apple/Macintosh	195,575	12.7	138	16
Use an IBM Compatible	697,604	45.3	121	14
VCR Recording	298,753	19.4	100	104

SOURCE: Reprinted with permission from SRDS.

Figure 3.1 Map of a Selected Life Style Incidence Across the U.S.



Source: http://www.srds.com/frontMatter/ips/lifestyle/images/LMAA_overview-with-map.jpg

Nielsen's Claritas division, a provider of solutions for geographic, lifestyle and behavioral target marketing, has developed a demographic widget that is available as a free download for personal electronics http://www.claritas.com/target-marketing/nielsen-claritas-demographic-widget.jsp .

Market size studies (e.g., size in sales dollars or units) often are conducted by trade associations, media, firms in the industry, and private research organizations. These studies are published and made available to interested parties. Industry-type studies may be concerned with such types of information as total market size, market characteristics, market segments and their size and characteristics, and similar types of information.

> Example: Mediamark Research, Inc. conducts a single-source continuing survey, primarily aimed at the advertising industry that provides demographics, lifestyles, product usage, and exposure to all advertising media data. One part of this study is a series of studies on specific products/services that is published as syndicated reports.
> http://www.mediamark.com

Information on new products and processes is available from such sources as patent disclosures, trade journals, competitors' catalogs, testing agencies, and the reports of governmental agencies, such as the Food and Drug Administration, the Department of Agriculture, and the National Bureau of Standards.

An extensive amount of information is available concerning advertising.

- Publishers Information Bureau provides a compilation of expenditures by medium for each competitor. http://www.magazine.org/advertising/revenue/index.aspx
- Audit Bureau of Circulations provides data on the numbers of magazine copies sold under specified conditions. http://www.accessabc.com/
- Standard Rate and Data Service provides complete information on the rates and specifications for buying advertising space and time. http://www.srds.com/portal/main?action=LinkHit&frameset=yes&link=ips
- Mediamark Research, Inc. publishes data on multiple major local media markets, relating detailed media behavior to demographic characteristics of readers/viewers/listeners. http://www.mediamark.com

- Arbitron Radio and Television Reports (http://www.arbitron.com/home/content.stm), the Nielsen Radio-Television Index http://ca.acnielsen.com/products/product_list.shtml, Starch Advertising Readership Service http://www.starchresearch.com/services.html, all measure of audience exposure to specific advertisements or programs.

EXHIBIT 3.2 Starch Readership Reports

The best way to create print ads for the future, and for the long term, is to get feedback on a constant basis in order to find out what works and what doesn't.

Each year, Starch measures over 25,000 ads in over 400 magazine issues. On the most basic level you get raw readership scores—the percent of readers who saw the ad and read the copy. Then the data are put into a context: The ad is ranked not only against other ads in the issue but also against other ads in its product category over the past two years. These norms are a fast and easy way to judge the performance of your ad over time and against the competition.

The Benefits of Starch Ad Readership

In-Depth Analysis

- Campaign analyses inform clients not only about the scores of the ads but also why they performed as they did and what can be done to improve the ads. Moreover Starch also is unique in its ability to tell clients about the best positions in various publications (e.g., whether far-forward positioning is superior to ads in the back of the book).

Extra Questions

- To give you information on advertising likeability, persuasiveness and intent to purchase
- Many times, if you ask a publisher to Starch an issue your ads will appear in, they will assume the cost and pass on the data to you for free.

The Starch Ad Readership Program

Through-the-Book, Recognition Method

- One-to-one in-person interviews
- Generally, 100–200 sample, but can be more if client desires
- Sample approximates readership of publication, but is not representative
- Reports present data on:
  - Noted: percent who saw any part of the ad
  - Associated: percent who saw advertiser's name
  - Read Some: percent who read any of the copy
  - Read Most: percent who read more than half the copy
- Most reports also offer indexed scores, based on ads of the same size, color, product category

SOURCE: Adapted from Roper, 2002.

Internet Databases

The Internet has become the staple of research and provides access to most commercial electronic databases. Thousands of such databases are available from numerous subscription systems, such as DIALOG (http://www.dialog.com), LexisNexis (http://www.lexisnexis.com/), or Dow Jones News/Retrieval
http://www.dowjones.com/Products_Services/ElectronicPublishing/EnterpriseMedia.htm.

In general, there are five categories of commercial databases:

1. Bibliographic databases that index publications
2. Financial databases with detailed information about companies
3. Statistical databases of demographic, econometric, and other numeric data for forecasting and doing projections
4. Directories and encyclopedias offering factual information about people, companies, and organizations
5. Full-text databases from which an entire document can be printed out.

The advantages of such current databases are obvious. All that is needed is personal computer with internet access or a CD-ROM/DVD.

Computerized databases have led to an expanded focus on database marketing. Database marketing has been defined as an extension and refinement of traditional direct marketing, but uses databases to target direct response advertising efforts and tracks response and/or transactions. In database marketing, the marketer identifies behavioral, demographic, psychographic, sociological, attitudinal, and other information on individual consumers/households that are already served or that are potential customers. Data may come from secondary and/or primary sources. Qualtrics clients are increasingly APIs (Application Programming Interface) to link and integrate customer databases with survey data and respondent panels. APIs can be used to link and integrate data from multiple sources in real time. Thus, information in database profiles is augmented by new contact and survey data, and can be viewed in dashboards that report current information and can be used to better target and predict market response. Databases can be used to estimate market size, find segments and niches for specialized offerings, and even view current customer use and spending (Morgan, 2001). In short, it helps the marketer develop more specific, effective, and efficient marketing programs.

Today, data mining is in high demand as a research focus. Data mining involves sifting through large volumes of data to discover patterns and relationships involving a company's products and customers. Viewed as a complement to more traditional statistical techniques of analysis, two of the more powerful data mining techniques are neural networks and decision trees (Garver, 2002). Further discussion of data mining techniques is beyond the scope of this text, but good discussions are found in Berry & Linoff (1997, 2002) and Dehmater & Hancock (2001).

EXHIBIT 3.3 Attaining Market Knowledge From Online Sources

Elkind and Kassel (1995) provided essential guidelines for attaining market knowledge from online sources:

- *Develop an online research plan.* The plan will outline all the key areas of inquiry and will provide a systematic pathway to search, retrieve, and arrive at the desired data, information, and knowledge.
- *Clearly define your information needs, knowledge gaps, and issue to be resolved.* One of the best ways is to do a knowledge inventory or review to determine what you already know or have in your possession, both primary and secondary research.
- *Focus the search.* Start by applying the learning from your knowledge inventory and specify the new areas that are critical to your project. The focus can be further enhanced by specifying key hypotheses regarding possible findings, information categories relevant to the issue, and other criteria such as product categories, consumer targets, market areas, time frames, and so on.
- *Search across multiple sources.* Don't expect to find what you need in single pieces of data or sources of information. You only rarely will find what you need in one place.
- *Integrate information from the multiple sources.* Use techniques of trend analysis to identify patterns that emerge when various information elements are combined; for example, content analysis, stakeholder analysis, paradigm shift, trendlines, critical path analysis, sector analysis (technological, social/cultural, occupational, political, economic), or other analytic techniques that facilitate integration of diverse data and information and identification of underlying patterns.
- *Search for databases that contain analyses rather than limiting the search to just data or information.* Many of the professional online database producers and vendors offer thousands of full-text articles and resources that contain analyses. You may be able to find material that already provides some interpretation that may be helpful.
- *Enhance the robustness of your data or information through multiple-source validation.* You can increase confidence in the validity of the findings of your secondary searches by looking for redundant patterns that cut across different sources and studies.

SYNDICATED SERVICES

Some of the aforementioned commercial services are examples of what are called syndicated services. Research organizations providing such services collect and tabulate specialized types of marketing information on a continuing basis for purposes of sale to a large number of firms. In general, syndicated data are made available to all who wish to subscribe. Reports are made available on a regular basis (for example, weekly, monthly, quarterly). Since these data are not collected for a particular firm for use in a specific research problem situation, they can properly be viewed as secondary data. Syndicated services are widely used in such areas as movement of consumer products through retail outlets, direct measures of consumer purchases, social trends and lifestyles, and media readership, viewing, and listening.

The syndicated Survey of American Consumers (based on surveying more than 25,000 adults) by Mediamark Research, Inc. provides an illustration of syndicated services. This survey provides data useful for detecting a marketer's best prospects by providing answers to such questions as the following:

- How many customers are there for the products or services we market? Is the size of the market growing? Stabilizing? Or shrinking?
- Who are the customers? How old are they? What do they earn? Where do they live?
- How do customers differ in terms of how often and how much they buy? Who are the heaviest purchasers of the product?
- What brands are customers buying? How have shares of the market changed? What differences are there among brand buyers?
- What's the best way of reaching prospects? Which media vehicles and formats are most efficient in delivering the message to the customer?

Mediamark is able to profile American consumers on the basis of more than 60 demographic characteristics and covers usage of some 500 product categories and services and 6,000 brands.

Types of Syndicated Services

Syndicated data may be obtained by personal interviews, direct observation, self-reporting and observation, or use of certain types of mechanical reporting or measuring devices. One of the most widely used approaches is the continuous panel, which refers to a sample of individuals, households, or firms from whom information is obtained at successive time periods. Continuous panels are commonly used for the following purposes:

1. As consumer purchase panels, which record purchases in a consumer diary and submit them periodically.
2. As advertising audience panels, which record programs viewed, programs listened to, and publications read.
3. As dealer panels, which are used to provide information on levels of inventory, sales, and prices.

Such panels have been established by many different organizations, including the federal government, various universities, newspapers, manufacturers, and marketing research firms. These types of panels furnish information on at regular intervals on continuing purchases of the products covered.

> For example, typical consumer panels might report the type of product purchased by brand, weight or quantity of unit, number of units, the kind of package or container, price per unit, whether a special promotion was in effect, store name, and date and day of week of purchase. Data are recorded in diaries, either online or are mailed in each month.

One of the largest consumer panels is maintained by NPD Research (http://www.npdor.com). This panel comprises 13,000 families and is national in coverage. NPD also maintains self-contained panels in 29 local markets. Other well-known national consumer panels are maintained by Synovate (synovate.com), NFO (mysurveys.com) ACNielsen (http://en-us.nielsen.com), and IRI (infores.com).

Advertising audience panels are undoubtedly more widely publicized than other panels. It is from these panels that television and radio program ratings are derived. These panels are operated by independent research agencies rather than the media—both for reasons of economy and to avoid any question of partisanship.

> For example, ACNielsen uses a metering device that provides information on what TV shows are being watched, how many households are watching, and which family members are watching. The type of activity is recorded automatically; household members merely have to indicate their presence by pressing a button. The sample is about 5,000 households. In local markets, the sample may be 300 to 400 households.

Single source data tracks TV viewing and product purchase information from the same household. Mediamark's national survey and IRI's BehaviorScan are examples of such single-source data. The single-source concept was developed for manufacturers who wanted comprehensive information about brand sales and share, retail prices, consumer and trade promotion activity, TV viewing, and household purchases.

The information obtained from the types of syndicated services described previously has many applications. The changes in level of sales to consumers may be analyzed directly without the problem of determining changes in inventory levels in the distribution channel. Trends and shifts in market composition may be analyzed both by type of consumer and by geographic areas. A continuing analysis of brand position may be made for all brands of the product class. Analyses of trends of sales by package or container types may be made. The relative importance of types of retail outlets may be determined. Trends in competitor pricing and special promotions and their effects can be analyzed along with the effects of the manufacturer's own price and promotional changes. Heavy purchasers may be identified and their associated characteristics determined. Similarly, innovative buyers may be identified for new products and an analysis of their characteristics made to aid in the prediction of the growth of sales. Brand-switching and brand-loyalty studies may be made on a continuing basis. One reported use of this syndicated service has been to design products for specific segments.

The products from syndicated services are continually changing with client needs and new technological opportunities.

SUMMARY

This chapter has been concerned with secondary information and sources of such information. We started with some reasons why secondary information is essential to most marketing research projects. Then, various sources and types of secondary information—internal and external—were discussed in some depth. Also given more than cursory treatment was syndicated data, a major type of service provided by commercial agencies.

REFERENCES

ACNielsen: http://en-us.nielsen.com

ACNielsen – Claritas Demographic Widget, http://www.claritas.com/target-marketing/nielsen-claritas-demographic-widget.jsp

Albaum, G., Duerr, E., & Strandskov, J. (2005). *International marketing and export management* (5th ed.). Harlow, UK: Pearson Education.

Andreasen, A. (1988). *Cheap but good marketing research.* Homewood, IL: Dow-Jones Irwin.

Berry, M., & Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support.* New York: Wiley.

Berry, M., & Linoff, G. (2002). *Mining the Web: Transforming customer data.* New York: Wiley.

Campbells 1959 attribute commercial http://www.youtube.com/watch?v=qMJtLP8jMWQ

Campbells 1985 Soup is good food commercial http://www.youtube.com/watch?v=4PtZYlcyXfI

Dehmater, R., & Hancock, M. (2001). *Data mining explained: A manager's guide to customer centric business intelligence.* Burlington, MA: Digital Press

DIALOG http://www.dialog.com

Direct Selling Association, 2003 http://www.dsa.org/pubs/

Dow Jones News/Retrieval http://www.dowjones.com/Products_Services/ElectronicPublishing/EnterpriseMedia.htm

Elkind, F., & Kassel, A. (1995, July 31). A marketer's guide for navigating the information superhighway. *Marketing News*, *29*, 2ff.

Garver, M. S. (2002, September 16). Try new data mining techniques. *Marketing News, 36,* 31–33.

IRI: infores.com

LexisNexis http://www.lexisnexis.com

Morgan, M. S. (2001, October 8). Research boosts database's power. *Marketing News, 35,* 16.

NFO: mysurveys.com

NPD Research http://www.npd.com

NPD Research, Inc. http://www.npdor.com

RoperASW. (2002). Available at RoperASW.com. Accessed January 9, 2002.

SRDS Life Style Market Analyst, http://www.srds.com/frontMatter/ips/lifestyle/index.html

SRDS Graphical Reports, http://www.srds.com/frontMatter/ips/lifestyle/images/LMAA_overview-with-map.jpg

Synovate: synovate.com

U.S. Census Bureau, Statistical Abstract, http://www.census.gov/compendia/statab/

U.S. Census of Population, http://www.census.gov/population/www/

U.S. Census of Housing, http://www.census.gov/hhes/www/housing.html

U.S. Current Population Reports, http://www.census.gov/main/www/cprs.html

U.S. Current Construction and Housing Reports, http://www.census.gov/prod/www/abs/cons-hou.html

APPENDIX 3.1
Selected Secondary Sources

Marketing/Advertising
Directories

American Marketing Association. *Green Book: International Directory of Marketing Research Houses and Services.* Annual. www.greenbook.org

The Green Book lists and describes services offered by major marketing firms. Foreign as well as U.S. research firms are entered alphabetically in a single list.

Broadcasting/Cablecasting Yearbook. Annual. http://www.bowker.com/index.php/component/content/article/2/33
In addition to its radio section, this source includes sections that survey the broadcasting industry, profile television and cable stations, and focus on programming, professional services, technology, advertising and marketing, and satellites.

The Direct Marketing Market Place. Annual. http://www.dirmktgplace.com/
Of possible interest to direct marketers, this directory lists the following: direct marketers of products and services (catalog and retail sales, financial and investment services, fund raising, etc.); service firms and suppliers to direct marketing companies (printers, list brokers and managers, computer services, etc.); creative and consulting services (agencies, artists, copywriters, etc.); and lists of organizations, courses, awards, and so on. For each it includes a description of the firm's product or service, names of key executives, sometimes number of employees, gross sales or billings, advertising budget and expenditures spent in direct marketing media, and countries in which business is conducted.

Rand McNally. *Commercial Atlas and Marketing Guide.* Annual. http://www.randmcnally.com/
Brings together current economic, demographic, and geographic information. In addition to a series of area maps highlighting major military installations, trading areas, retail sales, and manufacturing centers, it includes statistics on business and manufacturers, retail trade, sales, and the largest corporations in the United States.

Standard Directory of Advertising Agencies. Annual. http://www.redbooks.com/Nonsub/index.asp
Also known as the Agency Red Book, this directory contains current information about 4,500 American advertising agencies and their branches, including their names and addresses, specialization, major accounts, and key staff members.

Standard Directory of Advertisers. Annual.
A companion directory to the aforementioned Agency Red Book. It focuses on companies and other organizations that advertise and includes trade names, the types of media used for advertising, the advertising agencies employed, and, frequently, annual advertising budgets.

Indexes

ABI Inform CD Rom.. http://www.proquest.com/en-US/catalogs/databases/detail/abi_inform.shtml
Among the fields covered are marketing and advertising. Over 800 journals are indexed and each article is fully annotated. A full list can be printed from the CD-ROM. Current five years on disc

Business Index. Monthly.
This is a microfilm index that is a cumulative, three-year index to over 800 business periodicals and newspapers.

Business Periodicals Index. Monthly. http://www.ovid.com/site/catalog/DataBase/173.jsp
Indexes over 300 business periodicals. Most are general business periodicals; trade journals are not well represented.

Gale Group F & S Index United States. Weekly.
http://www.gale.cengage.com/servlet/BrowseSeriesServlet?region=9&imprint=000&cf=ps&titleCode=FSUS
Excellent for the latest news on product and industry developments or for information on specific companies, particularly those that are not well known. Indexes over 750 periodicals, reports, and trade journals.

Public Affairs Information Source Bulletin. Semimonthly.
The index covers periodical articles, books, government documents, conference proceeds, and other publications that touch on topics of interest to public policy. Over 1,400 journals are indexed. Also on CD-ROM.

Consumer Index. Monthly.
Topicator. Monthly.

Other Sources

New York Times Index. http://www.nytimes.com/ref/membercenter/nytarchive.html

Statistical Indexes
1. *American Statistics Index.* Monthly. http://academic.lexisnexis.com/cis/subject-area.aspx?pid=65&type=&parentid=58
The ASI is a master guide to statistical publications of the federal government. It is published in two parts, an Index Section and an Abstract Section. It is arranged by subject with references to microfiche reproduction of the publication found there.

2. *Statistical Reference Index.* Monthly.
Same format as aforementioned ASI except it indexes and abstracts statistics contained in publications not issued by the federal government. It includes statistics published by private organizations, corporations, commercial publishers, and independent and university-affiliated research organizations.

Government Publications
1. *Monthly Catalog of United States Government Publications.*
Records and indexes documents received by the Government Printing Office from all areas of government, legislative, judicial, and executive branches, and independent and regulatory agencies as well.

2. *Congressional Information Service.* CIS/Index. Monthly.
This is a comprehensive subject index and abstract to the working papers of Congress comprising committee hearings, reports, and prints, as well as publications of joint committees and subcommittees, executive documents, and special publications.

Bibliographies
1. Special Librarian Association. *What's New in Advertising and Marketing.* Monthly.
List by subject new books, periodicals, and trade publications in marketing and advertising. It includes many titles that are free, especially summaries of market research.

Special Marketing, Advertising, and Consumer Reference Sources

Editor and Publisher Market Guide. Annual. http://www.editorandpublisher.com/eandp/resources/market_guide.jsp
Market data for over 1,600 U.S. and Canadian newspaper cities covering facts and figures about location, transportation, population, households, banks, autos, gas and electric meters, disposable income, income per capita and per household, and many others.

Leading National Advertiser. *Ad. $ Summary.* Annual.
Lists brands alphabetically and shows total 9-media expenditures, media used, and parent company. Also included in this report are industry class totals and rankings of the 1,000 leading companies by total 9-media spending.

Sales and Marketing Management. Publishes an annual statistical issue and a loose-leaf data source.

1. *Survey of Buying Power.* Annual http://www.surveyofbuyingpower.com/sbponline/index.jsp
Contains geographically oriented demographic, income, and retail sales statistics for metropolitan areas and counties. Also lists areas by population, number of households, total retail sales, and much more.

2. *Survey of Buying Power Data Service.* Annual.
Has more detailed and additional information than the *Survey of Buying Power.*

Simmons Study of Media and Markets. http://www.libs.uga.edu/ref/simmons/index.html
Forty volumes devoted to the determination of use of various products and services.

Standard Rate and Data Service. http://www.srds.com/portal/main?action=LinkHit&frameset=yes&link=ips
This service offers separate directories giving advertising rates, specifications, and circulation for publications, broadcast stations, and so forth in the following media: business publications (monthly), community publications (semiannual), consumer magazines and agri-media (monthly), newspapers (monthly), print media production (quarterly), spot radio (monthly), spot television (monthly), and weekly newspapers (monthly). This service also publishes an annual *Newspaper Circulation Analysis* covering newspaper circulation and metro area, TV market, and county penetration. Entries include daily and Sunday circulation figures, by county, for each newspaper.

Company and Industry Data

Dun and Bradstreet. *Million Dollar Directory.* Annual.
Lists over 160,000 American companies having an indicated net worth of $500,000 or more.

Standard and Poor's Register of Corporations, Directors and Executives. Annual.
Lists over 45,000 U.S. businesses.

Directory of Corporate Affiliations. (Annual) and *America's Corporate Families: The Billion Dollar Directory* (Annual).
Each directory features lists of divisions and subsidiaries of parent companies.

MOODY's Investor Service. Various call numbers. Seven annual volumes plus twice-a-week supplements.
Includes Bank and Finance, Industrial, Municipal and Government, Over-the-Counter, Public Utility, Transportation, and International.

Standard and Poor's Corporation Records.

Value Line Investment Survey. Weekly.

Walker's Manual of Western Corporations and Services. Annual.

Standard and Poor's Corporation. *Industry Surveys.*
This loose-leaf service includes for 22 industry categories detailed analysis of each category and of the industries that comprise it.

Predicasts Forecasts. Quarterly.
Is useful both as a source of forecasts and projections of industries, products, and services and as a finding aid to other sources.

U.S. Industrial Outlook. Annual.
Contains short-term forecasts for specific industries.

International Business

Many sources listed previously also are relevant to foreign markets and international marketing activities.

Indexes and Guides

Predicasts Funk and Scott Indexes.

1. *F & S Europe.* Monthly.
Covers articles or data in articles on foreign companies, products and industries.

2. *F & S International Index.* Monthly.
Covers articles or data in articles on foreign companies, products, and industries.

Economic Trends

Asia Yearbook. Annual.
Contains information on individual countries plus finance, investment, economics, trade, and aid, and so on.
Europa Yearbook. Annual.
Includes history, economic affairs, economic statistics, constitution, government, and so on of individual countries.

Foreign Economic Trends and Their Implications for the United States. Semiannual.
Series of brief reports on more than a hundred countries that are prepared by U.S. embassies and consulates and provide current data on gross national product, foreign trade, unemployment figures, wage and price index, and so on.

Organization for Economic Cooperation and Development Economic Surveys of the O.E.C.D. Annual.
Economic surveys of each of the 24 OECD member countries that contain information on current trends in demand and output, price and wages, foreign trade and payments, economic policies, and future economic prospects.

Overseas Business Reports. Irregular.
Provides current and detailed marketing information for businesses evaluating the export market. Includes trade outlooks, statistics, advertising and market research, distribution and sales channels, regulations, and market profiles.

Price Waterhouse Guide Series. Continual update.
A series of guides on various aspects of doing business where Price Waterhouse has offices or has business contacts. Topics covered include investments, corporate information, business regulations, accounting, taxes, and so on. Over 75 countries are represented in this series.

World Economic Survey. Annual.
A comprehensive picture of the economic situation and prospects for the world as a whole and for major world regions. Analysis of inflation, rates of interest, exchange rates, trade balances, commodity prices, and indebtedness are included.

Directories

Bradford's Directory of Marketing Research Agencies and Management Consultants in the United States and the World.

Directory of American Firms Operating in Foreign Countries.
Gives alphabetical index of U.S. Corporations and international geographic distribution; alphabetized by country, then firm.

Directory of Foreign Firms Operating in the U.S.
Has:
1. Alphabetical listing by country—shows American Company and parent company.
2. Alphabetical listing of foreign parent companies and corresponding American subsidiaries or affiliates.
3. Alphabetical listing of American subsidiaries, branches, or affiliates of foreign companies.

Dun and Bradstreet. *Principal International Businesses.* Annual.
Complete legal name, parent company, address, cable or telex, sales volume, number of employees, SIC number, description of activities, and chief executives are given for over 50,000 businesses.

Foreign Commerce Handbook. Irregular.
Information on all phases of international business is included in this reference source, including foreign-trade services, daily language of foreign commerce, bibliography, and lists of organizations.

International Directory of Corporate Affiliations. Annual.

A comprehensive reference source of foreign companies with U.S. holdings and U.S. companies with foreign holdings.

Worldwide Chamber of Commerce Directory. Annual.
Listing of U.S. Chambers of Commerce, American Chambers of Commerce abroad and chief executive officer, and Foreign Chambers of Commerce in principal cities of the world.

Special Services

African Research Bulletin: Economic, Financial and Technical Series. Monthly.

Asian Recorder. Weekly.

Business International.
1. Business International *Loose-Leaf Services.* Weekly.
This is a newsletter service that offers short articles on capital sources, economy, industry, exporting, foreign trade, management, and marketing. These newsletters cover the following areas: Asia, China, Eastern Europe, Europe, International, and Latin America.

2. *Financing Foreign Operations.* Irregular.
Current guide to sources of capital and credit in 34 major markets.

3. *Investing, Licensing and Trading Conditions Abroad.* Annual.
Covers African-Middle East, Europe, Asia, North America, and Latin America.
Includes information on state role in industry, rules of competition, price controls, corporate taxes, personal taxes, incentives, labor, and foreign trade.

4. *Research Reports.* Irregular.
These are in-depth reports prepared by the B.I. service on various subjects and countries. "Marketing in China," "Andean Common Market," and so on.

5. *Worldwide Economic Indicators.* Annual
Includes key economic indicators for over 130 countries. Includes G.D.P., demographic and labor force data, foreign trade, and production and consumption data.

Commerce Clearing House. *Common Market Reporter.* Weekly.

Ernst and Whinney. *International Business* Series. Irregular.
Very much like Price Waterhouse Guides described previously but not as extensive in coverage.

Federation of International Trade Associations. *Really Useful (Web) Sites for International Trade Professionals.* Bi-weekly (http://fita.org, newsletter@fita.org)

Moody's International Manual. Twice weekly.
Provides financial and business information on more than 5,000 major foreign corporations and national and transnational institutions in 100 countries.

Statistics

Demographic Yearbook. Annual.
Comprehensive collection of international demographic statistics. Includes population, demographic and social characteristics, geographical, educational, and economic information.

U.S. Statistical Office. *Statistical Yearbook.* Annual.
Kept up to date by the *Monthly Bulletin of Statistics.* Comprehensive compendium of international comparable data for the analysis of socioeconomic development at the world, regional, and national levels.

International Labor Office. *Yearbook of Labor Statistics.*

Includes total and economically active population, employment, unemployment, hours of work, wages, and so on.

U.N. Department of Economic and Social Affairs. *International Trade Statistics Yearbook.* Annual.
Kept up to date by Commodity Trade Statistics and Direction of Trade.
U.N. Statistical Office. *National Accounts Statistics.* Annual.

UNESCO. *Statistical Yearbook.* Annual.
Contains tables grouped according to various subjects: population, education, libraries and museums, book production, newspapers and broadcasting, television, and cultural expenditure. Over 200 countries or territories represented.

U.N. Statistical Office. *Yearbook of Industrial Statistics.* Annual.
Covers general industrial statistics data for each country and commodity production data. The United Nations has separate Economic and Social Commissions for each geographical region. Each publishes an annual report that surveys the economic and social trends of that area, including economic prospects, foreign trade, investments, oil industry, and agricultural. The titles of these annual reports are listed next, followed by the title of the periodical and supplements that update them if one exists.

1. *International Financial Statistics.* Monthly.
Data on exchange rates, international reserves, money and banking, trade, prices, and production for all IMF member countries.

2. International Monetary Fund. *Balance of Payments Yearbook.* Annual.
Five-year detailed balance of payments statistics for about 100 countries. Includes statistics for goods, services, capital, SDRs, and so on. Pay special attention to notes that accompany each table.

This appendix was developed from information provided by Rodney Christensen, former
reference librarian, Knight Library, University of Oregon.

This page is intentionally blank

Chapter 4
CONDUCTING INTERVIEWS

THE INTERVIEW

The research interview is a means of obtaining information from respondents. We can define an interview as a form of dyadic (person-to-person) communication that involves the asking and answering of questions.

The dyadic nature of the interview is such that bias and error potentially will be present as a result of interviewer and the interviewee background characteristics (e.g., age, education, socioeconomic status, race, religion, gender, etc.), psychological attributes (e.g., perceptions, attitudes, expectations, motives), and behavior (e.g., errors in asking questions, probing, motivating, recording responses). Interviewer and respondent are perceive and react to the observable background characteristics and specific behaviors of each other. These factors have a direct influence when an interactive interview is conducted and are implied in self-report interviews (e.g., mail, e-mail).

Intervening variables are also part of the didactic interaction. The respondent's role expectations, the interviewer's task behavior, differences in social desirability of response alternatives, extent of topic threat, and salience of topic all affect the responses received.

The interviewer can likewise control the style of interviewing and thereby affect the quality of information obtained. For example, styles can range from socio-emotional (maintaining a warm, sympathetic, and understanding relationship with the respondent) to formal (where the person-oriented actions of the interviewer are held to a socially acceptable minimum). The researcher must assess the desirability of one style over the other in a given situation.

EXHIBIT 4.1 What's in an Interview?

Universal dimensions underlie the relationships that are shaped as part of every Interview.

- *Involvement* encompasses the degree to which each party wants to take part in the interview, including the degree of commitment of each to making it a success.
- *Control* refers to the degree of power the interviewer or interviewee has to affect the interview process and its outcome.
- *Relationship* is the degree of warmth or friendship between the interview parties.

A number of elements also define the environment in which each Interview takes place:

1. **Context.** The total situation in which an interview takes place, including location, physical arrangements, the people present, and those absent. This also includes status differences between parties, temperature, privacy, and time.
2. **Content.** What the parties talk about during the interview. It involves topic selection and treatment, arguments, supporting materials, language, and questions and answers. *Information From Respondents*

3. **Structure.** Includes the interviewer's or interviewee's basic organizational patterns, sequences of topics and questions, and the means used to open and close interviews.

4. **Disclosure.** The willingness on the part of both parties to reveal their "true" selves to one another.
5. **Feedback.** The continuous stream of verbal and nonverbal signals (e.g., smiles, puzzled expressions, raised eyebrows, moans) sent between interview parties that reveal feelings, belief or disbelief, approval or disapproval, understanding or misunderstanding, interest or disinterest, and awareness or
6. **Cooperation.** The degree to which the interview parties are willing and able to reduce the competition inherent in most interview situations and work together for their mutual benefit.
7. **Conflict.** The potential or actual struggle between parties because of incompatible or opposing needs, desires, demands, and perceptions.
8. **Trust.** Belief in the good, worth, ethics, believability, and reliability of the other party.

Involvement, Control and Relationships have some effect upon each of the elements. These dimensions and elements of relationships are present in each interview but are not of equal importance. Although they are independent of each other, they have strong interdependence as well.

SOURCE: From Stewart, C. J. and Cash, W. B., *Interviewing Principles and Practices, 4/e*. © 1985 William C. Brown, Publishers. Reprinted with permission of The McGraw-Hill Companies, pp. 9–13

Structure of the Interview

Interviews in marketing research and the behavioral sciences typically involve information gathering and are usually classified by two major characteristics. An interview is either **structured or unstructured**, depending on whether a formal questionnaire has been formulated and the questions asked in a prearranged order. An interview is also categorized as either **direct or indirect**, reflecting whether the purposes of the questions are intentionally disguised. Cross-classifying these two characteristics helps us to identify four different types of interviews

Objective Interviews	(a) structured and direct
	(b) unstructured and direct
Subjective Interviews	(c) structured and indirect
	(d) unstructured and indirect.

Types a and b are basically objectivist; types c and d, subjectivist. We discuss each type of interview in turn (although the discussion of the two indirect types of interviews is combined). We then discuss the media through which interviews may be conducted.

Structured-Direct Interviews

Structured-direct interviews are the usual type of consumer survey to "Get the Facts" and obtain descriptive information. A formal questionnaire is used consisting of nondisguised questions.

Example: A marketing research manager of a bedroom-furniture manufacturer wants to find out how many and what kinds of people prefer various styles of headboards and dressers. The question sequence is fixed only those questions are asked. The resulting interview is structured-direct in nature.

The next portion of this questionnaire is designed to obtain information about furniture styles that you own, bedroom furniture design preferences, and socioeconomic characteristics.

Which of the styles of furniture shown in these pictures is most nearly similar to your furniture? (Show folder with furniture pictures.)

◯ Country	◯ Modern
◯ Spanish	◯ Cape Cod
◯ Mediterranean	◯ Student Budget
◯ Traditional	◯ Shabby Chic

Which of the styles of bedroom sets shown in these pictures do you like best? (Show folder with bedroom set pictures.)

◯ Log Cabin	◯ Traditional
◯ Spanish	◯ Contemporary
◯ Modern	◯ Scandinavian Modern

Please indicate your occupation.

About how much was your total household income (for you and your spouse) last year from salary and other sources?

- ◯ $0 - $25,000
- ◯ $25,001 - $50,000
- ◯ $50,001 - $75,000
- ◯ $75,001 - $100,000
- ◯ $100,001 - $125,000
- ◯ $125,001 - $150,000
- ◯ $150,001 - $175,000
- ◯ $175,001 - $200,000
- ◯ $200,001+

The structured-direct interview has many desirable features. Since the questions are formulated in advance, all the required information can be obtained in an orderly and systematic fashion. The exact wording and phrasing of the questions can be worked out carefully to reduce the likelihood of misunderstandings or influencing the answer. Pretests can (and should) be made on the questionnaire to discover any problems in the wording or ordering of questions before the questionnaire is finalized.

In the structured-direct interview, the questionnaire is, in effect, the dominant factor in the interview. The interviewer's role is simply to ask questions. The same questions are asked of all respondents in the same order. This provides maximum control of the interviewing process and reduces the variability in results caused by differences in interviewer characteristics. This type of interview is less demanding insofar as the abilities of the interviewer are concerned, permitting the use of less-skilled interviewers and resulting in a lower cost per interview. The standardized, direct questions allow for uniform recording of answers, thereby reducing errors in editing, tabulating, and analysis of the information.

The major problems associated with this type of interview involve wording questions properly and the difficulties encountered in getting unbiased and complete answers to questions concerning personal and motivational factors. The structured-direct interview is by far the most commonly used type of interview in marketing research. An alternative approach is suggested in Exhibit 4.2.

EXHIBIT 4.2 Structuring Conversational Interviewing

Despite pretesting, every survey question contains terms that have the potential to be understood differently than the survey designer intends. For example, in a study of physicians conducted for a pharmaceutical company, the interviewer might ask, “During the past two weeks, did you prescribe any anti-inflammatory drug?” A respondent might answer, “Well, that depends. What exactly do you mean by anti-inflammatory?” The interviewer is now faced with a choice. Should he use his knowledge to answer the respondent’s question, or leave the interpretation of “anti-inflammatory” up to the respondent?

The normal (standardization approach) way of handling this situation would be to leave the interpretation of the question up to the respondent. Interviewers must read exactly the same question and never interpret the question in any way (Fowler, 1991; Fowler & Mangione, 1990). When a respondent asks for help, the interviewer should use so-called neutral probing techniques, like repeating the question, presenting response alternatives, and so forth.

Another school of thought holds that the interviewer in the example above should help the respondent and define “anti-inflammatory.” This group argues that response validity can be undermined if respondents interpret questions idiosyncratically. An approach suggested is that interviewers be allowed to use conversationally flexible interviewing techniques. This means that interviewers should engage respondents in a manner similar to ordinary conversation, deviating from the standardized script to ensure that respondents interpret questions consistently and correctly.

Online surveys have the ability to provide standard definitions through context sensitive help. The respondent uses the mouse cursor to touch a context laden word (underlined or otherwise designated) and a popup definition will appear. To the degree that respondents needing help would receive it, response validity would increase, but the interaction with the respondent would be standardized.

Unstructured-Direct Interviews

Unstructured-direct interviews are most often used in exploratory studies, and in qualitative research (see Chapter 5). In the unstructured-direct method of interviewing, the interviewer is given only general instructions to obtain the type of information desired. He or she is left free to ask the necessary direct questions to obtain this information, using the wording and order that seems most appropriate in the context of each interview.

Many research projects go through an exploratory phase in which researchers contact respondents and hold unstructured interviews. These interviews are useful for obtaining a clearer understanding of the problem, and determining what areas to investigate. This type of interview is also often useful for obtaining information on motives. Following the exploratory interviews, a formal questionnaire is developed for the final interviews.

To use the bedroom furniture example again, if the owner of a bedroom set is asked the free-answer question, “Why did you buy your bedroom set?” the answer is almost certain to be incomplete, provide proximate causes and may be worthless.

If the interviewer were seeking motivations, consider answers such as “Because we needed a bed,” “Our old bed was worn out,” or “Because it was on sale.” When motivations are given, such as “We enjoy a comfortable mattress that gives us a good night’s sleep,” they are rarely complete.

The added enjoyment may be because the mattress is firmer, because of the pillow top, because of the prestige the owner attaches to having a carved oak bedroom set, or some combination of these and other factors. In addition, it is probable that motives other than “enjoyment” influenced the purchase.

When used to establish motives, the unstructured-direct interview is known as a ***depth interview***. The interviewer will continue to ask probing questions: “What did you mean by that statement?” “Why do you feel this way?” “What other reasons do you have?” The interviewer continues with similar questions until satisfied that all the information that can be obtained has been obtained, considering time limitations, problem requirements, and the willingness and ability of the respondents to verbalize motives.

The unstructured interview is free of the restrictions imposed by a formal list of questions. The interview may be conducted in a seemingly casual, informal manner in which the flow of the conversation determines which questions are asked and the order in which they are raised. The level of vocabulary used can be adapted to that of the respondent to ensure that questions are fully understood and rapport is developed and maintained. The flexibility inherent in this type of interview, when coupled with the greater informality that results when it is skillfully used, often results in the disclosure of information that would not be obtained in a structured-direct interview.

In the unstructured interview, the interviewer must both formulate and ask questions. The unstructured interview can therefore be only as effective in obtaining complete, objective, and unbiased information as the interviewer is skilled in formulating and asking questions. Accordingly, the major problem in unstructured direct interviews is ensuring that competent interviewers are used. Higher per-interview costs result, both as a result of this requirement and the fact that unstructured interviews generally are longer than those that use a questionnaire. In addition, editing and tabulating problems are more complicated as a result of the varied order of asking questions and recording answers.

Structured-Indirect and Unstructured-Indirect Interviews

A number of techniques have been devised to obtain information from respondents by *indirect* means. Both structured and unstructured approaches can be used. Many of these techniques employ the principle of *projection*, in which a respondent is given a non-personal, ambiguous situation and asked to describe it. It is assumed that the respondent will tend to interpret the situation in terms of his or her own needs, motives, and values. The description, therefore, involves a projection of personality characteristics to the situation described. These techniques are discussed in more depth in Chapter 5.

REDUCING RESPONSE AND NONRESPONSE BIAS

A major concern of the research planner when choosing which interview medium to use is the potential systematic error (i.e., bias) that might arise. In Chapter 2, we discussed total error and looked at its major components. At this point, it is useful to explore how to reduce the non-sampling based error that occurs during the interview process. In communication, error may be due to the nature of response given (**inaccuracy** and **ambiguity**) or due to the fact that the sample member has not responded. The following sections will discuss these topics.

Inaccuracy

Inaccuracy refers to intentional and unintentional errors made by the respondent when they provide information. There are two types of inaccuracies: predictive and concurrent. Predictive inaccuracy is a special case or response error caused by inaccurate intentions:

> Suppose a respondent indicates that he intends to buy a new sports car within 6 months; and then does not. Or, alternatively, he does not now intend to buy, he answers "No" to the question, and then buys a car within the six-month period. In each case, the respondent intention was clear, but was not followed. This situation is a predictive inaccuracy.

A similar type of predictive inaccuracy can occur when marketing researchers try to predict actual market response to a price by asking consumers, "How much are you willing to pay for Product X?" Differences between predicted and actual price acceptability may occur because the true range of acceptable prices may change between the time of data collection and the point of purchase for a number of reasons, such as budget constraints or windfalls, the price of substitutes at point of purchase, search costs, and purchase urgency.

Concurrent inaccuracy occurs when the respondent intentionally does not provide accurate information. Everyday experiences and empirical evidence suggest that inaccurate information results from the respondent's **inability** or **unwillingness** to provide the desired information.

> For our car purchase example, suppose the respondent answers "Yes", but really has no intention of buying a sports car within this period or, conversely, answers "No", but does intend to buy a sports car. In this case, we may say that there is concurrent inaccuracy in his statement. The intention is to provide inaccurate information.

Concurrent inaccuracies are a major concern for many kinds of information obtained from respondents (information on past behavior, socioeconomic characteristics, level of knowledge, and opinion-attitude). Concurrent inaccuracies may also apply to instances where observation is used; the observer charged with reporting the events or behavior may be unable or unwilling to provide the desired information.

It is clear from this brief introduction to predictive and concurrent inaccuracies that inability and unwillingness to respond are major contributors to response bias and warrant more detailed attention to understand how they can be controlled.

Inability to Respond

Even such a simple and straightforward question as “What is the model year of your family car?” may result in an information-formulation problem, particularly if the car is several years old. If respondents were asked, “What brand or brands of tires do you now have on your car?” most would have even more difficulty in providing an accurate answer without looking at the tires. Finally, if respondents were asked, “What reasons did you have for buying Brand “A” tires instead of some other brand?” most respondents would have even more difficulty in providing an accurate answer. Semon (2000a, 2000b) suggests that inaccuracies due to inability to respond stem from three major conditions:

- Memory error: A respondent gives the wrong factual information because he or she simply does not remember the details of a specific event. Often time since an event occurred (purchase) are underestimated or overestimated. While better questionnaire and survey design can help reduce this error, such proven techniques are not used because they add to the length of the survey. For instance, in a personal or telephone interview survey, follow-up calls are often not made to confirm the answers given.

- Ignorance error: This refers to the respondent’s lack of understanding, awareness or perception of irrelevance for a question and is due to a poor research design in terms of question content and sampling. A question (or even an entire questionnaire) may be unrealistic, deficient, or directed to the wrong persons.

- Misunderstanding: This can be a matter of careless question design. Poorly defined terms or words with different meanings can lead to inaccurate, or even deliberately falsified responses. Proper question design would avoid words with multiple meanings and definitions, or should clearly define the context in which the word is being used in the questionnaire.

In addition, the following items also create inaccuracies through the inability to respond accurately:

Telescoping

Questions that ask respondents to reconstruct past experiences run a high risk of response bias. More specific is the possibility of a respondent *telescoping*, or misremembering when an event occurred during a short recent time period. In a study of durable-goods purchases in the United States, respondents on average displayed forward-telescoping biases (reporting something happened more recently than it did), and the magnitude of this reporting increased (Morwitz, 1997). Overall, the tendency to make forward-telescoping errors may differ by the demographic of the respondent and by the event being studied— recall of purchase, reading of an ad, or some other event will have an effect on the nature of any telescoping errors.

Telescoping can be reduced by using bounded recall procedures, which involve asking questions about the events of concern in previous time periods as well as the time period of research interest (Sudman, Finn, & Lannam, 1984). Other approaches include asking respondents to use finer time intervals, and to use a landmark event such as New Year’s Day or Easter, or an individual event landmark such as the date of a child’s wedding.

Exhibit 4.3 Response Error and Questionnaire Design

One of the keys to minimizing concurrent errors is for researchers to better select questions that fulfill the clients' information needs.

Important/Urgent questions are a first priority and easy to identify. Unimportant/Non-urgent questions are likewise easy to identify and exclude. It is the other two categories that can cause most of the problems.

Remember that Urgent/Unimportant questions may be best answered by a judgment-call than by extensive research, and that the Important/Non-urgent questions are the ones that often need to be addressed. In the short run, a company will carry on without answers to Important/Non-urgent questions. But these answers may be essential to the long-term future direction of the company.

Unwillingness to Respond

When we move to the problem of unwillingness of respondents to provide accurate information, the topic is more complex. Here we are dealing with the motivations of people: why they are not willing to accurately provide the information desired.

Except in those instances where the respondent provides information by being observed in a natural situation, there are always costs (negative utilities) attached to his or her formulating and sharing information.

No fully accepted general theory to explain this behavior, but we again apply everyday experiences to this problem and add some research findings to suggest why people may not be willing to make accurate information accessible.

Investigator Expectations

A complex source of inaccuracy in response stems from the respondents' appraisal of the investigator and the opinions and expectations imputed to him or her.

> A classic example is a cosmetics study that showed an unexpectedly high reported usage of luxury cosmetics among women from low-income families. In this case, one exceptionally well-dressed, carefully groomed, competent interviewer conducted all of the interviews. The study was repeated with a matronly woman, in dress similar to the women interviewed, calling on the same respondents on the following days. The reported brands of cosmetics used were much less expensive, in this series of interviews.

Investigator Unwillingness

Sometimes, what appears to be respondent unwillingness to provide accurate data is actually a case of interviewer cheating where the investigator is unwilling to obtain accurate information. This happens when an interviewer finds a particular question too embarrassing to ask; when the interview finds it easier to self complete the survey forms rather than conduct the interviews; or when interviewers have friends complete the survey. Interviewers may also complete some questions legitimately and then make an estimate or inference of other questions such as age, income, and certain attitudes or behaviors of respondents.

Interviewer cheating can be kept to a low level of incidence but not eliminated completely. Careful selection, training, and supervision of interviewers will eliminate much of the problem. In addition, control procedures can and should be established to reduce it even more.

The simplest control procedure is to *call-back* a subsample of respondents. If the information on an initial interview is found to disagree significantly with that on the call-back interview, additional call-backs may be made on respondents originally interviewed by the same person. The fear of being caught will discourage cheating.

Other control procedures include the use of “cheater” questions and the analysis of response patterns. Cheater questions are informational questions that will disclose fabricated answers with a reasonably high probability of success. Likewise, the analysis of patterns of responses for interviewer differences will disclose interviewer cheating when significant variations from expected norms occur. Such analyses can be made at very little additional cost.

Time Costs

Perhaps the most common reason for respondent unwillingness to provide accurate information, or any information for that matter, is the result of the time required to make the information available. Respondents often give hasty, ill-considered, or incomplete answers and resist probing for more accurate information. When possible to do so, a respondent will tend to act in a manner that will reduce time costs. Such behavior often results in inaccurate or missing information.

When conducting telephone and personal interviews the interviewer might ask “Is this a good time to answer some questions, or would you rather set a time when I could contact you again?” Experience has shown this latter technique only slightly lowers response rates.

Perceived Losses of Prestige

When information attributing prestige to the respondent is sought, there is always a tendency to receive higher-prestige responses. All researchers experience difficulty both in recognizing the items that demand prestige content, and in measuring the resulting amount of inaccuracy. Information that affects prestige is often sensitive information, including socioeconomic (age, income, educational level, and occupation), place of birth or residence.

> An example of a still more subtle prestige association occurred in a study on nationally known brands of beer. One of the questions asked was, “Do you prefer light or regular beer?” The response was overwhelmingly in favor of light beer. Since sales data indicated a strong preference for regular beer, it was evident that the information was inaccurate. Subsequent investigation revealed that the respondents viewed people who drank light beer as being more discriminating in taste. They had, therefore, given answers that, in their view, were associated with a higher level of prestige.

Measuring the amount of inaccuracy is a difficult task. One solution to this problem is to ask for the information in two different ways. For example, when obtaining information on respondents’ ages, it is a common practice to ask early in the interview, “What is your present age?” and later “In what year did you graduate high school?”

In one study, when respondents were asked, “Are you afraid to fly?” Very few people indicated any fear of flying. In a follow-up study, when they were asked, “Do you think your neighbor is afraid to fly?” (a technique known as the *third-person technique*), most of the neighbors turned out to have severe anxieties about flying.

Invasion of Privacy

Clearly, some topics on which information is sought are considered to be private matters. When such is the case, both nonresponse and inaccuracy in the responses obtained can be anticipated. Matters about which respondents resent questions include money matters or finance, family, life, personal hygiene, political beliefs, religious beliefs, and even job or occupation. It should be recognized however, that invasion of privacy is an individual matter. Thus, information that one person considers sensitive may not be viewed that way by others.

The investigator should attempt to determine sensitivity if it is suspected to be a problem. One way of handling this is to add questions in the pretest stage which ask about the extent of sensitivity to topics and specific questions. A comprehensive treatment of sensitive information and how to ask questions about it is given by Bradburn and Sudman (1979).

Ambiguity

Ambiguity includes errors made in interpreting spoken or written words or behavior. Ambiguity, therefore, occurs in the transmission of information, through either communication or observation.

Ambiguity in Communication

Ambiguity is present in all languages. Unambiguous communication in research requires that the question asked and the answers given each mean the same thing to the questioner and the respondent.

The first step in this process is the controlling one. If the question is not clearly understood by the respondent, frequently the answer will not be clearly understood by the questioner. To illustrate this point, after pretesting in an actual research project on tomato juice, the following question change occurred after pretesting.

Do you like tomato juice?			
	Yes	No	Neither like nor dislike
Yes	○	○	○

➔

Do you like the taste of tomato juice?		
Yes	No	Neither like nor dislike
○	○	○

Even a careful reading of these two questions may not disclose any real difference in their meaning. The analyst who drew up the question assumed that "like" refers to taste. In pretesting, however, it was discovered that some housewives answered "Yes" with other referent in mind. They "like" the amount of Vitamin C their children get when they drink tomato juice, they "liked" the tenderizing effect that tomato juice has when used in cooking of meat dishes, and so on. If the wording of the question had not been changed, there would have been a complete misunderstanding in some cases of the simple, one-word answer "Yes."

A related issue is one where a shortened form of a sentence is used. Examples are, "How come?", "What?", and "How?" This "elliptical sentence" requires the respondent to first consider the context of the sentence and then add the missing parts. When the mental process of transformation is different for the researcher and respondent, communication is lost and interpretation of a person's response is faulty and ambiguous.

The understanding of questions is an issue that goes beyond ambiguity. All too often a respondent may not understand a question, but may have no opportunity to request clarification. Most personal and telephone interviewing uses *standardized interviewing*, meaning that the interpretation of questions is left up to the respondent. As discussed earlier in Exhibit 4.2, one interesting approach taken in online surveys by Qualtrics.com is to use context-sensitive help that provides standardized clarification or instruction for a given question.

Procedures for Recognizing and Reducing Ambiguity in Communication

Every research design that uses communication to obtain information should have as many safeguards against ambiguity as possible. Procedures should be employed to recognize where ambiguity may be present and to reduce it to the lowest practicable level.

Three procedural steps are useful for these purposes and should be considered in every project:

1. Alternative question wording
2. Pretesting
3. Verification by observation

1. Alternative question wording. We have already seen that the present state-of-the-art question formulation cannot guarantee unambiguous questions. In questions where ambiguity is suspected, it is advisable to consider alternative wordings and forms of questions to be asked of sub-samples of respondents.

The use of this simple experimental technique costs no more for online surveys (randomly assign respondents to different blocks of questions). In personal and telephone interviewing situations the interviewers can likewise be instructed to change the order for one-half the interviews. Where significant differences in response are discovered, it will be even more worthwhile as a warning in interpreting the information.

2. Pretesting. Pretesting of questionnaires is a virtual necessity (Converse & Presser, 1986, pp. 51–75). The only way to gain real assurance that questions are unambiguous is to try them. Pretesting is usually done initially by asking proposed questions of associates. To be truly effective, however, pretesting of questions should be conducted by asking them of a group of respondents who are similar to those to be interviewed in the final sample.

A typical way to assess problems with individual questions included in the questionnaire is to ask those participating whether they had any trouble with each of the questions. Be sure to ask about the exact nature of the problem.

If the pretest is done by an interviewer, each respondent can be asked about each question and probing can get more depth in the response. It is the rule, rather than the exception, that questions will be revised as a result of pretesting. Several versions of a question may need to be considered as a result of pretesting before the final version is decided upon.

3. Verification by observation. Information obtained through communication should be verified by observation whenever cost, time, and the type of information desired permit. Clearly, verification by observation is not always possible or practical. For example, a housewife may object to a pantry audit to verify that brands she indicated as preferred are on hand.

Ambiguity in Observation

Although it has been suggested that, where practical to do so, information obtained by communication should be verified by observation, the implication should not be drawn that observation is free of ambiguity. In making observations we each select, organize, and interpret visual stimuli into a picture that is as meaningful and as coherent to us as we can make it. Which stimuli are selected and how they are organized and interpreted are highly dependent on both the expertise, background and frame of reference of the observer.

As an illustration, a cereal manufacturer ran a promotional campaign involving a drawing contest for children. Each child who entered was required to submit (along with a box top) a picture he or she had drawn that depicted Brand X cereal being eaten. The contest was run, the prizes awarded on the basis of artistic merit, and the brand manager turned his attention to other matters. Later, a psychologist who worked for the company happened to see the pictures and was permitted to study them. He found that a sizable proportion of them showed a child eating cereal alone, often with no other dishes on the table. This suggested to him that cereal is often eaten by children as a between-meal snack. Later studies by the company's marketing research department showed that cereals are eaten between meals by children in greater amounts than are eaten for breakfast. The advertising program of the company was subsequently changed to stress the benefits of its cereals as between-meal snacks.

Nonresponse Error

A nonresponse error occurs when an individual is included in the sample but, for any of many possible reasons, is not reached or does not complete the survey. In most consumer surveys this is a source of a potentially sizable error.

Non-response errors differ in nature depending on the mode of survey administration. For example, when using telephone or personal interview methodologies, families who cannot be reached generally have different characteristics than those who can be reached. They may be away from home during the day and differ from those in which at least one member can usually be found at home with respect to age, number of small children, and the proportion of time in which the wife is employed. Similarly, fathers who are unwed, poor, and live in large cities; busy executives and professionals, or occupational group such as hospital purchasing agents responsible for chemical agents provide examples of hard-to-reach-populations that are difficult to locate and interview (Teitler, Reichman, & Sprachman, 2003).

Internet surveys have the potential of increasing contact because the survey invitation appears in the inbox awaiting the potential respondent's reply. However other respondent differences (time pressure, occupation, or lack of interest) and technological issues (spam filters) may similarly increase non-response.

The seriousness of nonresponse error is magnified by the fact that the direction of the error is often unknown. Hansen and Smith (2009) recently showed an 8% increase in response rates by using a two stage presentation of a highly interesting question (a highly interesting question was asked at the beginning of the survey and the respondent was told that the second part of the question would appear at the end of the survey). These additional respondents were shown to provide more centrally distributed responses, thus producing no additional error.

Researchers believe that major reasons for refusal include the public's concerns about data privacy and personal protection; a negative association with telemarketing efforts of all types; consumers' natural aversion to telephone surveys combined with a lack of survey choices for consumers; low salaries of interviewers; and the fact that financial remuneration is not widely used in surveys to compensate consumers for their time.

Evangelista, Albaum and Poon, (1999) suggest that four general motivations drive survey response. Exhibit 4.4 suggests that response rates may be increased (and non-response bias decreased) by using specific motivational techniques and inducements.

EXHIBIT 4.4 Theories of Survey Response

Why do people participate as respondents in a survey? The question is often asked by marketing researchers, perhaps all too often implicitly, and seldom is an answer provided other than in terms of specific techniques (including inducements) that have been used to increase participation. The following theories are among those proposed (and studied to varying degrees) as answers to this question (Evangelista, Albaum and Poon, 1999).

Exchange

The process of using survey techniques to obtain information from potential respondents can be viewed as a special case of *social exchange*. Very simply, social exchange theory asserts that the actions of individuals are motivated by the return (or rewards) these actions are expected to, or usually do, bring from others. Whether a given behavior occurs is a function of the perceived costs of engaging in that activity and the rewards (not necessarily monetary) one expects the other participant to provide at a later date. In order that survey response be maximized by this theory, three conditions must be present:

1. The costs for responding must be minimized.
2. The rewards must be maximized.
3. There must be a belief that such rewards will, in fact, be provided.

Cognitive Dissonance

Cognitive dissonance theory appears to provide a mechanism for integrating, within a single model, much of the empirical research that has been done on inducement techniques for survey response. As used to explain survey response, the theory postulates that reducing dissonance is an important component of the “respond/not respond” decision by potential survey respondents.

The process is triggered by receipt of a questionnaire and invitation requesting participation. Assuming that failure to respond might be inconsistent with a person’s self-perception of being a helpful person, or perhaps at least one who honors reasonable requests, failure to respond will produce a state of dissonance that the potential respondent seeks to reduce by becoming a survey respondent. Since the decision process involves a series of decisions for some people, delaying the ultimate decision may be a way to avoid completing the questionnaire without having to reject the request outright (and thus experience dissonance). Delaying a decision, therefore, may in itself be a dissonance-reducing response.

Self-Perception

Self-perception theory asserts that people infer attitudes and knowledge of themselves through interpretations made about the causes of their behavior. Interpretations are made on the basis of self-observation. To the extent that a person’s behavior is attributed to internal causes and is not perceived as due to circumstantial pressures, a positive attitude toward the behavior develops. These attitudes (self-perception) then affect subsequent behavior.

The self-perception paradigm has been extended to the broad issue of survey response. To increase the precision of this paradigm, the concepts of *salience* (behaviors one has attended to), *favorability* (the affect or feeling generated by a given behavioral experience), and *availability* (information in memory) are utilized. In addition, to enhance the effects, researchers should create labels. *Labeling* involves classifying people on the basis of their behavior such that they will later act in a manner consistent with the characterization. Self-perception would predict that labeling one's behavior would cause that person to view himself or herself as the kind of person who engages in such behavior; therefore, the likelihood of later label consistent behavior is increased.

Commitment and Involvement

Of concern here is the range of allegiance an individual may be said to have for any system of which he or she is a member. Consistent behavior is a central theme, including the following characteristics:

1. Persists over some period of time
2. Leads to the pursuit of at least one common goal
3. Rejects other acts of behavior

Consequently, the major elements of commitment are viewed as including the following:

1. The individual is in a position in which his or her decision regarding particular behavior has consequences for other interests and activities not necessarily related to it.
2. The person is in that position by his or her own prior behavior.
3. The committed person must recognize the interest created by one's prior action, and realize it as being necessary.

A person who is highly committed to some activity is less likely to terminate the activity than one who is uncommitted.

The theory of commitment (or involvement) can be extended to explain survey response behavior. To do this requires recognition that commitment can be attached to many different aspects of a survey, such as the source or the sponsor, the researcher, the topic and issues being studied, and/or the research process itself. To a large extent, commitment is manifested by interest in what is being asked of the potential respondent. The following hypotheses (untested) can be proposed:

1. The less favorable the attitude toward a survey's sponsor, topic, and so forth, the less involvement with, and thus commitment to, anything related to that study.
2. The less the extent of involvement, the more behavior productive of disorder (e.g., nonresponse, deliberate reporting of false information, etc.) is perceived as legitimate.
3. The more behavior productive of disorder is perceived as legitimate, the less favorable the attitude toward the survey.

REDUCING INTERNET SURVEY ERROR

Conducting online surveys has become not only accepted, but the dominant form of conducting structured/direct interviews. This shift to online research is due largely to reduced cost, the availability of dynamic surveys using advanced survey flow logic, the ability to display visually interesting and even interactive graphics, the ease of survey creation and administration, and the ability to eliminate errors associated with data entry, coding and transcription.

In this light, we will focus our attention on online surveys to discuss how we can further reduce or manage four other major sources of error in survey interviewing:

- Coverage error
- Sampling error
- Nonresponse error
- Measurement error

These same sources of error must be addressed regardless of the mode of survey data collection.

Coverage Error

Coverage error occurs when the sample frame (the group from which the sample is drawn) does not represent the population as a whole. For example, a random sample of Apple Mac users would be a mismatch for the adult population of the United States. In more traditional research methods such as mail or telephone methodologies, samples are drawn from sources such as telephone directories, driver's license records, rolls of property owners, credit reports, and so forth. However such sampling frames are very information specific and often do not contain email addresses.

E-mail list brokers offer panels and e-mail address lists that may be targeted to reduce coverage error. Respondent lists can be selected by many variables, including gender, interests (computers, electronics, family, finance, Internet, medical, and travel), and online purchasing. These lists are typically double opt-in, meaning that the users have specifically indicated their agreement to receive surveys or other promotional materials.

When the researcher requires a more detailed set of sample criteria, the cost of reducing coverage error increases. Targeted specialty lists such as physicians of a given specialty are expensive, costing as much as $100 per completed response. While this amount seems large, the cost is much less than other methods of data collection. E-mail name brokers make a practice of not providing the list, but of sending the survey invitation out, thereby controlling their list and avoiding survey abuse of the potential respondents on the list.

Online sampling frames rarely include all elements of the target population. Therefore coverage error will continue to be the greatest source of inaccuracy for online surveys for many years to come. While this same problem is often encountered in the use of mail and phone lists, it is not as severe as with online e-mail lists, which are often based on lists from online websites, including magazines that have specialized hobby and interest affiliations. Carefully selecting lists from well constructed probability panels or panels having millions of members will help to reduce coverage error.

Sampling Error

Sampling error occurs when a non-representative sample is drawn from the sampling frame. The estimation of sampling error requires that probability sampling methods be used, where every element of the frame population has a known nonzero probability of being selected, which may be made the same (i.e., equal) for all. However when the relationship between the sample frame and the target population is unknown, statistical inferences to the target population using confidence intervals may be inaccurate or entirely misleading. In online surveys the degree of sampling error is generally unknown unless the sample is drawn from an online panel or other frame with known size and characteristics. This information is rarely found in consumer research and is rarely estimated.

Online surveys are therefore subject to certain amounts of sampling error. Sampling error may be reduced in part by increasing the sample size. This is an easy task, especially when using an online panel.

Nonresponse Error

Internet researchers are confronted with many non-respondent problems that have elements both unique and common to those faced in telephone surveys. Spam filters, like caller ID monitoring, prevent many survey requests from reaching the "In Box." Internet users often have limited discretionary time resulting in decreased willingness to participate in surveys. This self-selection bias is manifest as potential respondents consider the appeal of the survey topic, survey length, and incentives to complete the survey. The net impact is that without adequate survey response and sample representativeness, non-response error will reduce validity and accuracy of results (Shaffer and Dillman, 1998).

Increasing response rates and reducing nonresponse error most often includes the use of multiple notifications and requests, and the use of personalization in the contact email requesting completion of the interview. Sometimes even these techniques to not produced the desired results. When a population of interest is not adequately represented online or is particularly difficult to interview, a mixed-mode survey strategy should be considered to reduce nonresponse error: a combination of e-mail and telephone, mail, or mall intercept techniques.

> For example, many airline passengers making a connection in Cincinnati during July, 2009 encountered an interviewer in the terminal who was giving travelers a business card with instructions and an online survey code. She requested that the traveler complete the airline satisfaction survey when they returned home or to their office. The contact was quick, novel and non-intrusive. Most travelers kindly accepted the card.

The single most important factor in reducing survey nonresponse is the number of attempts to make contact with each prospective respondent. While many studies have confirmed this fact, one of the more rigorous studies compared response rates for mail and e-mail surveys (Shaffer and Dillman, 1998). In this field study, respondents in the mail and e-mail treatment groups were contacted four times through (1) pre-notifications, (2) letters and surveys, (3) thank-you/reminder notes, and (4) replacement surveys. Results showed no statistically significant difference between the 57.5 percent response rate for the mail group, and the 58.0 percent response rate for the e-mail group.

EXHIBIT 4.5 Increasing Response Rates

The variation in response rates for surveys is enormous, especially when interest and incentives are considered. Ryan Smith, director of sales and marketing at Qualtrics.com, relates his experience with three client surveys that differed greatly in their respective response rates (Smith, 2007). These three very different surveys provide insight into the types of variables that influence response rate:

1. The first survey consisted of a short 10-question survey entitled “What Do Women Want . . . For Valentine’s Day?” This somewhat whimsical survey was sent using a single e-mail blast (with no second communication) to a “random sample” of Internet users through an e-mail list broker. Recipients of the survey were offered the chance to win $500 cash in a random drawing and in addition were promised a copy of the results. This combination of incentives plus a short, interesting survey produced an amazing 43 percent response rate.

2. A second e-mail survey, a very long academic survey of more than 100 questions, focused on developing a demographic, psychographic, and technological expertise profile of the online shopper. This survey measuring attitudes and behaviors was sent through the same broker to a random sample of “Internet shoppers.” Respondents were promised the chance to win $500 cash in one of seven random drawings. The university sponsorship of the survey was identified in the cover letter that contained the professor’s name, contact information, and link to the survey. The response rate was 11 percent.

   A parallel paper and pencil survey was conducted for comparison purposes using a national sample provided by Experian, a provider of credit rating reports. This mail survey was implemented using three separate mailings (1) a pre-notification, (2) the survey, and (3) a follow-up reminder. The mail version produced a 20 percent response rate. Comparison of the mail and online survey results showed that demographic profiles were very different. This difference was attributed to the difference in sampling frames. Respondents to the mail sample were older, had different family structures and were more financially secure. However, demographic differences aside, the psychographic profiles related to online shopping were nearly identical.

3. A third survey sent to brokers by a leading investment firm resulted in a .002% response rate after two mail outs. Further follow up revealed that this fast paced group of potential respondents was too busy to be bothered with a survey.

Smith believes that five actions will greatly increase your online survey response rates:

(1) make your survey as short as possible by removing marginal questions
(2) make your survey interesting to the respondent
(3) include an offer of incentives
(4) use group affiliations whenever possible
(5) use requests that focus on altruistic self perception appeals (I need your help)

It should be noted that although Dillman's response rates for a university faculty population were considerably higher than would be expected for a consumer survey, the similarity across survey modes stands as a solid finding. Perhaps most noteworthy is the finding that when compared with the mail survey, the survey administered by e-mail produced 12.8 percent more respondents who completed 95 percent or more of the questions. Individual item response rates and item completion rates were also higher. For the e-mail based open-ended text responses, the same 12% increase in completion rates was observed, but in addition, responses were longer, averaging 40 words versus 10 words for the paper-and-pencil survey.

Non-response is also important at the question level. Response rates can be improved by using forced responses (the respondent cannot continue until all questions are answered). Albaum et. al. (2010) show that contrary to expectations, using forced response increases both the total response rate and the data quality improve.

It is clear that response rates are improved through the use of multiple contacts to secure cooperation, sending reminders to complete the survey, and forcing a response. This applies not only in traditional mail surveys but also in e-mail surveys. Yet, as shown in Exhibit 4.5 great response rate variations can exist.

Management of the data collection process through state-of-the-art online survey technology offers many new capabilities like survey tracking and personalization of invitations and survey questions. The integration of panel information with the survey (using embedded codes and data) facilitates the identification and tracking of survey respondents and nonrespondents. Also with this integration, personalized follow-up mailings and reminders can be sent to nonrespondents, and survey questions can be personalized to further increase response rates.

Additional technologies enable the careful tracking of respondents. Statistics can be compiled about the data collection status, including the number of surveys e-mailed, the number received by potential respondents, the number of e-mails opened, the number of surveys viewed (link clicked on), and the number of surveys completed. While technological advances help the researcher to reduce non-response rates, it is clear that multiple factors are responsible for nonresponse rates, many of which are not addressable through the administration and handling of the survey.

Measurement Error

Measurement error is a result of the measurement process itself and represents the difference between the information generated on the measurement scale and the true value of the information. Measurement error may be due to such factors as faulty wording of questions, poor preparation of graphical images, respondent misinterpretation of the question, or incorrect answers provided by the respondent.

Measurement error is troublesome to the researcher because it can arise from many different sources and can take on many different forms. For telephone and personal interviews, measurement error will often occur when the interviewer misinterprets responses, makes errors recording responses, or makes incorrect inferences in reporting the data.

Technical issues may similarly create measurement error in online surveys. The size and resolution of the monitor, browser, operating system (Mac, Microsoft Windows, Linux), and even web page color pallet may change the appearance of the survey. Additionally, skins or templates may affect the actual survey's appearance by adjusting the spacing between questions, the appearance of horizontal lines separating questions or sections, the use of horizontal versus vertical scales, drop-down boxes versus checkboxes or radio buttons, and even font characteristics including size, typeface, the use of boldface and italics, and spacing between scale items.

Researchers have for decades compared measurement error differences for the various modes of data collection. While differences do exist, online surveys compare favorably with traditional paper-and-pencil and telephone surveys. Standard surveys follow the structured/direct approach for data collection and change little when transitioning from paper and pencil to radio button or checkbox formats. However, as will be discussed later, the differences between the less personal online surveys and the in-person qualitative research are far more extreme.

Many of the traditional measurement errors associated with transcription and recording of data are eliminated with electronic real-time entry of the data. With Internet surveys, the survey as well as the analysis of results can be conducted in real-time and posted to a secure Web site in hours. In one recent survey of programmers and software developers conducted by Qualtrics.com for Microsoft, 6,000 invitations were sent out with the promise of a $20 Amazon.com gift certificate. Nine hundred completed surveys were received within 48 hours as results were monitored online.

Online studies may be completed in 24 hours, as compared to the four to ten weeks required for paper-and-pencil methodologies. Mail surveys must be prepared, printed, mailed, followed up with mail reminders, manually coded, typed or scanned into the database, analyzed and then compiled into a managerial report. These many steps involve many participants with varying levels of expertise, and each may introduce error. Internet based surveys eliminate many of these steps and combine other steps to complete the research much more quickly and easily. Efficiencies aside, no matter which mode is used for survey completion, error control must be addressed to assure quality results.

SUMMARY

This chapter first examined the various types of information that can be obtained from respondents. It then considered communication as a means to obtain information from respondents. The types of respondent interviews—structured-direct, unstructured-direct, and structured- and unstructured-indirect—were discussed.

The next section introduced the concepts of inaccuracy and ambiguity as the major sources of response and non-response bias. Predictive and concurrent sources of inaccuracy were discussed in the context of respondent inability or unwillingness to respond. Methods of reducing non-response error were then discussed in the context of theories of survey response. Finally our discussion focused on how to reduce coverage, sampling, non-response and measurement errors in online surveys.

The objective of marketing research is to understand the consumer and apply information and knowledge for mutual benefit. Technological advances in online marketing research provide the ability to monitor customer knowledge, perceptions, and decisions to dynamically generate solutions tailored to customer needs. In this chapter we have stressed the need to improve the research process by reducing errors in the research process. Perhaps the biggest mistake the market researcher can make is to view research options as time- and cost-saving tradeoffs across the data collection options. New technologies continue to be developed, but each must be tested for applicability under specific research conditions, and refined so that marketers are able to better identify and measure the constructs being investigated.

REFERENCES

Gerald Albaum, Catherine A. Roster, James Wiley, John Rossiter, and Scott M. Smith (2010). Designing Web Surveys in Marketing Research: Does Use of Forced Answering Affect Completion Rates?, Working paper.

Evangelista, F., Albaum, G., & Poon, P. (1999, April). An empirical test of alternative theories of survey response behavior. *Journal of the Market Research Society, 41*, 2, 227–244

Bradburn, N., & Sudman, S. (1979). *Improving interview method and questionnaire design.* San Francisco: Jossey-Bass.

Stewart, C. J. and Cash, W. B., *Interviewing Principles and Practices, 4/e*. © 1985, William C. Brown, Publishers. Reprinted with permission of The McGraw-Hill Companies, pp. 9–13

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Beverly Hills, CA: Sage.

Fowler, F. J. (1991). Reducing interviewer-related error through interviewer training, supervision, and other means. In P. P. Bremer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman, *Measurement errors in surveys* (pp. 259–278). New York: Wiley.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related errors*. Newbury Park, CA:

Hansen, Jared M., Scott M. Smith, and Michael D. Geurts (2009), "Improving Survey Completion Rates and Sample Representativeness Using Highly-Interesting Questions: A National Panel Experiment Comparing One and Two Stage Questions" http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1313265

Morwitz, V. G. (1997). It seems like only yesterday: The nature and consequences of telescoping errors in marketing research. *Journal of Consumer Psychology*, *6*(1), 1–29.

Semon, T. T. (2000a, August 14). Better questions means more honesty. *Marketing News, 34*, 10.

Semon, T. T. (2000b, January 17). If you think a question is stupid—it is. *Marketing News*, *34*, 7.

Schaffer, D. R., & Dillman, D. A. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly, 62*, 378–397.

Smith, R. (2007). Personal communication [interview].

Stewart, C. J. and Cash, W. B., *Interviewing Principles and Practices*, *4/e*. © 1985. William C. Brown, Publishers. Reprinted with permission of The McGraw-Hill Companies, pp. 9–13

Sudman, S., Finn, A., & Lannam, L. (1984, Summer). The use of bounded recall procedures in single interviews. *Public Opinion Quarterly*, *48*, 520–524.

Teitler, J. D., Reichman, N. E., & Sprachman, S. (2003). Cost and benefits of improving response rates for a hard-to-reach population. *Public Opinion Quarterly*, *67*, 126–138.

Chapter 5
INTERVIEWING MODES: PERSONAL-CALL-SEND

Survey information is obtained from respondents through communication in several alternative media modes. Respondents may be interviewed in person, by telephone, or they may be sent a questionnaire. Mail, FAX and the Internet surveys are similar in that non-personal self reporting is involved.

PERSONAL INTERVIEWS

As the name implies, the personal interview consists of an interviewer asking questions of one or more respondents in a face-to-face situation. The interviewer's role is to get in touch with the respondent(s), ask the desired questions, and record the answers obtained. Recording of the information may be done either during or after the interview. In either case, it is the interviewer's responsibility to ensure that the content of the answers is clear, unambiguous and that information has been recorded correctly.

While it is significantly more expensive on a per-completed-interview basis, the personal interview, as a collection medium, has several advantages relative to telephone interviews and mail questionnaires. It provides the opportunity to obtain a better sample, since virtually all the sample units can be reached and, with proper controls and well-trained interviewers, nonresponse to the survey or to individual questions can be held to a minimum. It also gives the opportunity to obtain more information, as a personal interview can be of substantially greater length than either a telephone interview or mail questionnaire. Finally, it permits greater flexibility. More freedom is provided for adapting and interpreting questions as the situation requires, especially in the case of unstructured personal interviews where visual, auditory, or olfactory aids are used.

The limitations of the personal interview include time, cost, and the response bias that may be induced by poorly trained or improperly selected interviewers. Problems with personal interviews arise from its very nature in that it is a social interaction between strangers, often on the respondent's territory, initiated by an interviewer who may have little in common with the respondent.

In addition to the home and workplace, many studies conduct consumer interviews in malls, where the so-called mall-intercept method is used. This method avoids the logistic, financial and time costs of travel to meet with respondents. The mall intercept method involves having interviewers stationed at selected places in a mall who request interviews from people passing by. Presumably the people are chosen on the basis of a predetermined sampling plan. At times, monetary incentives may have positive effects (Wiseman, Schafer, & Schafer, 1983).

The mall-intercept method is a widely used method of data collection in marketing research. Many malls within the United States have permanent research facilities located within them. These facilities may be equipped with videotape equipment, private interviewing compartments, food preparation facilities for taste tests, and a variety of other research equipment. Soundproof rooms free from distractions and equipped with proper lighting and materials can contribute to reliable data collection. Researchers can observe an interviewer's technique and check the completed work immediately.

Overall quality of data (completeness, depth) appears to be about equal to that of other methods, since mall-intercept respondents are more frequent users of shopping centers, and they may be better able to provide more brand and store-oriented information than respondents contacted by other means.

EXHIBIT 5.1 Mall Intercepts Are Widely Used

According to Katherine Smith (1989), mall intercepts have the following advantages:

1. They allow researchers to conduct visual, auditory and taste tests of ads, products and other physical stimuli.
2. They offer an opportunity to obtain immediate response.
3. They potentially provide more depth of response than non-face-to-face interviews.
4. Researchers can use equipment to analyze responses (for example, voice-pitch or eye movement tracking analysis).
5. A large number of respondents from a wide geographic area can be interviewed in a limited time.
6. Researchers can control the interviewing environment and supervise the interviewer.

Mall intercept studies are less expensive than door-to-door interviewing, because travel time and the "not-at-home problem" are eliminated. However, it is becoming increasingly more difficult to locate people at home, and even more people are hesitant to let strangers inside.

Using the mall intercept, interviewing often takes place where members of the population of interest are doing something related to what is being measured. For studying certain types of products or behaviors, the mall is a more realistic setting when a respondent is being asked to make choices. Finally, using certain sampling methods, the mall-intercept procedure may give a better distribution of respondents.

Despite all these virtues, mall intercepts have limitations:
1. The mall customer may not reflect the general population.
2. The intercept is not well-suited to probability sampling.
3. Shoppers in a hurry may respond carelessly.
4. The interview time constraint is more severe with mall intercepts than with other personal interviewing methods.

A variation of the mall intercept that is often used in business is to interview at conferences, sales meetings, or other gatherings representing the population of interest.

In more general terms, research is better conducted on site whenever the topic is about the business, when the purchase decision is made on the premises, or when the population of interest is represented. Respondents are most likely to recall and discuss their experiences during the experience, not days later during a survey. David Kay (1997), a partner in Research Dimensions International, suggests there are five types of interviews for on-site research:

1. **Stream of consciousness interview.** This is a conversation with questions designed to elicit what the respondent is experiencing at every moment of shopping.
2. **Spontaneous reaction interview.** This asks for spontaneous, minimally prompted reactions of customers to their environment.
3. **Directed general-response interview.** Useful to assess effectiveness of strategy, this method asks general questions directed to the strategy.
4. **Directed specific-response interview.** This is useful to determine why consumers feel as they do, as indicated by answers to other questions.

5. Prompted reaction to execution elements. This is designed to elicit response to specific elements. For example, an in-store taste test might include the question "What do you think about the taste of China Sea brand Spring Rolls?"

The obvious advantages of on-site interviews are that the respondent is usually in a proper state of mind and has better task or product recall. In addition it is easier to contact the actual target group, making the response rates are higher. On-site interviews seem to produce more robust information.

Paying people to participate in surveys, in the form of prepaid incentives, tends to increase overall response rates for personal interviews, as well as for other types of interviews. But, do not seem to influence the quality of the data collected (Davern, Rockwood, Sherrod, and Campbell, 2003).

THE TELEPHONE INTERVIEW

Telephone interviews are often used in lieu of personal interviews, especially when personal contact is desired, when the information must be collected quickly and inexpensively, and when the amount of information required is relatively limited. Compared to e-mail or mail surveys, telephone interviews often are more costly in terms of total costs of data collection. However, when cost is figured on a per completed questionnaire basis, telephone interviews may be less costly than mail, but more costly than e-mail. In addition, telephone surveys offer the opportunity to probe for clarification or further information.

It is generally recognized that for business to business and consumer research, telephone interviewing is as effective as personal interviewing for scope and depth of information obtained. In addition, when a telephone survey is conducted from a call center, they can be better supervised than personal interviews.

Virtually all telephone interviews are structured direct interviews. However, when the population to be studied is business decision makers, some research practitioners believe that more information may be obtained using the telephone than by conducting focus groups (Eisenfeld, 2003). For business people and consumers alike, it is frequently easier to get 10 minutes of telephone cooperation, than a longer personal interview or attendance at a focus group.

With a detailed database to use as a sample frame, interviews of business people, current customers, former customers, and prospects all can be contacted relatively easily. Furthermore, pre-notification letters can also be sent. A recent study of political telephone surveys concluded that advance pre-notification letters can significantly increase response rates (Goldstein & Jennings, 2002).

The likelihood of the potential respondent refusing to be interviewed is always present when starting a telephone interview. Telephone surveys are unique in that they allow the interviewer to respond to the potential respondent and attempt to turn a refusal into a completed interview. In his classic treatise on telephone surveys, Dillman (1978) identifies common reasons people give for refusals and suggests some possible responses the interviewer can give. These responses can help the researcher handle objections and refine their interviewing skills. These are shown in Table 5.1.

TABLE 5.1 Possible Answers to Reasons for Refusals

Reasons for Refusing	Possible Responses
Too busy	This should only take a few minutes. Sorry to have caught you at a bad time, I would be happy to call back. When would be a good time for me to call in the next day or two?
Bad health	I'm sorry to hear that. Have you been sick long? I would be happy to call back in a day or two. Would that be okay? (If lengthy or serious illness, substitute another family member. If that isn't possible, excuse yourself and indicate they will not be called again.)
Too old	Older people's opinions are just as important in this particular survey as anyone else's. In order for the results to be representative, we have to be sure that older people have as much chance to give their opinion as anyone else. We really do want your opinion.
Feel inadequate: Don't know enough to answer	The questions are not at all difficult. They mostly concern your attitudes about local recreation areas and activities, rather than how much you know about certain things. Some of the people we have already interviewed had the same concern you have, but once we got started they didn't have any difficulty answering the questions. Maybe I could read just a few questions to you and you can see what they are like.
Not interested	It's awfully important that we get the opinions of everyone in the sample; otherwise the results won't be very useful. So, I'd really like to talk with you.
No one else's business what I think	I can certainly understand, that's why all of our interviews are confidential. Protecting people's privacy is one of our major concerns, and to do it people's names are separated from the answers just as soon as the interview is over. And, all the results are released in a way that no single individual can ever be identified.
Objects to surveys	We think this particular survey is very important because the questions are ones that people in parks and recreation want to know answers to, so they would really like to have your opinion.
Objects to telephone surveys	We have just recently started doing our surveys by telephone, because this way is so much faster and it costs a lot less, especially when the survey is not very long, like this survey.

SOURCE: Reprinted from *Mail and Telephone Surveys*: *The Total Design Method* by Dillman, D. Copyright © 1978. This material is used by permission of John Wiley & Sons, Inc.

As with all modes of surveying, telephone surveys benefit from the use of inducements or incentives—monetary or nonmonetary—to encourage potential respondents to participate. Incentives may be promised or sent in advance with a preliminary letter when the mailing address of the potential respondent is known, or they may also be offered when the initial request for participation is a refusal. When used this way it is known as a *refusal conversion incentive*.

The main purpose of such incentives is to generate a greater response rate with the effect of reducing nonresponse error. But the use of incentives has implications as well. First, total cost will increase, although cost per response may decrease depending on how effective the incentive is. Second, data quality may be affected, leading to a change in response bias, which may be a positive or negative change. Third, sample composition may be affected, again with a positive or negative effect. Fourth, expectations of interviewer and respondent may be changed. Finally, interviewer effort may be affected.

The telephone survey may be a good approach to reach specific market segments, particularly when door-to-door interviews are not possible or might lead to serious distortions in response. It is obvious that there must be sufficiently high telephone penetration in the segment for this mode of data collection to be advantageous. For example, the use of surname sorts makes telephone surveys the most efficient way to locate, contact and survey ethnic groups in the United States.

The basic limitations of telephone interviews are the relatively limited amounts of information that can be obtained (at least compared with alternative methods) and the bias that exists in any sample of home telephone subscribers. More than 25 percent nationally and more than 50 percent in large cities are not listed in a published directory, either because they have an unlisted number or as a result of moving (www.busreslab.com/articles/article3.htm). A technique for including unlisted telephone numbers in the sample frame is called random digit dialing (RDD).

Another additional problem for telephone researchers is that home telephone subscribers are disappearing. Currently about 1 in 5 homes do not have a "landline" telephone, but rely instead on cell phones or computer based phone services.
http://www.azcentral.com/business/articles/2009/07/20/20090720biz-cellnation0721side.html

Additional problems associated with telephone interviewing are those of sample control and interviewer performance. Often this is manifested by inadequate efforts to complete interviews with some of the harder-to-reach respondents. Adding another sample is no substitute for dealing properly with the original sample.

Another aspect of interviewer performance that can influence response rates and data quality is actually something beyond any given interviewer's control. This is the interviewer's accent. There is evidence that accent can influence respondents' participation. Linguistic experts have found that listeners form mental impressions of people who speak with an accent different from theirs, impressions that may lead to a refusal or bias the responses. In the United States there are many region-specific accents (New England, the Deep South) and also those that are cultural (Hispanic, Asian, Indian). When faced with an unfamiliar accent, people may have trouble communicating. When communication becomes difficult, refusals will increase. Alternatively, some accents increase curiosity (British, Irish) and can actually increase response rates. In addition, respondents who identify an interviewer's accent may apply preconceived biases to the interviewer and to the survey. Accent-free interviewing eliminates one potential source of nonresponse and bias. At the very least, if a study is regional in nature, then having interviewers from that region will also reduce nonresponse and bias.

The so-called "caller-id" telephone technology problem has emerged for telephone surveys as the number of households having caller-id and answering machines has increased. One way people use answering machines is to screen incoming calls. The second, of course, is to allow those calling to leave a message when members of the household are not at home. One might be tempted to assume that screening and leaving messages would allow potential respondents to choose to not participate—a form of refusal. Interestingly, some research conducted on this issue has found that households with answering machines were more likely to complete the interview and less likely to refuse to participate compared to households where there was no answer on the initial call attempt (Xu, Bates, & Schweitzer, 1993). This suggests that where answering machines are operating, the call can represent a form of pre-notification. It is generally believed that pre-notification in any type of survey increases participation rates. But this phenomenon also has the potential to generate bias in the sample (Oldendick & Link, 1994).

In the United States, major legislation affecting telemarketing went into effect in 2003. This was the creation of the National Do Not Call Registry. Research practitioners are exempt from this law. Of primary concern to the research community is the use by telemarketers of selling-under-the-guise-of-research techniques. For those interested, the European Society for Opinion and Marketing Research (ESOMAR) has responded by issuing guidelines for distinguishing telephone research from telemarketing (ESOMAR, 1989). A good overview of telephone surveys is given by Bourque and Fielder (2003a).

THE MAIL INTERVIEW

Mail interviews have in the past been widely used. Mail questions provide great versatility at relatively low cost and are particularly cost effective when included as part of a scheduled mailing, such as a monthly correspondence or billing. A questionnaire may be prepared and mailed to people in any location at the same cost per person: the cost of preparing the questionnaire, addressing the letter or card sent, and the postage involved. Respondents remain anonymous unless a name is requested, the questionnaire is openly coded, or some ethically-questionable practice is employed.

Timeliness of responses is critical in mail surveys. If the time given is reasonable, say one or two weeks, stating a deadline should not adversely affect the response rate. Stating such a deadline may encourage the potential respondent not to postpone the task indefinitely.

The overall process of data collection from a mail survey is summarized below as a sequence of contact activities for an optimal mail survey. With minor modifications this general sequence is applicable to personal interview, telephone, and e-mail surveys. When designing a survey, the researcher must consider issues that can affect response rate and data quality, including those shown in Table 5.2.

TABLE 5.2 Selected Dimensions of a Mail Survey and Alternatives for Choice

Dimension	*Alternatives*
Preliminary notification	Letter, postcard, telephone call, e-mail, none
Reminder	Letter, postcard, telephone call, e-mail, none
Cover letter	Separate item, included as first page of questionnaire Personalized, non-personalized Color of ink in signature (black, blue)
Length of questionnaire	Number of pages
Format of questionnaire	Print front, print front and back, individual stapled pages, booklet
Type of outgoing postage	First-class stamp, first-class metered, bulk, nonprofit (where appropriate)
Return envelope postage	First-class stamp, metered, first-class permit, none
Inducements	Monetary (amount), nonmonetary (pen, silver jewelry, trinkets of all types), contribution to charity, none when given (prepaid, promise to pay)
Coding with a number	Yes (on questionnaire, on return envelope), none
Anonymity/Confidentiality	Yes, no
Endorsement	Yes, no

Increasing Response Rates

Perhaps the most serious problem with mail surveys is that of nonresponse. Typically, people indifferent to the topic being researched will not respond. It is usually necessary to send additional mailings (i.e., follow-ups) to increase response. But even with added mailings, response to mail questionnaires is generally a small percentage of those sent; the modal response rate is often only 20 to 40 percent. On the front end of the surveying effort, response rates are increased through preliminary contact by letter or telephone call, cover letters, and monetary or non-monetary (a gift) inducements.

Other experimental research has evaluated response effects of questionnaire format and length, survey sponsorship, endorsement, type of postage, personalization, type of cover letter, anonymity and confidentiality, deadline date premiums and rewards, perceived time for task, and the use of a follow-up reminder,. Reported nonresponse rates and accuracy of data for experiments involving these techniques vary, and there appears to be no strong empirical evidence that any one is universally better than the others, except that it is better to use a follow-up and use monetary or nonmonetary incentives.

Cover letters are included as the first page of a questionnaire is shown in Exhibit 5.2. This letter could have been a separate page if more space was needed for the questionnaire. Note that email invitations should be much shorter, perhaps limited to only one short paragraph. There is no evidence that any alternative is universally better than another within each dimension. The best rule of thumb is to use common sense. Further discussion of these will be found in the many review articles and studies published in such sources as the *Journal of Marketing Research* and *Public Opinion Quarterly*, in the book by Bourque and Fielder (2003b), and in the classic works of Dillman (1978, 2000).

EXHIBIT 5.2 Example of a Cover Letter

My colleague, Dr. David Boush, and I are engaged in a study of consumers' use of financial services. The broad objective of this study is to gain an understanding of how people use banks and similar financial organizations, and what characteristics influence their behavior with such companies. The Bank of Anytown has agreed to cooperate with us in this endeavor by assisting us in data collection.

The enclosed questionnaire is being sent to a large number of the customers of the Bank of Anytown, each of whom has been selected by a random process. I would greatly appreciate your completing the questionnaire and returning it in the envelope provided. Please note that you do not have to add postage to this envelope.

All individual replies will be kept in strictest confidence. No person associated with The Bank of Anytown will see any questionnaire. Only aggregate results will be shown in our write-up of the results. No person other than Dr. Boush, myself, and our research assistant will ever see a completed questionnaire. If you do not wish to participate in this survey simply discard the questionnaire. Completing and returning the questionnaire constitutes your consent to participate.

The code number at the top of the questionnaire will be used only for identifying those people who have not responded so that he or she will not be burdened by receiving a follow-up mailing. After the second mailing has been made, all records that match a number with a person's name will be destroyed.

The success of this project depends upon the assistance of persons such as yourself. If you have any questions, please call me at 503-346-4423.

Sincerely,
Gerald Albaum
Professor of Marketing

Endorsements are an intriguing dimension of survey research. An endorsement is an identifying sponsorship that provides "approval and support for a survey from an individual or organization." An endorsement can be included as a company logo or a person under who's signature the letter is sent. Unknowingly, an endorsement may have a positive, neutral, or negative effect, depending on how the endorser is perceived by a potential respondent.

Rochford and Venable (1995) found that significantly higher response rates were observed when there was endorsement by an external party associated with the targeted audience than when there was no such endorsement. In addition, endorsements by locally known individuals produced higher response rates than endorsements by highly placed but less well-known individuals from a national headquarters office.

Since people responding to a mail questionnaire tend to do so because they have stronger feelings about the subject than the nonrespondents, biased results are to be expected. To measure this bias, it is necessary to contact a sample of the nonrespondents by other means, usually telephone interviews. This is a type of nonresponse validation. The low level of response, when combined with the additional mailings and telephone (or personal) interviews of nonrespondents, results in substantial increases in the per-interview cost. The initial low cost per mailing may therefore be illusory. On the other hand, the nonresponse validation may indicate that populate subgroups have not been omitted and that results may not be biased.

Variations on Mail Interviews

Many variations of the mail interview are frequently used. These include the warranty card, hand delivered surveys, newspaper/magazine surveys, the fax back survey, survey on the back of checks, website polls (a one question survey) and of course, the email survey.

Warranty cards often ask for information about where the item was purchased, what kind of store or outlet sold it, when it was purchased, and other variables such as demographics and life style. Although warranty cards do not provide an extensive amount of information, response rates are substantially higher than for the usual mail questionnaire and the cards are useful for creating a customer database.

Another variation is the questionnaire that is either printed or inserted in a newspaper or magazine. Potential respondents are requested to mail or fax this back to a designated address/fax number. Among the many problems with this approach is the lack of any formalized control over the sample. Yet, despite this major limitation, the approach does have a possibility of better hitting the target population.

Another modification of including a questionnaire in a printed publication asks respondents to return the questionnaire by fax. One advantage of this approach is that those who do respond will be truly committed to the project. But, again there is little control over the sample. One major Airlines routinely printed questionnaires to be returned by fax in its in-flight magazine.

Potential advantages of using fax include quick contact, rapid response, retention of the original document format and visual images used, modest cost, and automated faxing that works directly with the computer (Baker, Hozier, & Rogers, 1999).

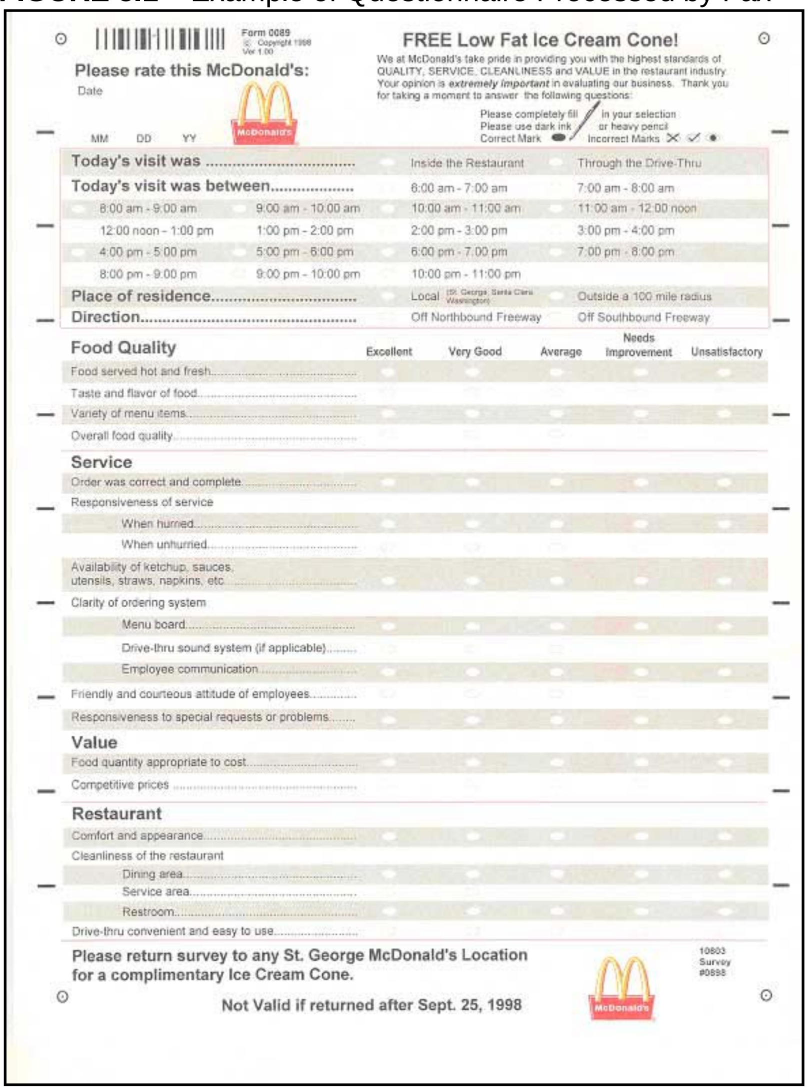
One hybrid approach, used by McDonalds, had patrons to take a brief in-store survey on a mark sense sheet, as shown in Figure 4.3. When completed, the store faxed the sheets for processing using a technique that converted the image directly to data without being printed to paper on the receiving end. The results were then available online in real time.

WEB AND E-MAIL INTERVIEWS

As computer coverage in home markets increase, the use of electronic surveys has increased. Web and e-mail surveys are fulfilling their promise to be a driving force in marketing research. Currently (2009) 73 percent of all U.S. households have Internet access in their home. http://www.marketingcharts.com/interactive/home-internet-access-in-us-still-room-for-growth-8280/

The Internet has experienced a growth rate that has exceeded any other modern technology, including the telephone, VCR, and even TV. The Internet has diffused from a highly educated, white-collar, upper-income, male dominated core. At the opposite end of the spectrum, the approximately 25% who are non-adopters include a disproportionate number of elderly, single mothers, African Americans and Hispanics, and lower-income individuals. For some studies, this may be a serious limitation. However even today, most groups like company employees, students, or association members have nearly 100 percent Internet access and check e-mail on a daily basis.

FIGURE 5.1 Example of Questionnaire Processed by Fax



Form 0089
© Copyright 1998
Ver 1.00

Please rate this McDonald's:

Date

MM DD YY

FREE Low Fat Ice Cream Cone!

We at McDonald's take pride in providing you with the highest standards of QUALITY, SERVICE, CLEANLINESS and VALUE in the restaurant industry. Your opinion is ***extremely important*** in evaluating our business. Thank you for taking a moment to answer the following questions:

Please completely fill in your selection
Please use dark ink or heavy pencil
Correct Mark Incorrect Marks

Today's visit was .... Inside the Restaurant / Through the Drive-Thru

Today's visit was between........ 6:00 am - 7:00 am / 7:00 am - 8:00 am / 8:00 am - 9:00 am / 9:00 am - 10:00 am / 10:00 am - 11:00 am / 11:00 am - 12:00 noon / 12:00 noon - 1:00 pm / 1:00 pm - 2:00 pm / 2:00 pm - 3:00 pm / 3:00 pm - 4:00 pm / 4:00 pm - 5:00 pm / 5:00 pm - 6:00 pm / 6:00 pm - 7:00 pm / 7:00 pm - 8:00 pm / 8:00 pm - 9:00 pm / 9:00 pm - 10:00 pm / 10:00 pm - 11:00 pm

Place of residence........ Local (St. George, Santa Clara, Washington) / Outside a 100 mile radius

Direction........ Off Northbound Freeway / Off Southbound Freeway

	Excellent	Very Good	Average	Needs Improvement	Unsatisfactory
Food Quality					
Food served hot and fresh					
Taste and flavor of food					
Variety of menu items					
Overall food quality					
Service					
Order was correct and complete					
Responsiveness of service					
When hurried					
When unhurried					
Availability of ketchup, sauces, utensils, straws, napkins, etc.					
Clarity of ordering system					
Menu board					
Drive-thru sound system (if applicable)					
Employee communication					
Friendly and courteous attitude of employees					
Responsiveness to special requests or problems					
Value					
Food quantity appropriate to cost					
Competitive prices					
Restaurant					
Comfort and appearance					
Cleanliness of the restaurant					
Dining area					
Service area					
Restroom					
Drive-thru convenient and easy to use					

Please return survey to any St. George McDonald's Location for a complimentary Ice Cream Cone.

10803 Survey #0898

Not Valid if returned after Sept. 25, 1998

Electronics driven lifestyles that include online social networks, massive use of texting and pervasive internet connections are no doubt responsible in part for seemingly responsible for attitude and behavioral changes in the way we view our increasingly virtual world. Strong upward trends are observed in the percentage of Internet purchases for airline tickets, CDs, DVDs, books, computer software, hardware and systems. These online customers provide excellent access for research purposes.

Advocates of online surveying quickly point to the elimination of mailing and interviewing costs, elimination of data transcription costs, and reduced turnaround time as the answer to client demand for lower cost, timelier, and more efficient surveys. As a result, online marketing research has become so widely accepted that online research has been optimistically projected to account for as much as half of all marketing research revenue, topping $3 billion. While these numbers appear to be overly optimistic, it is clear that online research is growing and that researchers operate in a much faster-paced environment than ever before. This pace will continue to increase as new modalities for research open: wireless PDAs, Internet-capable mobile phones, Internet TVs, and other Internet-based appliances yet to be announced. Each is an acceptable venue for interacting with the marketplace and conducting online research.

Substantial benefits accrue from the various approaches to computer-assisted data collection in surveys as shown below:

- Respondents need few computer related skills
- Respondent choose their own schedule for completing survey
- Can easily incorporate complex branching into survey
- Can easily pipe and use respondent generated words in questions throughout the survey
- Can accurately measure response times of respondents to key questions
- Can easily display a variety of graphics and directly relate them to questions
- Eliminates need to encode data from paper surveys
- Errors in data less likely, compared to equivalent manual method
- Speedier data collection and encoding compared to equivalent manual method.

ACNielsen (Miller, 2001) reported the results of 75 parallel tests comparing online and traditional mall intercept methods. Researchers noted high correlations in aggregate purchase intentions. While online measures may yield somewhat lower score values, recalibration of averages against appropriate norms produced accurate sales forecasts. Wilkie further reported that while responses gathered using different survey modes may be similar, the demographic profiles of online and traditional respondents groups do differ.

Given that the current percentage of households online is approximately 75 percent, statistical weighting of cases could be used to adjust demographic differences of online groups to match mall intercept or telephone populations. However, the possibility of weighting actually raises the question of whether it is better to model phone or mall intercept behavior (which are also inaccurate) or to attempt to independently model the actual behavior of the respondents.

TABLE 5.3 Comparative Evaluation of Alternative Survey Methods of Data Collection

Criteria	*Telephone CATI*	*In-Home Interviews*	*Mall-Intercept Interviews*	*CAPI*	*Mail Surveys*	*Mail Panels*	*Internet/ Web*
Flexibility of data collection	Moderate to high	High	High	Moderate to high	Low	Low	Moderate to high
Diversity of questions	Low	High	High	High	Moderate	Moderate	Moderate to high
Use of physical stimuli	Low	Moderate to high	High	High	Moderate	Moderate	Moderate
Sample Control	Moderate to high	Potentially high	Moderate	Moderate	Low	Moderate to high	Low to moderate
Control of data collection environment	Moderate	Moderate to high	High	High	Low	Low	Low
Control of field force	Moderate	Low	Moderate	Moderate	High	High	High
Quantity of data	Low	High	Moderate	Moderate	Moderate	High	Moderate
Response rate	Moderate	High	High	High	Low	Moderate	Very low
Perceived anonymity of respondent	Moderate	Low	Low	Low	High	High	High
Social desirability	Moderate	High	High	Moderate to high	Low	Low	Low
Obtaining sensitive information	High	Low	Low	Low to moderate	High	Moderate to high	High
Potential for interviewer bias	Moderate	High	High	Low	None	None	None
Speed	High	Moderate	Moderate to high	Moderate to high	Low	Low to moderate	Very high
Cost	Moderate	High	Moderate to high	Moderate to high	Low	Low to moderate	Low

SOURCE: From Malhotra, N., *Marketing Research: An Applied Orientation,* 4th edition, Copyright © 2004. Reprinted with permission of Pearson Education, Inc., Upper Saddle River, NJ.

PROBABILITY AND NONPROBABILITY SURVEY APPROACHES

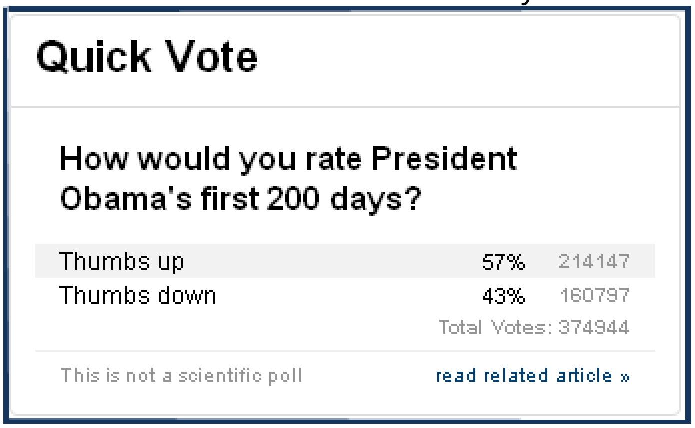
A variety of approaches to presentation of surveys and recruitment of respondents are used on the Web. Surveys based on probability samples, if done properly, provide a bias-free method of selecting sample units and permit the measurement of sampling error. However the majority of online research can be typified as a one-shot mailout, the objective of which is to obtain a sufficient number of completed responses. With these types of studies, there is rarely much thought given to the issues of representativeness of the sample, or validity and accuracy of the results. The question is, then, what is required for effective online research?

Online Nonprobability Surveys

Nonprobability samples offer neither of these features. Nonprobability-sample–based surveys, generally for entertainment or to create interest in a Web site, are self-selected by the respondent from survey Web sites either for interest or compensation, or are provided to members of volunteer panels.

As an example of the Web site interest variety of survey, Figure 5.2 shows a CNN.com "Quick Vote" survey, which includes a link to the online results page.

FIGURE 5.2 Web Site Interest Survey



Source: http://edition.cnn.com/

The National Geographic Society Web site offers surveys that focus on a variety of educational, social and environmental issues at www.nationalgeographic.com/geosurvey/. Surveys include lengthy inventories covering demographics, Internet usage, and attitudes about such topics as geographic literacy, conservation and endangered species, culture, and a variety of other topics. The 2006 Global Geographic Literacy Survey was conducted jointly with Roper Research to assess the geographic knowledge of people ages 18 to 24 across the United States. Specific questions focused on benchmarking attitudes towards the importance of geography and how aware young adults are of geography in the context of current events.

Other well-recognized nonprobability surveys include the ACNielsen BASES (test marketing) and Harris-Black panels. Although nonprobability surveys, these panels are continually redefined to match the demographic characteristics of telephone and mall intercept surveys. The parallel telephone and mall intercept studies provide weighting to proportionately adjust online samples to reduce selection bias.

Online Probability Surveys

Probability-based surveys allow the researcher to estimate the effects of sampling error and thereby provide inferences about the target population through hypothesis testing. Coverage errors, nonresponse errors, and measurement errors still apply and may reduce the generalizability of the data. Online probability samples generally result where e-mail surveys are sent to comprehensive lists that represent the target population. When the target population is large, random samples from the list will be used. For smaller populations such as employees of a company, the survey may be sent to the entire population, thus representing a census.

Where the target population of interest is visitors to a given Web site, pop-up surveys may be presented randomly to visitors during their visit to the site. In this case, the target population is well defined and the sample element has a known nonzero probability. The Qualtrics.com Site Intercept tool allows the market researcher to control all pop-up and pop-under content without the assistance of the IT department. Surveys, white papers, and messages can be distributed based on a variety of conditions, including multiple pages visited that contain specific key words. For example alternative surveys about kitchen appliances could be administered if the visitor viewed pages dealing with major appliances (stoves and refrigerators) rather than counter top appliances (mixers and waffle irons).

Pre-recruited online panels, when recruited using probability-based sampling methods such as random-digit telephone dialing, also produce probability surveys. In this case, random digit dialing would be used to contact the prospective panel members who would be qualified as Internet worthy before being recruited for the panel. The initial interview may also include a variety of demographic, lifestyle, and computer usage questions that would help in weighting the panel, thereby reducing selection bias.

Mixed Mode Studies

Mixed-mode designs provide another alternative for the researcher, presenting respondents with a choice of responding via online survey or via another mode. Respondents contacted by mall intercept, telephone, mail, or other probability-based sampling mechanism, are given the opportunity to respond in several modes, including online. It is common for businesses or individuals to prefer the online survey format.

Wisconsin cheese producers respond annually to an industry group survey that reports production by the type of cheese. This more-than-90-page survey details the desired information for a separate type of cheese product on each page. When asked if they would prefer a paper-and-pencil or online survey, more than 50 percent favored the online mode. While the online methodology may be preferred, access to the survey must be provided to all cheese producers, even those without Internet access. A mixed-mode survey design is the obvious choice.

When online samples are used to make inferences about the general population, we must account for the multiple factors that distinguish online samples from the general population. These factors include nonsampling errors unique to the Internet methodology: for example, fewer households have adopted the Internet than have telephone or mail and researchers lack control of the respondent's computer configuration (browser, operating system, fonts, and resolution). In addition, refusals, partial completions, and all other nonsampling factors that bias traditional survey measurement and results still apply to online surveys. (Sources of nonsampling error were discussed in Chapter 2.)

Online survey techniques are also subject to many of the other errors that affect telephone and mail surveys. Marketing researchers, both professional and casual, often neglect to consider the implications that nonprobability sampling and surveys have on the ability to make inferences regarding the target population. While this brief review has done little more than identify the topic areas to be considered, much research on the topic has been completed for both traditional and online surveys. In Chapter 10, we build upon this discussion of general survey and sampling methodology.

STRATEGIES OF DATA COLLECTION

When designing a survey, our concern should be for the total package of survey procedures rather than any single technique (Dillman, 1978, 2000). The total package concept underlying survey design strategy goes beyond contacts with respondents. All aspects of a study must be considered when comparing alternative strategies. This is the essence of a total design. Pretesting and conducting pilot surveys are part of this package.

Pretesting and Pilot Survey

A distinction should be made between a pretest and a pilot survey. Pretesting is an activity related to the development of the questionnaire or measurement instrument to be used in a survey or experiment. In contrast, a pilot survey is a small-scale test of what the survey is to be, including all activities that will go into the final survey. Pretesting a questionnaire answers two broad questions:

1. Are we asking "good" questions?
2. Does the questionnaire flow smoothly, and is the question sequence is logical?

Pretesting does not, however, ensure that the questionnaire (or even the survey) will be valid, particularly in its content. A general rule of thumb for most surveys is that a pretest of about 30 to100 interviews is adequate, provided this covers all subgroups in the main survey population. Ideally, the sample for the pretest should mirror in composition that of the main survey.

The pilot study is designed to ascertain whether all the elements in the survey fit together. Thus, questionnaire pretesting may be part of the pilot study but normally should not be. One aspect of the pilot survey is that it can help researchers decide the size of the original sample for the main survey. Response to the pilot can be used, together with the desired sample size, to determine the size of the required sample.

Both pretesting and pilot surveys can provide information helpful to manage some of the sources of potential research error. Moreover, in the long run they can both make a survey more efficient and effective.

LONGITUDINAL DATA COLLECTION WITH PANELS

Panels are widely used in marketing research. In the preceding chapter we discussed the continuous panel as used by syndicated services. In this chapter we have discussed methods for reducing response error along with some general characteristics of panels. Although the panel concept has been used in business-to-business marketing research, its greatest application has been in studying consumer purchase, consumption and behavioral patterns.

> For example, panels have been effectively used to develop early forecasts of long-run sales of new products. There are major commercial consumer panel organizations, and many consumer product companies that maintain their own panels or create short-term ad hoc panels as the need arises to test new products. In addition, several universities maintain consumer panels to obtain research data and generate revenues by providing data to others. Moreover, the application of Internet technology has helped many corporations to reduce data costs while increasing customer feedback and research.

The distinguishing feature of a panel is the ability to repeat data collection and collect longitudinal data from a sample of respondents. The repeated collection of data from panels creates both opportunities and problems. Panel studies offer at least three advantages over one-time surveys:

1. Deeper analysis of the data is possible so that the researcher can determine if overall change is attributable primarily to a unidirectional shift of the whole sample or only reflects changes in subgroups.
2. Additional measurement precision is gained from matching response from one interview data collection point to another.

3. Panel studies offer flexibility that allows later inquiries to explain earlier findings.

When responses are obtained at two or more times, the researcher assumes that an event happens or can happen (i.e., changes may occur) during the time interval of interest. In fact, it is just such changes, analyzed in the form of a turnover table, that provide the heart of panel analyses.

> Assume that we have changed the package in one market for a brand of paper towels called Wipe, and that we run a survey of 200 people purchasing the product two weeks before the change (*T*1) and a similar measure for the week after (*T*2). The results are shown in Table 4.4. Both (A) and (B) tell us that the gross increase in sales of Wipe over X (this represents all other brands) is 20 units (or 10 percent). However, only the turnover table from the panel in (B) can tell us that 20 former buyers of Wipe switched to X and that 40 former buyers of X switched to Wipe. In those instances where there is experimental manipulation, such as the introduction of a new product or the use of split-cable advertising, the manipulation is presumed to cause changes between time *x* (when the change is made) and time $x + 1$.

TABLE 5.4 Changes in Sales of Wipe Paper Towels between T_1 and T_2 (Hypothetical)

(A) Cross-Sectional	T_1	T_2	
Bought Wipe	100	120	
Bought X	100	80	
Number of Purchasers	200	200	
(B) Panel	**At T_1 Bought Wipe**	**At T_1 Bought X**	**Total**
At T_2 Bought Wipe	80	40	120
At T_2 Bought X	20	60	80
Total	100	100	N=200

Panel studies are a special case of longitudinal research, where respondents are typically conscious of their ongoing part in responding to similar questions over a period of time. This consciousness of continuing participation can lead to panel conditioning, which may bias responses relative to what would be obtained through a cross-sectional study. As in any effort at scientific measurement, the researcher should be concerned with threats to internal validity, since internal validity is a precondition for establishing, with some degree of confidence, the causal relationship between variables. Another issue of concern is panel attrition, the extent of nonresponse that occurs in later waves of study interviewing. Some persons who were interviewed at the first time may be unwilling or unable to be interviewed later on.

There are many distinguishing characteristic of panel types. We have already mentioned different types of sponsoring organizations (such as commercial), permanence (continuous or ad hoc), and research design (non-experiment). Panels can also be characterized by geographic coverage (ranging from national to local), whether a diary is used, data collection method (all types are used), sampling method employed for a given study (probability or not), and type of respondent.

A unique type of panel is the scanning diary panel. This panel involves recruiting shoppers in the target market area to participate in the panel, and each person typically is compensated for participation. An identification card (similar to a credit card) is given to each member of the panel household. At the end of a normal shopping trip in a cooperating store, the card is presented at the start of the checkout process. This identifies the respondent for input of purchases into the computer data bank. The types of information available from this sort of panel are similar to those discussed in the preceding chapter for scanner-based syndicated services. An added advantage here, of course, is that there is a carefully designed sample providing purchase data.

One last comment about panels is that they are often used for a cross-sectional study. When used this way and only one measurement is made, the panel is merely the source of a sample (the sample frame).

SUMMARY

This chapter first examined the media through which interviews may be conducted. The personal interview, the telephone interview, the mail interview and online e-mail interview were discussed, including the merits and limitations of each. Variations of these basic methods, including electronic-based variations, were briefly described. Finally, the use of panels was presented in a general context.

The objective of marketing research is to understand the consumer and apply information and knowledge for mutual benefit. Technological advances in online marketing research provide the ability to monitor customer knowledge, perceptions, and decisions to dynamically generate solutions tailored to customer needs. In this chapter we have stressed the advantages as well as the caveats associated with online research. Perhaps the biggest mistake the market researcher could make would be to view online research as simply a time- and cost-saving extension of traditional modes of data collection. As new technologies continue to be developed, they are tested for applicability in marketing research settings, and refined so that marketers are able to better identify the needs and wants of today’s consumers.

REFERENCES

www.azcentral.com/business/articles/2009/07/20/20090720biz-cellnation0721side.html

Baker, K. G., Hozier, G. C., Jr., & Rogers, R. D. (1999, February). *E-Mail, mail and fax survey research: Response pattern comparisons.* Paper presented at the 6th Annual Meeting of the American Society of Business and Behavioral Sciences, Las Vegas, NV.

Becker, E. E. (1995, January 2). Automated interviewing has advantages. *Marketing News*, *29,* 9.

Bourque, L. B., & Fielder, E. P. (2003a). *How to conduct telephone surveys*. Thousand Oaks, CA: Sage.

Bourque, L. B., & Fielder, E. P. (2003b). *How to conduct self-administered and mail survey*s. Thousand Oaks, CA: Sage.

www.busreslab.com/articles/article3.htm

www.cnn.com/

Davern, M., Rockford, T. H., Sherrod, R., & Campbell, S. (2003). Prepaid monetary incentives and data quality in face-to-face interviews. *Public Opinion Quarterly*, *67*, 139–147.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley-Interscience.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York:Wiley.

Eisenfeld, B. (2003, March 3). Phone interviews may garner more data. *Marketing News*, *37*, 57.

European Society for Opinion and Marketing Research. (1989). *Distinguishing telephone research from telemarketing: ESOMAR guidelines*. Amsterdam: ESOMAR.

Goldstein, K. M., & Jennings, M. K. (2002). The effect of advance letters on cooperation in a list sample telephone survey. *Public Opinion Quarterly*, *66*, 608–617.

Kay, D. (1997, January 6). Go where the consumers are and talk to them. *Marketing News, 31,* 14.

Malhotra, N. K. (2004). *Marketing research: An applied orientation* (4th ed.). Upper Saddle River, NJ: Pearson Education.

www.marketingcharts.com/interactive/home-internet-access-in-us-still-room-for-growth-8280/

Miller, T. W. (2001, September 24). Make the call: Online results are mixed bag. *Marketing News,* 30.

www.nationalgeographic.com/geosurvey/

Oldendick, R. W. (1994). The answering machine generation. *Public Opinion Quarterly, 58,* 264–273.

Rochford, L., & Venable, C. F. (1995, Spring). Surveying a targeted population segment: The effects of endorsement on mail questionnaire response rate. *Journal of Marketing Theory and Practice,* pp. 86–97.

Smith, K. T. (1989, September 11). Most research firms use mall intercepts. *Marketing News, 23,* 16.

Wiseman, F., Schafer, M., & Schafer, R. (1983). An experimental test of the effects of a monetary incentive on cooperation rates and data collection costs in central-location interviewing. *Journal of Marketing Research, 20,* 439–442.

Xu, M., Bates, B. J., & Schweitzer, J. C. (1993). The impact of messages on survey participation in answering machine households. *Public Opinion Quarterly, 57,* 232–237.

Chapter 6
QUALITATIVE RESEARCH AND OBSERVATION

In Chapter 4 we discussed communication as a means of obtaining information from respondents. Respondents provide information by answering questions (via an interview) or by having their behavior observed. It will be recalled that at the extremes, an interview will be either structured or unstructured in organization and that it may be either direct or indirect in questioning. In this chapter we consider the use of indirect interviews of all kinds, special types of unstructured-direct interviews, and observation as means of obtaining information from respondents. We conclude the chapter with an assessment of direct and indirect research techniques.

FOCUS GROUP INTERVIEWS

Perhaps the best-known and most widely used type of indirect interview is that conducted with a focus group. As we mentioned in Chapter 2, a focus group interview involves group of people jointly participating in an interview that does not use a structured question and- answer method to obtain information from these people. A trained moderator conducts the interview with a group of, ideally, 8 to 12 (but increasingly only 6 to 8) willingly recruited participants. The composition of the group varies according to the needs of the client, especially the problem under study. Although the technique is widely used in exploratory research, it also is useful in nonexploratory research. Such applications include gaining greater understanding of consumer wants, thoughts and needs; idea generation (e.g., problem, unmet needs, ideas for new products), concept development and screening, tests for comprehension of promotion and communication materials, and establishment of "opinion leader" panels.

It seems to be a well-accepted fact that focus groups work, especially when used with other techniques. Practitioner researchers accept the idea that qualitative techniques like focus groups, explore, define and describe. In contrast, quantitative methodologies measure, estimate and quantify. These techniques are complements, not substitutes. Typically, the qualitative research precedes quantitative research, but, when a company already has a substantial amount of existing research data to prepare a quantitative survey questionnaire, the focus group can add meaning to the survey findings (Garee & Schori, 1996).

> **Example**: The Federal Duck Stamp Office of the Department of Interior, U.S. Fish and Wildlife Service (Rydholm, 2000) focus group study. Migratory Bird Hunting and Conservation Stamps, known commonly as duck stamps, are required for hunters of ducks. They are also popular among stamp collectors. Funds from the sale of these stamps are used to help fund the preservation of wetlands in the United States. Because the number of hunters has been declining, there was a need to find a way to reach a new audience and to broaden the market for the stamps. Focus groups were used during the development of a marketing campaign in which a certificate bearing a duck stamp and stating that the recipient has helped preserve one-tenth of an acre of wetlands could be purchased for $30. The objective was to appeal to bird watchers, hikers, and other naturalists. Focus groups showed that the stamp itself was not enough to make the sale. The certificate idea came directly from the focus group research. Focus groups were conducted throughout the country, two groups per city—one with environmentally active people, the other with people who were not predisposed against environmental issues. Once a range of appeals

was identified in the groups, the questioning centered on isolating the elements that made people change their minds.

Example: is the Jacksonville, Florida, symphony orchestra's use of focus groups to identify lifestyle marketing issues to explore entertainment alternatives, and to provide some ideas about what future audiences would want and expect from the orchestra (LaFlamme, 1988).

Raymond Johnson (1988) has identified four distinctive categories of focus groups on the basis of examining tapes from the project files of several research companies. Johnson, a practitioner, has defined each type of focus group by the adaptation of an interviewing technique to answer one of four basic research questions. The focus group types are as follows:

- *Exploratory studies* of consumer lifestyles and probing to "just find out what's on the consumers' minds these days."

- *Concept testing studies* of how a group, without prompting, interprets a deliberately sketchy idea for a new product or service. Potential users are able to react to a concept still in its formative or experimental stage.

- *Habits and usage studies* deal with the real world of actual consumers. The topic is framed by the moderator's instructions to describe, usually by situation-specific narratives, the details of personal experiences in using a particular product or service.

- *Media testing* in which participants are asked to interpret the message covered in media usually seen in rough form is the fourth type. All types of media may be covered. Group members talk about their understanding of the message and evaluate the extent to which they find it credible, interesting, and emotionally involving.

The use of focus groups is not limited to consumer products and services. This technique can provide a relatively easy and cost-effective way to interact with business consumers in industries ranging from pharmaceuticals to computer software. The ways in which focus groups are structured and conducted are similar for consumer-based and business-to-business groups, except as identified by Fedder (1990).

A natural question, of course, is, "Why do focus groups work?" One view is that clients are provided with a gut-level grasp of their customers. This means that a sense of what is unique about customers is gained —their self-perceptions, desires, and needs that affect everything they do. For more detailed discussions, see Bloor, Frankland, Thomas, and Robson (2001), and Fern (1981). Qualitative research offers not just an intellectual comprehension of consumers but a vivid, visceral recognition that affects, on a very deep level, how clients see, feel about, and deal with their customers from then on. Some guidelines and questions to assist clients in observing focus groups more effectively are discussed briefly in Exhibit 6.1.

One critical aspect of a focus group's success is the moderator (Exhibit 6.2). The moderator's job is to focus the group's discussion on the topics of interest. To accomplish this, the moderator needs to explain to the group the operational procedures to be used, the group's purpose, and why the group is being recorded. Rapport with the group must be established and each topic introduced. The respondents need to feel relaxed early on, and often moderator humor helps this to happen. Once the respondents are comfortable, the moderator needs to keep the discussion on track while not influencing how the discussion proceeds. A moderator has done a good job if, after the topic has been introduced, the discussion proceeds with little direction from the moderator. There must be synergy among the group members. Thus, a critical measure of success is that the group members have talked with each other, not with the moderator.

EXHIBIT 6.1 Observation of Focus Groups by Clients

As an observer, a client should be cognizant of certain things as he or she observes the dynamics of a focus group in the "back room" from behind a one-way mirror. According to Judith Langer (2001), a client should consider the following:

1. Determine your overall impression of the people in the group by looking at their sophistication level, appearance, and the way they express themselves.
2. Do the respondents' reactions support your assumptions? A way to assess whether the people are atypical or less than fully honest is to have outside data.
3. Are there segments that seem to exist in the focus groups, perhaps based on psychographics?
4. Are there patterns that emerge after several groups? Watch out for making conclusions after just one session. Look for variance! Do not count numbers.
5. A single comment by a respondent may be quite insightful.
6. Look at how people say things, not just what they say. Nonverbal communication can be valuable in interpreting the verbal responses.
7. If a new product or product concept is involved, are the respondents enthusiastic or are they neutral about it?
8. Although certain statements by respondents may appear to be complimentary, they may not really be. This is known as a false positive and is something to be avoided.
9. Be aware of any contradictions that arise between what respondents say and what they report as behavior.
10. Are respondents open to changing their minds, given other information?
11. After the session, talk with the moderator and ask him or her to put the responses in perspective. The last suggestion may be difficult for a client to accept. Never take what is said as being personal. Forget about ego and company politics.
12. Joining the participants of the focus group (after the focus group script is ended). Seeing and hearing consumers, up close, has an impact that no set of data and no written report alone can have. It makes the abstract real because it is human and individual.

When conducting focus groups with professionals, the practical aspects of focus group use are somewhat unique. Special attention must be given to recruitment, type of compensation or gratuity, convenience of the facility to be used, and the moderator. When professionals are involved, the moderator's interaction must be such that he or she is presented as an authority on research in the field of the professional. At least a working knowledge of current technical terminology is necessary.

EXHIBIT 6.2 The Make or Buy Decision for Focus Groups

Some marketing managers believe that almost one-half the cost of the research can be saved by having internal marketing research people conduct the groups. As the costs of doing focus groups increase, some companies consider hiring an "outsider" to conduct groups. In general, a managers' argument for having the groups done internally is based on the following:

1. Money can be saved as the cost of the moderator is eliminated.
2. Company personnel are more familiar with the product and will be able to ask the right questions, recognize the nuances, and know when the group is not telling the truth.
3. The groups can be scheduled as the managers' schedules permit.
4. There have been bad experiences with the outside moderators used in the past.

To an extent, these points are valid and could represent an advantage for doing it internally. However, Thomas Greenbaum (1991) believes that when the total situation is considered, it is preferable to hire the right outsider to conduct the focus groups. He lists the following reasons:

1. An outside moderator can be more objective than an insider, and is less likely to lead the group in a specific direction.
2. Focus group moderation is a learned skill that is developed over time.
3. An outside moderator can be helpful in designing the groups and in developing an effective moderator guide.
4. The lack of detailed product knowledge often is an advantage, not a disadvantage. Since moderators pretend not to know much about the category being discussed, it is possible to ask the seemingly dumb question or seek "help" from the participants that will generate information to achieve the goals of the research.
5. There is less chance that the participants will refrain from showing concern or reacting negatively to ideas when an outside moderator is used.
6. An outsider will be more objective in interpreting results than will an insider.
7. Clients work better by doing what they need to do from the back room, behind the mirror, rather than by conducting the groups themselves.

It seems clear that much of the argument for using outsiders is based on their greater experience and objectivity. Of course, there may be insiders who possess the experience and can, indeed, be objective in doing focus group research. Greenbaum (1999) presents a practical perspective and a more detailed guide of moderators.

In addition to in-person group interaction, a focus group can be conducted over the telephone by use of a conference call. Respondents are recruited from across the country and are told to call a toll-free number at a certain time to participate. Groups have included doctors, car dealers, accountants, travel agents, and others for projects relating to product development, promotion feedback, reasons why a product was not selling, and similar issues. Simon (1988) listed some of the advantages of this approach:

1. Groups can have tremendous geographic diversity.
2. Travel costs can be virtually eliminated.
3. Recruitment is easier because you do not ask a respondent to spend an evening traveling to, sitting in, and returning from a facility.

4. Mixed groups are no problem.
5. Bad weather generally has no effect on the carrying out of a group session.
6. The information from a telephone group is clean, concise, and to the point.
7. Overbearing respondents can be better handled without disrupting the group.
8. Concept testing is easy.
9. Researchers and clients do not have to go all over the country to put together a sufficient number of representative groups from among the smaller sub-specialties.

One disadvantage of the telephone focus group is that the researcher loses the ability to evaluate verbal response within the context of nonverbal communication. More recent technology has allowed clients to view focus groups live without traveling to the geographic areas where they are held. One company has developed video software capable of broadcasting live focus group interviews from a nationwide network of independently-owned focus group facilities to a conference room in a company's own office. Clients view all of the action on a large monitor and control two cameras that allow a full group view, close-up, zoom, or pan. They can maintain audio contact with the moderator, video the complete sessions, and hold "open-mike" post-group debriefings. Major advantages of this system include saving time and money. Such video-system focus-group facilities have expanded to include global options.

Technology has had a dramatic impact on focus group structure and method. Videoconferencing is being used to reduce costs. Related to this is conducting focus groups at conventions, which has application for business-to-business situations and to situations where the target market consists of professionals. With the right convention, a large concentration of the target market can be accessed (Greenbaum, 1993). Focus groups conducted via the Internet can replace face-to-face research. This so-called virtual research has the appeal of groups being run more quickly and at lower cost. But, there is a cost! This cost is that the researcher loses the ability to relate nonverbal communication by respondents to the verbal aspects. That is, body language, voice inflection, facial expression, and interaction between people cannot be observed (Miller, 1994).

INDIRECT INTERVIEWS AND QUALITATIVE RESEARCH

A number of techniques have been devised to obtain information by indirect means. Most of these techniques employ the principle of projection. That is, the subject is given a nonpersonal, ambiguous situation and asked to describe it, expand on it, or build a structure around it. The person giving the description will tend to interpret the situation in terms of his or her own needs, motives, and values. The description therefore involves a projection of the respondent's own personal characteristics to the situation described.

Projection techniques include word association tasks, sentence completion tests, and interpretation of pictures and pictorial representations that have been developed as a means of inducing people to project their feelings (see Table 6.1). These techniques have been most widely used for studies of consumer products that are similar in quality, performance, and price—notably for such products as automobiles, soaps and detergents, gasoline, cigarettes, food products, beverages, and drug sundries. Projection techniques can stimulate a relaxed free flow of associations that tap and identify deep, unacknowledged feelings to a degree not usually possible by other research techniques. Because projective techniques are designed to bypass people's built-in censoring mechanisms, they are useful in eliciting information about sensitive or threatening topics and products.

Table 6.1 Classification of Projective Techniques

Technique	*Response Requested of Subjects*
Construction	To create a story based on a stimuli presented
Ordering Item preference test	To order stimulus items by preference
Expressive techniques	Play a role; draw a picture of a person doing something; describe a character in a simulated situation
Association Word-association test	To reply to a stimulus with the first word, image, or percept that comes to mind
Completion Sentence-completion test	To complete incomplete expressions, images, or situations

Most indirect interviews are at least partially structured in that they are conducted using a predefined set of words, statements, cartoons, pictures, or other representation to which the subject is asked to respond. However, the interviewer is usually allowed considerable freedom in questioning the respondent in order to ensure a full response. Indirect interviews, therefore, are commonly neither fully structured nor fully unstructured; ordinarily they utilize both types of question. Within the marketing research community these techniques constitute qualitative research techniques.

The Third-Person Technique

The simplest way of obtaining information through indirect questioning of a respondent is to ask for the view of a neighbor, an (unnamed) associate, or some other person whose views on the subject at hand might reasonably be known. This permits the respondent to project his or her own views with no feeling of social pressure to give an "acceptable" answer.

An early study using a variation of this technique that has come to be regarded as a classic is the study by Mason Haire (1950) on instant coffee. This study was conducted when instant coffee was first being introduced. The purpose of the study was to determine the motivations of consumers toward instant coffee in general and Nescafe, a brand of instant coffee, in particular. Interviews of consumers had been conducted using a questionnaire employing direct questions. Among the questions asked were "Do you use instant coffee?" and (if "No") "What do you dislike about it?" The majority of the unfavorable responses were of the general content "I don't like the flavor." This answer was suspected to be a stereotype rather than revealing the true reasons. An indirect approach was therefore chosen.

Two shopping lists were prepared that were identical in every respect except that one contained "Nescafe instant coffee" and the other "Maxwell House coffee (drip grind)." These shopping lists were shown alternately to a sample of 100 respondents, each being unaware of the other list. Each subject was given the following instructions:

> Read the shopping list below. Try to project yourself into the situation as far as possible until you can more or less characterize the woman who bought the groceries. Then write a brief description of her personality and character. Wherever possible indicate what factors influenced your judgment.

The results were quite revealing. The descriptions given were summarized as follows (Haire, 1950, p. 652):

- Forty-eight percent of the people described the woman who bought Nescafe as lazy; four percent described the woman who bought Maxwell House as lazy.
- Forty-eight percent of the people described the woman who bought Nescafe as failing to plan household purchases and schedules well; 12 percent described the woman who bought Maxwell House this way.
- Four percent described the Nescafe woman as thrifty; 16 percent described the Maxwell House woman as thrifty; 12 percent described the Nescafe woman as spendthrift; 0 percent described the Maxwell House woman this way.
- Sixteen percent described the Nescafe woman as not a good wife; 0 percent described the Maxwell House woman this way; 4 percent described the Nescafe woman as a good wife; 16 percent described the Maxwell House woman as a good wife.

The implications of these findings seem clear. The woman using the instant coffee was characterized as being lazier, less well organized, more of a spendthrift, and not as good a wife as the one using the conventional coffee. These imputed characteristics must have been the result of the respondents' projecting their own feelings toward instant coffee in their descriptions of the woman using it.

This study has been replicated a number of times. The general acceptance of instant coffee and the change in dietary habits since the original study have resulted in different findings in the more recent studies. Nevertheless, the original study remains as a classic application of the third-person technique of research.

Word Association Tests

Word association tests consist of presenting a series of stimulus words to a respondent who is asked to answer quickly with the first word that comes to mind after hearing each. The respondent, by answering quickly, presumably gives the word that he or she associates most closely with the stimulus word.

Word association tests are simple and easy to use, and offer powerful insights into the perceptions and associations related to the concepts being tested.

Sentence Completion Tests

Sentence completion tests are similar to word association tests, both in concept and in use. A sentence stem (the beginning phrase of a sentence) is read to the respondent, who is asked to complete the sentence quickly and with the first thought that occurs to him or her. Recognizing that people may react in more than one way to a sentence stem, participants are asked to fill in the sentence several times rather than once. This increases the likelihood of uncovering all major feelings on a topic.

Sentence completion tests provide a top-of-mind association between the respondent and the topic/product/subject being investigated. This data is easy to collect, but is difficult to analyze so that an accurate perspective is obtained of the differences between groups and the meanings of the comments.

> Example: A sentence completion was used in a study of automobile buying to probe the motivations of automobile buyers and thereby provide a sounder basis for advertising. Analysis of selected responses of men and women to two of the sentence stems illustrates how inferences of motivational influences can be drawn through the use of this technique (Newman, 1957, pp. 227–228).

Sentence stem: *When you first get a car . . .*	Sentence stem: *A car of your own . . .*
Women's responses:	Women's responses:
. . . you can't wait till you drive.	. . . is a pleasant convenience.
. . . you would go for a ride.	. . . is fine to have.
. . . you would take rides in it, naturally.	. . . is nice to have.
. . . you would put gas in it and go places.	
Men's responses:	Men's responses:
. . . you take good care of it.	. . . I would take care of it.
. . . I want to make darn sure it has a good coat of wax.	. . . is a good thing.
. . . check the engine.	. . . absolutely a necessity.
. . . how soon can I start polishing it.	

Women's responses indicated that for them a car is something to use and that pride of ownership stresses being seen in the car. For men a car was something for which they should be protective and responsible. Their emphasis was on examining the car and doing things to it. Men appeared to feel closer to their car and regarded it as more of a necessity than did women.

The Depth Interview

The unstructured informal interview in marketing research is referred to as a depth interview. It is used to explore the underlying predispositions, needs, desires, feelings, and emotions of the consumer toward products and services. Depth interviews, if conducted in sufficient detail, produce accurate and understandable qualitative information about the research problem (exploratory studies), or for concept or habits and uses studies.

Insofar as obtaining information on motivations is concerned, the concept of "depth" refers to the level at which underlying motivations are uncovered. When conducting depth interviews, the key is to ask questions that probe for more depth in underlying relationships and motivations. Mariampolski (1988) created a useful list of different probes that can be used:

1. The silent probe (use eye contact and body language)
2. Request elaboration
3. Request definition
4. Request word association
5. Request context or situation of use, occurrence or activity
6. Shift the context or situation
7. Request a clarification
8. Request a comparison
9. Request classification or typology
10. Compare and contrast to a previous statement
11. Challenge veracity
12. Challenge completeness
13. Confrontational probe
14. Echo probe
15. Interpretive probe
16. Summary probe
17. Purposive misunderstanding
18. Playing naive
19. Projective probe

These alternative techniques of probing have varied effectiveness, but are all useful tools in interviewing. By following leads and cues provided by respondents, phrasing questions to continue the flow and pattern of the conversation and to maintain the rapport established, the competent interviewer can explore and probe the underlying motivations of the respondent.

Dissatisfaction with the group influence and the high cost of focus groups, together with certain evolving factors in the marketing environment, have led to recent increased use of the individual depth interview. The depth interview is ideal for obtaining from consumers anecdotes of times they used a product or service, and such "stories" provide the marketer with a good view of what products mean to consumers. When used this way, the result of the interview is usually presented verbatim. Telephone interviewing has proved to be effective in obtaining such consumer stories. The depth interview offers insights with depth, whereas focus groups offer information with breadth. Thus, each has its place in the total set of methodologies from which marketing researchers can draw.

There are three distinct research situations where depth interviews can be used (Kates, 2000):

1. Depth interviews are useful in exploratory research to obtain background information to support the development of a quantitative survey instrument.

2. Depth interviews may be the sole research method used such as hard-to-reach groups, or when focus groups are not feasible. Obviously, results normally cannot be statistically significant, but results can be projected if the size of the sample interviewed is large enough (at least 60 percent of the population, indicating, of course, a relatively small population).

3. Depth interviews can obtain information on a subject without being biased by the group dynamic that often occurs in a focus group. For instance, in the case of a rollout of a new product the marketer may not want respondents to be influenced by the views of others.

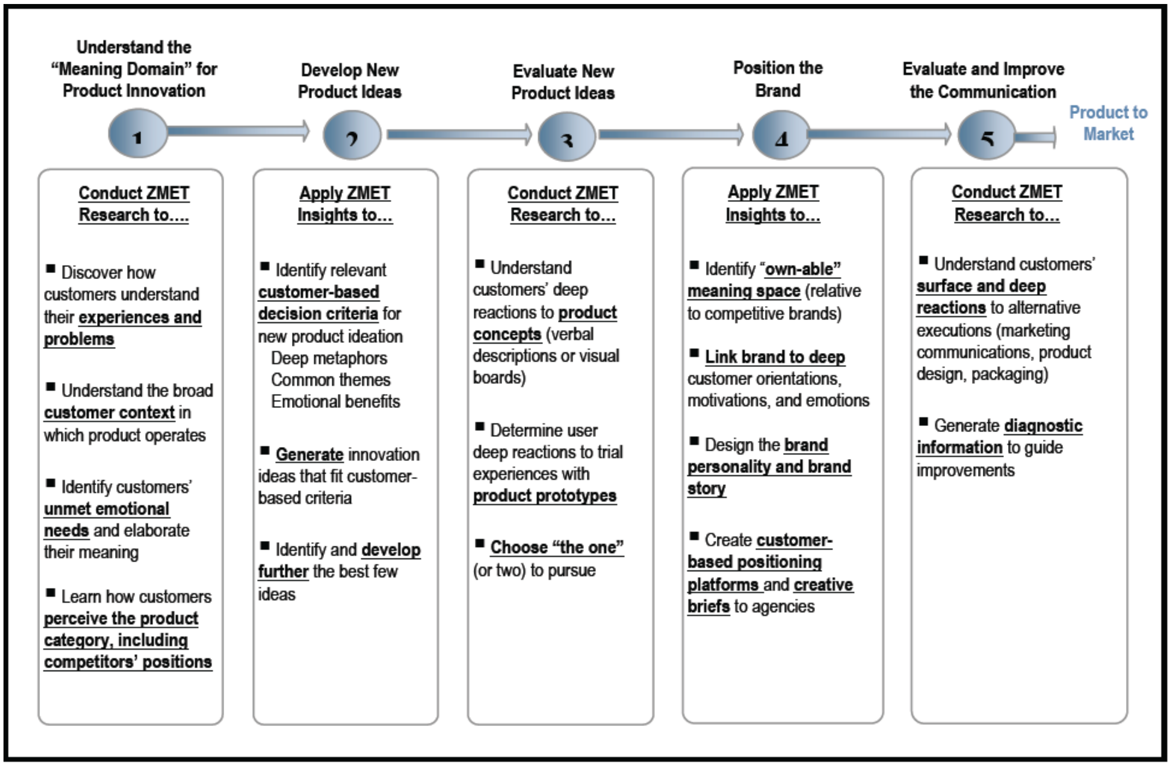
Another approach, labeled ZMET, for Zaltman Metaphor Elicitation Technique was developed by Zaltman (2003). This approach uses a rigorous in-depth interviewing procedure to understand the deep meaning customers attach to product decisions. Founded in the assumption that 95% of human thoughts and feelings are unconscious, we process current information through the context of past experiences to provide their meaning and interpretation to us.

ZMET identifies the metaphors that shape decision making and behavior. For example, we can identify the surface metaphors (statements like "He invests his time wisely"), thematic metaphors (themes: "time is like money") and deep metaphors ("time is a resource") that help us understand and evaluate the thinking and emotions that precede our behavior (what we hear, feel, think, say and do).

The ZMET interview focuses on photos or drawings that subjects are asked to bring with them, representing their feelings about the topic of discussion for example, donating to a charity or university, or purchasing a specific product or brand). These representations are discussed in depth to identify the deep metaphors associated with the topic.

To better understand this result, Olson, Waltersdorff and Forr (2008) provide, in Figure 6.1, a tabular explanation of how a ZMET analysis can benefit the marketing process.

Figure 6.1 Applications of ZMET in Product Marketing



Source: Olson, Waltersdorff and Forr (2008)

Table 6.2 shows examples of these deep metaphors that each of us in expressing thoughts and emotions. Interviewees are asked to explain their true feelings about objects of concern with images, and not just words.

Table 6.2 Deep Metaphors and Expressions in Conversation

DEEP METAPHORS	EXPRESSIONS IN THE VIEW
Physicality	References to bodily functions and senses such as taste it; feel it; pick up, ingest, see my point; hurts me
Orientation	References to spatial orientation--up/down, higher/lower; bigger/smaller; upright/lie down; front/back.
Pleasure/pain	References to the positive (or negative); to things that give one pleasure; feeling good or bad; References to enjoyment, fun, happiness, euphoria, well-being ... or to he opposite (fear, disgust, pain); Feeling good versus hurting, physically or emotionally.
Entity	Considering an intangible idea, concept as a thing, a physical object, an entity; e.g.; "I can't get my breath"
Balance	References to equilibrium, stability, equalize or compensate; Including both sides; Images of scales, teeter-totter, balance beam; References to reciprocity--give and take; References to 'stable" emotional states such as calm, relaxed, serene; Feeling 'right" with the world.
Connection or Linkage	References to connecting to things or people; Making an association; References to linking or attaching; To be a part of; to not be isolated from; Liking or loving someone or something; References to getting in touch with yourself; find your true self.
Resource	References to having/getting the requisite knowledge, energy, tools, or materials to accomplish some task; Having or getting help and assistance from others (we are a team, I need support);

Container	References to being in (or out) of a place (house, room); References to keeping or storing; References to "in" and "out;" Keeping things out as well as in; Being wrapped up ... or out in open.
Motion or Movement	References to moving (flowing, traveling, running or walking); References to movement actions (hurrying, getting going); Keep moving; Keep it going.
Journey	References to taking a trip; Following a path, choosing a direction; Getting there; Journey of life.
Transformation	References to changing from one state to another-physical or emotional; Becoming something or someone else; References to evolving, maturing, growing;
Time	References to the passage of time; Reference to past events or historical perspectives; Images of clocks and watches; References to old memories, remembering past events.
Nature	References to nature, outdoors, natural world; Specific images of nature--rain forest, desert, woods; References to pure, unadulterated, pristine, uncontaminated and to wild, untamed, chaotic, stormy. References to breeding, evolving, growing.
Force	References to power, a powerful presence, or a source of energy; References to the consequences of force (getting hit; slammed, impact)
Fight vs. Flight	References to war; fights, battle, attacking, aggression; Choose your battles; References to weapons; Avoid a fight; Don't get involved; Running away or hiding from something; Ignoring an issue.
Knowing and knowledge	References to knowing, understanding, learning. Gaining knowledge and understanding through study; Knowledge and wisdom; Comprehension; insight. In contrast, the unknown, ignorance, stupidity; Inability to comprehend and understand something.
Contrast or Paradox	References to looking at both/two sides; Comparing opposites; Juxtaposition of opposites; References to paradox--being one thing and its opposite at same time;
Personal expression	Reference to things that express one's personal goals, values, points of view ... to one's self or to others.
The Ideal	Reference to the ideal object, situation, feeling. Statements about one's ideal self. References to perfection, the perfect one.
System	References to machine metaphors (wheels and gears, well-oiled machine); or a constructed process or approach for solving a problem; A set of rules or procedures to accomplish a task; Following a ritual
Sacred (Profane)	References to divine or spiritual qualities; Symbolic cleanliness, purity, vs. dirty, impure, devilish, malevolent
The Mother; Care giving	References to love, fondness; warm relationships with family and friends; Caring for or nurturing others, animals, plants; Being a caring resource for others; taking responsibility for and supporting and helping others
Masculine/ Feminine	References to gender--male or female; Having masculine or feminine traits; Attributions of gender to nonliving things, ideas, concepts.
Birth/Death	References to beginning or end of life, of an idea, or concept; Being rejuvenated or brought back to life (reborn); References to decline, dying, or death.
Quantify/ Measure	References to quantification; Keeping track; Counting; Yardsticks, scales, sextons; Computing amounts and quantities
Holism/ Completeness	References to being whole, entire, complete, not lacking anything, not having weaknesses that compromise any other part of the whole.

Source: Jerry C. Olson, Olson Zaltman Associates (www.olsonzaltman.com).

As a final note for conducting in-depth interviews, it is important to use empathy in understanding and appreciating someone else's beliefs, attitudes, feelings, and behaviors, as well as the social factors that influence their behavior. Many standard qualitative techniques neglect empathy. Specific guidelines for conducting empathic interviews include the following (Lawless, 1999):

1. The researcher needs to imagine himself/herself in the respondent's situation and must listen to the respondent fully.
2. Do not be hindered by the discussion guide; react and improvise as needed.
3. Ask open-ended, non-leading questions that start with how, what, and why.
4. Avoid self-referencing by setting aside thoughts, preconceptions, and interpretations.
5. Challenge generalizations by asking for specific examples.
6. Probe non-judgmentally to understand the person's beliefs, feelings, and behaviors.
7. Let the respondent reveal himself/herself through personal studies.

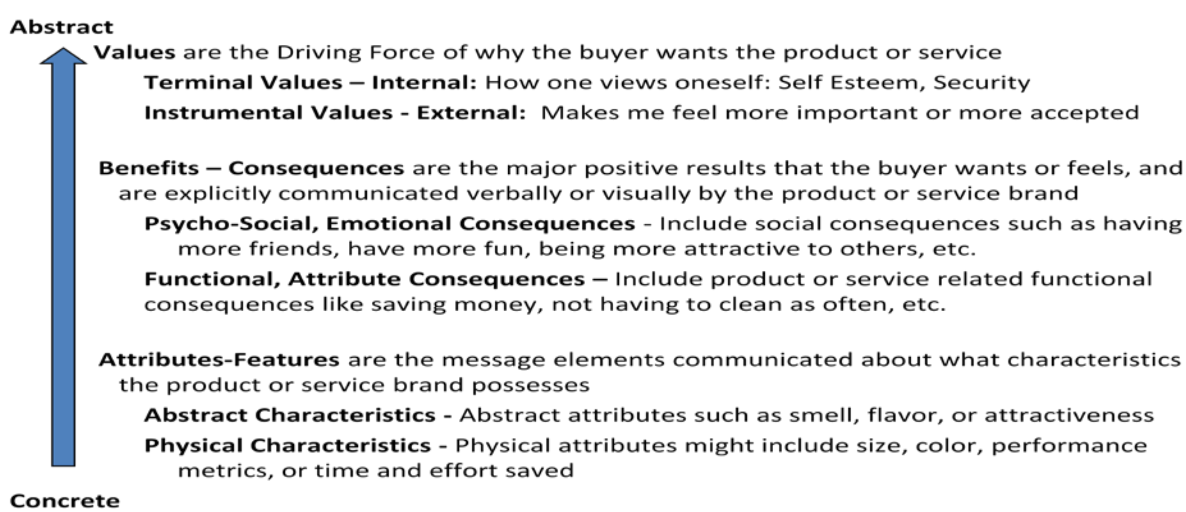
The next section demonstrates a depth interviewing technique called means-end analysis, which focuses on discovering the product attributes/features, benefits and values that motivate the use or purchase of a product.

MEANS-END ANALYSIS

Means-end analysis, also known as Laddering and Means-End Chain, is an in-depth, one-on-one interviewing technique that is directed at identifying the linkages people make between product attributes (the means), the benefits derived from those attributes (the consequences), and the values or end state (the ends) that underlie the desire to achieve the consequences.
The premise of means-end analysis is that consumers make attributed based decisions, selecting those attributes that lead to their desired benefits or consequences and that are consistent with their achieving, fulfilling, or identifying with specific personal values.

Laddering interviews, as an in-depth interviewing technique, employ structured dialogues that identify the most important attributes, the benefits (consequences) derived from those attributes and the values linked to the consequences in a given usage situation (Figure 6.2). The interviewer moves up and down the means-end chain, identifying the hierarchical structure of the product attribute / product benefit / personal benefit / personal value components and linkages. The interviewer usually asks questions like "Why is that important to you?" or "What do you gain from having that characteristic?" Interviews typically last between 45 and 90 minutes and are recorded for analysis and preparation of the resulting laddering maps of the components and linkages.

Figure 6.2 Understanding Consumer Decision Making



Laddering focuses on both the positive and negative linkages between attributes and consequences that are important in choosing a brand. This laddering of the reasons that underlie a decision provides a much deeper understanding of the consumer than does a traditional "ratings survey" of product attributes, but does not penetrate as deeply as does the ZMET analysis.

The first task of a laddering exercise is to elicit the attributes that are important in distinguishing between brands. Exhibit 6.3 identifies a series of approaches that might be used to elicit the attributes that are most important in distinguishing between brands. In practice, several different methods may be used to capture a full range of meaningful distinctions between a brand and its competitors. For example, a laddering interview might start with top-of-mind imaging to understand general product-category beliefs, then increase in brand-related specificity by asking about the usage context (contextual environment), and finally about alternative usage occasions. Other tools, such as a worksheet (Exhibit 6.4) might also be used.

EXHIBIT 6.3 Methods for Eliciting Brand Attitudes

A variety of methods can be used in marketing research to elicit brand attitudes:

- *Top-of-mind imaging.* The respondent gives positive and negative associations for the brand or product category, along with reasons why the characteristic is viewed that way. This line of questions uncovers the attributes and consequences that distinguish the characteristic.

- *Grouping similar brands.* Grouping identifies similar and dissimilar brand groupings within a product category and the reasons for this perceived similarity or dissimilarity. The primary reasons, most important attributes, and most representative brands are identified and attributes and consequences are laddered.

- *Contextual environment.* The usage context for a brand or product can be described either as physical occasions (place, time, people), or need state occasions (relaxing, rejuvenating, building relationships, feeling powerful, reducing stress, and getting organized). A brand or product is associated with a usage context.

- *Preference, usage, similarity and dissimilarity differences*. Comparing brands based on personal preference or usage is commonly used to distinguish between brands. Similarity and dissimilarity groupings also provide a direct method of distinguishing between brands. Questions of why Brand A was grouped differently or ranked higher than Brand B produce elicitations of attributes and consequences.

- *Timing of purchase or consumption*. Timing issues are often related to product or brand choice and usage. For example, a respondent might be asked to identify products used for relief of a stuffy nose into several stages like onset, full-blown, and on-the-mend, or daytime and nighttime. Then the respondent would relate which brands were preferred for each time-related stage.

- *Usage trends.* Dialogues about past and expected future usage of a brand help to elicit attributes and consequences that lead to different usage patterns. For example, respondents may be asked, "Will this brand be used more often, less often, or about the same as you have used it in the past?" Then, reasons for increased, decreased, or unchanged usage are discussed.

- *Product or brand substitution.* Product and brand substitution methods elicit the degree of similarity of perceived attributes and consequences associated with usage. When questions are asked about the degree of substitutability, attributes and consequences are discovered that inhibit or promote substitution (attributes or consequences that need to be added or removed for substitution or trial to occur. The respondent can first sample or be given a brand descriptions, followed by questions like: how likely would you be to substitute (name of the new brand) for your current brand for this occasion—why is that?

- *Alternative usage occasions*. Alternative uses are presented to determine if and why the brand is present or absent from the choice set. Questions might be phrased to ask: why would you consider using Brand A for this occasion, or what is keeping you from using Brand A for this occasion now? Both positive reasons why a brand fits a new occasion and negative reasons why it does not fit can be elicited and laddered. .

Adapted from by Reynolds, Dethloff, and Westberg, 2001, and Reynolds and Whitlark, 1995

One study using the Reynolds and Gutman (1988) approach to analyzing and quantifying the results of laddering exercises, attempted to understand the development of a new bank credit card.

Using the following types of probing questions, the study identified nine critical consumer attributes for a new card: no annual fee, status, low interest rate, added value features, acceptance, credit limit, ability to carry a balance, location of the sponsoring bank, and availability.

- Why do you prefer these cards?
- How do they benefit you?
- How are they different from others?
- What is your least favorite banking or credit card and why?
- What specific things did you dislike about this card?
- Have you had negative experiences that affect your current preferences?
- Under what conditions would you consult these rejected sources for a card?
- How does your preferred card affect your life?
- How does this card make you feel as a consumer?
- What does this preferred card do for you that the rejected cards don't?
- During what period of your life are you most likely to need these services?
- During which period has it most benefited you? Why?
- Describe the best personal experience you have ever had concerning credit cards.
- How has that experience changed your life?
- Do you seek out similar experiences today?
- How would you describe a card that could help you achieve this?

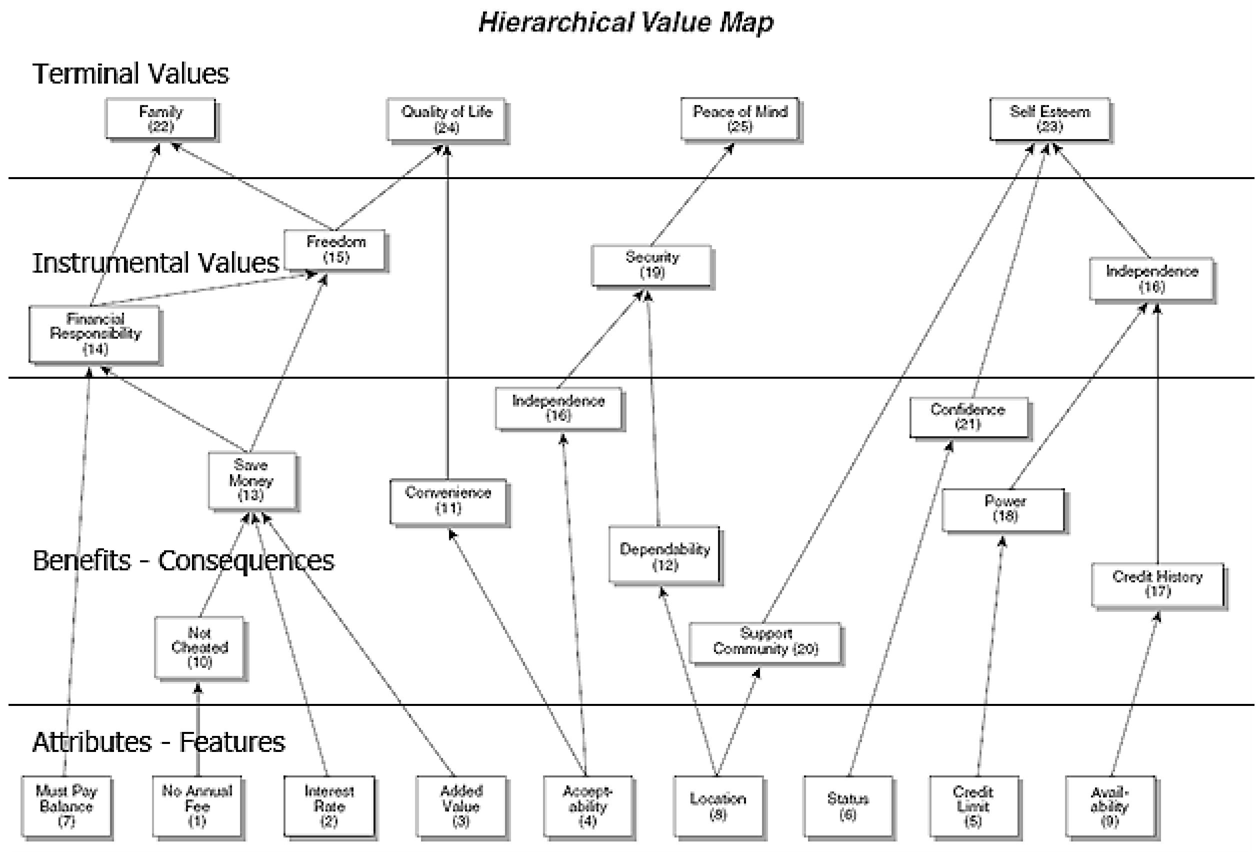
These attributes were found to be linked to 12 benefits (consequences) that were perceived as part of card usage: not feeling cheated, independence, convenience, dependability, saving money, financial responsibility, freedom, establishing a credit history, power, security, supporting the community (local card), and confidence.

Finally the consequences were linked to four personal values: family concerns, improved self-esteem, improved quality of life, and peace of mind. This analysis is actually mapped to show the hierarchy of attribute relationships.

In Figure 6.3, four distinct sub-hierarchies are shown that focus respectively on the four values. Means-end analysis provides a valuable platform upon which media can be developed. In this case, it is easy to imagine four separate commercials, each stressing a different set of attributes and consequences with overall personal value appeals to "Providing for the Family," "Superior Quality of Life," "Peace of Mind," and "Heightened Self-Esteem." Reflection upon the most effective credit card commercials shows that they are well targeted toward these very same appeals.

The advantage of the benefit chain is that it is relatively easy to administer without a highly sophisticated interviewer. The only key decision is when to stop probing.

Figure 6.3 Hierarchical Value Map



SOURCE: Gibby, Holland, Lee, Lundell, Merrill, and Robison, 1995.

OTHER TECHNIQUES

Other variations of projective techniques have been used that are similar in nature but different in form from those already described. The *repertory grid,* for example, is a partially structured technique that requires the respondent to compare objects along dimensions that he or she selects. The usual procedure is to present the interviewee with a pack of cards on which brand names or other stimuli are printed. The respondent is asked to cull unfamiliar brands (stimuli) from the pack, and three cards with familiar brand names (stimuli) are selected at random. Following this, the respondent is asked to describe a way in which any two of the familiar brands (stimuli) are like each other and different from the third. The respondent is then asked to rate all the brands (stimuli) with respect to this dimension. The response may be in the form of a paired comparison, a ranking, or a numerical rating on a scale. This process is repeated using three different brands (stimuli) until the dimensions of the respondent are exhausted. Additional respondents are interviewed until no new dimensions are given. On the average, 40 interviews are required to identify most of the relevant dimensions.

An example of the use of the repertory grid is a study done for a major producer of frozen foods who was interested in introducing a new type of toaster item that could be made to taste like a variety of freshly baked products. The research director was interested in the attributes of bakery items that consumers use to distinguish one product from another. The stimuli were 20 products as shown in Table 6.3. The study resulted in a list of 22 attributes that were scaled as bipolar scales (see Chapter 7) for further analysis, as shown in Table 6.4. The stimuli and attributes can be set up as a grid whereby respondents can then rate each stimulus (i.e., bakery item) on each attribute.

Table 6.3 List of Food-Item Stimuli Used in Repertory-Grid Study

Stimulus	*Stimulus*
1. Toast pop-up	11. Cinnamon bun
2. Buttered toast (white)	12. Danish pastry
3. English muffin and margarine	13. Buttered toast (rye)
4. Jelly donut	14. Chocolate chip cookie
5. Cinnamon toast	15. Glazed donut
6. Blueberry muffin and margarine	16. Coffee cake
7. Hard rolls and butter	17. Apple strudel
8. Toast and marmalade	18. Toasted pound cake
9. Buttered toast and jelly	19. Corn muffin and butter
10. Toast and margarine (white)	20. Bagel and cream cheese

SOURCE: Green, Tull, and Albaum, 1988, p. 712.

Table 6.4 Twenty-Two Bipolar scales Found in a Repertory Grid

1.	Nonfruity flavor	1	2	3	4	5	6	7	Fruity flavor
2.	Easy to prepare	1	2	3	4	5	6	7	Hard to prepare
3.	Low crispness	1	2	3	4	5	6	7	High crispness
4.	Natural flavor	1	2	3	4	5	6	7	Artificial flavor
5.	Dry texture	1	2	3	4	5	6	7	Moist texture
6.	Complex flavor	1	2	3	4	5	6	7	Simple flavor
7.	Complex shape	1	2	3	4	5	6	7	Simple shape
8.	Not very filling	1	2	3	4	5	6	7	Highly filling
9.	Appeals mainly to kids	1	2	3	4	5	6	7	Appeals mainly to adults
10.	Served formally	1	2	3	4	5	6	7	Served informally
11.	Primarily breakfast item	1	2	3	4	5	6	7	Primarily non-breakfast item
12.	Soft texture	1	2	3	4	5	6	7	Hard texture
13.	High perishability	1	2	3	4	5	6	7	Low perishability
14.	Mostly eaten at home	1	2	3	4	5	6	7	Mostly eaten away from home
15.	High calories	1	2	3	4	5	6	7	Low calories
16.	Highly nutritious	1	2	3	4	5	6	7	Low in nutrition
17.	Drab appearance	1	2	3	4	5	6	7	Colorful appearance
18.	Usually eaten alone	1	2	3	4	5	6	7	Usually eaten with other foods
19.	Low general familiarity	1	2	3	4	5	6	7	High general familiarity
20.	Highly liked by men	1	2	3	4	5	6	7	Highly disliked by men
21.	Ordinary-occasion food	1	2	3	4	5	6	7	Special-occasion food
22.	Expensive	1	2	3	4	5	6	7	Inexpensive

SOURCE: Green, Tull, and Albaum, 1988, p. 713.

Exhibit 6.4 discusses the use of *protocols* (Ericsson & Simon, 1984). This technique allows respondents to respond freely without intervention of an interviewer.

EXHIBIT 6.4 Protocols for the Qualitative Research Tool Kit

A protocol is a record of a respondent's verbalized thought processes while performing a decision task or while problem solving. This record is obtained by asking the respondent to "think out loud" or talk about anything going through his or her head while performing the task. Protocols can be collected either in a laboratory situation while the respondent is making a simulated purchase or in the field while an actual purchase decision is being made.

Protocols can be recorded concurrent with the task or retrospective when the verbalizing aloud is done just after the task has been finished.

In contrast to traditional survey methods, protocol methodology allows a person to respond freely in his or her own terms in relation to the actual choice task or decision situation. The form and particular stimuli to which the research subject should respond is not defined or specified by the researcher.

Protocols can be useful in studying brand choice, product categorization, product usage patterns and attitudes, and the impact of shopping environment and situational variables on behavior.

Example: Comcast, a major provider of cable TV, telephone and Internet services offered an unsuccessful self install kit that often resulted in customers calling for a technician to complete the installation. A protocol analysis was conducted when individuals in a laboratory situation were give a self-install kit and asked to complete an install while narrating their activities. Comcast identified the key problem areas in the install process and was able to revise installation instructions and provide an 800 support line, thereby making significant reduction in install service calls.

Story completion, an extension of the sentence completion technique, consists of presenting the beginning of a situational narrative to a respondent, who is asked to complete it. The general underlying principle is that the person will project his or her own psychological interpretation of the situation into the response. For example, the situation could be formulated as follows: "Last weekend my partner and I were deciding which jewelry store to visit for a purchase. When I mentioned XYZ, my partner remembered the last visit there. Now you complete the story."

OBSERVATION

The remaining major method of collecting qualitative information is through observation. Observation is used to obtain information on both current and past behavior of people. Rather than asking respondents about their current behavior, it is often less costly and/or more accurate if the behavior is observed. We clearly cannot observe past behavior, but the results of such behavior are often observable through an approach known as the case study or customer case research. This exploratory qualitative methodology traces the stories of people, circumstances, decisions and events leading to actual purchase decisions through one-on-one interviews (Berstell & Nitterhouse, 2001). The case study approach allows for determining any underlying patterns and may uncover unforeseen problems and unexpected opportunities. Some key characteristics of this approach are (Berstell, 1992):

- Case studies uncover motivations through demonstrated actions, not through opinions.
- Studies are conducted where a product is bought or used.

- Observation and documentation are used to stimulate questions and corroborate responses.
- Case studies can access multiple decision makers because they are done on site.
- Case studies require researchers who are essentially “market detectives” rather than “census takers.” Such detectives must have the skills necessary to continue asking “why” questions until answers emerge that show and explain motivations.

Observation may be used as the sole means of collecting data or, as is frequently the case, it may be used in conjunction with other means. It is a method that should always be considered in designing marketing research investigations that call for information on past or current behavior. In some circumstances, observation is the only means of collecting the data desired. Respondents often cannot and sometimes will not report information accurately. Brands usage reports of well-established brands generally show a “halo effect,” an upward bias reflecting the prestige the respondent associates with the use of the brand. Many companies, for example, have found that respondent reports of brand purchases of products vary widely from the actual brand of product that the consumer has on hand.

Other examples include a food retailer who tested a new type of shelving for canned goods by observing shoppers as they used the new shelves, and a toy manufacturer who, in a laboratory setting, observes children playing with product prototypes. In all these cases other research techniques could have been used, but observation was preferable.

There are some problems in using observation. One concern is with selective perception. Since human perception is quite selective, what people observe depends upon their backgrounds. Thus, what is observed, and reported as such by an observer, depends upon who is doing the observing. The second potential problem is that the behavior being observed is not representative. It may be that what has been observed is the “exception” rather than the rule. That is, the behavior observed may be a unique incident. Of particular concern is whether those being observed know they are being observed. If a person is being observed, his or her behavior may not be a “true” behavior. Also, the presence of others (e.g., in a crowded retail store) may influence behavior at the time. Thus, the situation and setting are critical to the observation experience.

The Audit

Audits of both distributor inventories and consumer purchases are widely conducted to understand purchase patterns. The distributor audit is the more widely known of the two. The commercially available Nielsen Retail Index, an audit of retail stores performed regularly, was described in Chapter 2. As indicated there, data from this and audits available through other research agencies provide estimates of market size, market share, geographic pattern of the market, seasonal purchasing patterns, and results of promotional and pricing changes.

The pantry audit of consumer homes is the second type of audit that is sometimes performed. In this type of audit, the field worker takes an inventory of the brands, quantities, and package sizes that the consumer has on hand. When this type of audit is performed on a recurring basis, inconspicuous labels may be attached to the package showing the date the item was first included in the inventory. When the audit is combined with questioning of the consumer, an estimate of usage may be made. The pantry audit is relatively expensive in terms of data obtained, compared with a self-reporting consumer panel, however. Its use has declined as the use of consumer panels has increased.

Recording Devices

A number of electromechanical devices for "observing" the behavior of respondents are in use in marketing research. Some of these devices are used primarily in laboratory-type investigations and others are used to record behavior in its natural setting. Types of recording instruments used in laboratory studies are the eye/pupilometric camera, and the psychogalvanometer. Three of the devices used in noncontrived situations are the video camera, and the Audimeter.

The "observing" of respondent behavior in a laboratory situation with the aid of recording devices has been largely confined to the pretesting of advertising. Eye/pupilometric cameras, for example, are specially designed cameras that record the pupil size changes and eye movements to specific materials on a page. Subjects are fitted with headgear and software is trained to the individual user. Subjects are then given an advertisement or other print/text media. The computer records and maps their line of eye travel, the ad sections that attract attention initially and sequentially, the relative amounts of time and points of focus in looking at images, and which portions of the copy are actually read, and so on.

Pupil size fluctuations are also recorded. The dilation and restriction of the pupil of the eye has been found to correlate with the degree of interest aroused by the visual stimulus. Interesting or arousing stimuli result in the dilation of the pupil. An advertisement or a product that is of interest to the subject will be evidenced by greater dilation of the pupil, indicating degree of interest.

The psycho galvanometer measures the extent of the subject "responds" to the advertisement. The principle involved is that the perspiration rate of the body is increased by the amount of stimulation provided by an advertisement. This relationship is measured as the change in electrical resistance in the palms of the subject's hands.

Other devices are also used for "observing" behavior under laboratory conditions. In general, all such devices have the advantage of permitting careful and detailed observations of behavior that could not be made otherwise. They have the added advantage of providing permanent records of the behavior observed. In using these devices, however, one should always keep in mind two important questions:

1. Is the behavior we are observing a valid predictor of the behavior we want to predict?

2. Are the subjects behaving as they would in a natural situation?

The answer to the second question can clearly be in the affirmative if the observation is made outside the laboratory and in the natural situation, such as in ethnographic research. Hidden video cameras, for example, are used in many situations to record respondent behavior.

Direct Observation

Direct observation of people and how they behave in situations of interest is a commonly used method of collecting information. Many studies have been made of shopping behavior to determine the relative effects of such variables as displays, availability, and reliance on salesperson advice on the brand selected and the quantity purchased. Supermarkets and department store managers continually rely on observation of traffic flows and length of waiting lines to determine the proper location of the various lines of products and the number and location of salespeople and cash registers. An important consideration in the location of banks, retail stores, and entire shopping centers is the amount and pattern of traffic at alternative sites.

Information obtained from direct observation of purchasers can be highly useful in helping to answer such questions as

- Who actually buys the product?
- Do they appear to be influenced by an accompanying person?
- To what extent do brand choices appear to have been made earlier versus at the point of purchase?
- What proportion of shoppers appears to check prices?
- What proportion of shoppers studies the package before purchase?

Unobtrusive Measures

Observation is the method of data collection underlying a set of methods known as unobtrusive measures. By their very nature these are nonreactive measures. Included are all the types of data collection mentioned in this section.

> One such application was a study performed for a manufacturer of frozen juice concentrates who was considering changing the design and amount of information given on the label. Before this change was made, information was needed on the extent to which consumers actually read information on labels. Hidden cameras were stationed in a sample of supermarkets in front of the frozen food cases, and pictures were taken of consumers selecting frozen juice concentrates. An analysis of these pictures indicated that far more time was spent in the selection and more careful attention given to the label than had previously been believed to be the case. It is not necessary for the camera to be hidden in order for it to be a useful device for recording behavior.

In addition, other types of unobtrusive measures are traces and archives. Regarding traces, studies of garbage can tell much about consumers. Garbage data can be used to examine many aspects of consumption behavior (Reilly, 1984):

1. Brand and product type switching patterns
2. Market share estimation
3. Lifestyle patterns
4. Ethnic and social group differences
5. Media usage patterns
6. Free samples, deal packs, trial sizes, coupons

As an example, consider a “classic” study where alcoholic beverage purchases are tracked in a “dry” community. This unobtrusive research would show which types and brands are purchased and consumed by examining residents’ garbage. In this type of unobtrusive trace research, there are some biases that can arise. While the chose example and explanation focuses on “garbage research”, the biases and advantages outlined below apply more generally to all unobtrusive research. In general, biases may include the following:

> **Missing evidence**. Compost piles, pets, garbage disposals, and recycling all affect estimates of product usage and waste and can lead to inaccurate estimates.
>
> **Incomplete evidence**. Discard of packaging material may not correspond to usage.

Despite these limitations, it appears they are minor relative to the advantages, particularly when the refuse analysis is part of an accepted project. Reilly (1984, 127–128) listed the following advantages of garbage analysis:

- **Unobtrusiveness**. Under the current practices used (random sampling of household refuse within selected neighborhoods), individuals are not aware that their refuse is being analyzed.
- **Nonreactivity**. Because they are not aware that their consumption is being monitored, consumers are unlikely to alter their behavior to appear more rational, more socially acceptable, or more economical.
- **Nonresponse**. There is no selective bias in participation. Problems in estimating the effects of an imperfect sample are not evident.
- **Interviewer effects**. Garbage is coded according to objective standards. There is limited possibility for the recorder to consciously or unconsciously bias the outcome of the analysis.
- **Response effects**. As a result of the unobtrusiveness of the procedure, respondents are not capable of misrepresenting their behavior, either because they can't remember accurately or because they wish to create some type of favorable impression.
- **Longitudinal analysis**. The behavior of the same household can be observed over time. Patterns of brand/type/product switching can be accurately observed.
- **Satisfaction**. Refuse provides an accurate measure of the waste of the product, which is a good indicator of the consumer's liking for the product.
- **Completeness**. Garbage provides information on products that are difficult to monitor through traditional means. Illegal behaviors, purchases from unscanned stores, outlier behaviors (such as beer consumption at parties), and socially sensitive aspects of behavior are all amenable to quantitative analysis using refuse evidence.
- **Consumption**. Garbage analysis includes accurate measures of when products were used, as opposed to measures of when the products were purchased. This makes measurement of stock-up effects and cross-consumption possible.

For a more detailed discussion of unobtrusive methods, we refer the reader to the writings of Webb, Campbell, Schwartz, and Sechrest (1966); Sechrest (1979); and Bouchard (1976). In many cases it is desirable to use these methods in conjunction with other more traditional ones. This is the process known as *triangulation*.

DIRECT VERSUS INDIRECT RESEARCH TECHNIQUES—AN ASSESSMENT

Opinion has been divided among practitioners about the role and relative merits of indirect research techniques in marketing research. This division reflects in marketing research the objectivist-subjectivist debate in the behavioral sciences in general. The controversy has largely centered on three areas:

- The applicability of the techniques
- Sample selection and sizes employed
- Accuracy of utilizing disguised modes of obtaining such information

We will discuss each of these areas.

Applicability of Indirect Research Techniques

The basic premises leading to the use of indirect research techniques are as follows:

1. The criteria employed and the evaluations made in most buying and use decisions have subconscious thoughts and emotions.

2. This subconscious content is an important determinant of what we hear, feel, think, say and do when performing a choice behavior such as buying a product.

3. Such content is not adequately or accurately verbalized by the respondent through direct communicative techniques.

4. Such content is adequately and accurately verbalized by the respondent through indirect communicative techniques.

How valid are these premises? We have already seen that they are valid for some problems. While correct, this is not a very satisfactory answer. It is more useful to review situations in which indirect research information might reasonably be sought from respondents. Four situational categories can be distinguished in which information might be sought from respondents.

Category one is where the information desired is known to the respondent and he or she will give it if asked. Direct questioning will therefore provide all of the needed information in this situation. If the reason a consumer does not buy brand X tires is because he believes they do not wear as well as they should, he will willingly say so given the opportunity.

Category two is where the information desired is known to the respondent, but he or she does not want to divulge it. Matters that are considered to be private in nature, that are believed to be prestige- or status-bearing, or that are perceived as presenting a potential respondent-investigator opinion conflict may not be answered accurately. That otherwise quiet and retiring people sometimes buy powerful cars because it gives them a feeling of superiority on the highway are not reasons that will likely be expressed openly and candidly in response to direct questions. When underlying motivations of this general nature are believed to exist, indirect techniques are well suited to elicit such information.

Third, the information desired is obtainable from the respondent, but he or she is unable to verbalize it directly. When respondents have reasons they are unaware of, such as the association of the use of instant coffee with lack of planning and spendthrift purchasing, properly designed and administered indirect techniques can be highly useful for uncovering such motivations.

Fourth, the information desired is obtainable from the respondent only through inference from observation. In some cases motivations of respondents are so deep-seated that neither direct nor indirect methods of questioning will bring them to the surface.

> An experiment in which the same detergent in three different-colored boxes resulted in the opinion of housewives using them that the detergent in the blue box left clothes dingy, that the one in the yellow box was too harsh, and that the one in the blue-and-yellow box was both gentle and effective in cleaning is an illustration of color association and its effect on assessment of

> product quality that very likely would not have been discovered through either direct or indirect questioning.
>
> In another experiment, orange-scented nylons were placed on a counter in a department store next to identical, but unscented, hosiery. The scented hosiery was bought by approximately 90% of the women making purchases. Questioning of the women who bought the scented hose as to why they preferred the hose they bought resulted in answers such as "of better quality," "sheerer," and the like.

Of these four informational categories, only the second and third lend themselves to the use of indirect research techniques. It remains, of course, for the analyst to decide in which one or more of these categories the information he or she requires will fall.

While neither the universally applicable methodology nor the panacea that some proponents have claimed, indirect research techniques can provide information on some types of marketing problems that is not now obtainable by other means.

Sample Selection and Sizes in Qualitative Research

The subject of sampling is considered in detail in the next chapter. However, it is desirable to examine here the typical sampling procedures and practices that have been used in qualitative research studies, as this has been an area of considerable controversy.

Sample selection in qualitative research studies has tended to be done on nonprobabilistic (purposive) bases rather than by probabilistic methods. Typically, selection has been on a judgment or quota basis. Serious sampling errors can result from purposive sampling and the extent of the sampling error is unknown. There are times, however, when the sample is not statistically representative of the target market, but is close enough to be used as a basis for judgment.

A second area of controversy over the samples typically taken in qualitative research studies relates to their size. Generally, samples have been small, often ranging from 20 to 50 in size. The use of a small sample in a qualitative research study suggests that the population of psychological attributes and motivations being sampled is sufficiently homogeneous that only a limited sample is required to provide an adequate representation of the population. However, the bulk of the evidence amassed by psychologists suggests that motivations are myriad and varied in their effect on behavior. To assume that the motivations of a very small group of people adequately represent those of the population at large is to ignore the high degree of variability that empirical studies have substantiated.

The Validity of the Findings

What about the validity of indirect research findings? How has their performance in these respects compared with that of the more conventional research methods? The question of validity of findings is, of course, the heart of the issue here, as it is in the general objectivist-subjectivist controversy. Unfortunately, to raise the question is to beg it; no definitive answer can be given. As has already been indicated, the answer is necessarily conditional on the nature of the problem being investigated.

An observation does need to be made, however, about the differences in judging validity by the "clients" of basic research versus those of decisional research projects. The client of the basic research project is the professional in the field. Judgments of a study's validity are a highly impersonal process and one that is seldom urgent. When the purpose of the project is either to make the best estimate of a population parameter or to conduct the best test of a hypothesis within the constraints of available resources, the rules of evidence for a basic research study require that the procedures be public, the results investigator-independent, and the project replicable.

Since indirect research methods violate each of these requirements to some extent, there has been reluctance on the part of some basic researchers to give even tentative acceptance to unvalidated findings of studies that employ indirect methods. They tend to look upon indirect research methods as a means of generating hypotheses for testing by objectivist methods rather than as a source of valid findings. In the absence of data that can be used for direct validation, the basic research project is judged tentatively on the basis of method.

The client for a decisional research project has a different set of requirements. Rather than wanting to be assured that the best estimate of a parameter or the most definitive test of a hypothesis has been made, the client needs information that will assist him or her in making the best decision possible in the circumstances. The procedures of the investigation need not be public, and there is seldom a need for replication. The client works directly with the researcher and is able to raise any questions he or she has about the project. The client usually will have had the opportunity to judge the validity of the findings of past research projects conducted by either the researcher or the organization for which he or she works. An assessment of validity of the findings must be made now; for a decision to await the outcome to determine if the findings are valid would obviate the very purpose for which the research was conducted. Judgment of degree of validity therefore turns out to be a much more subjective process in decisional than in basic research.

Indirect techniques serve several useful purposes in marketing research. They can be used to obtain information from respondents unwilling or unable to provide it by direct methods, to check the validity of responses to direct techniques, and to provide supplemental information. Included in the supplemental information that is of value is that which suggests hypotheses that can be tested by direct methods.

SUMMARY

In this chapter we first examined the various types of indirect interviews and qualitative research techniques that can be used to obtain information from respondents. In the indirect types of interviews, we described the more commonly used projective techniques, including the third person technique ("what does your neighbor think of ___?"), word association, sentence completion and depth interviews. Also discussed were focus groups and in-depth interviews, including ZMET and means ends laddering interviews.

We then considered the means of obtaining information through observation of people. The use of audits, recording devices, and direct observation were described and their applications discussed. Of particular interest is the use of trace analysis (analysis of post-consumption garbage) to study behavior.

Finally, an assessment was made of direct (discussed in Chapters 4) versus indirect research techniques from the standpoints of applicability to marketing problems, sample selection and sizes, and validity of findings.

REFERENCES

Berstell, G. (1992, January 6). Study what people do, not what they say. *Marketing News, 26,* 15ff.

Berstell, G., & Nitterhouse, D. (2001, February 26). Getting whole story can reveal surprise opportunities. *Marketing News, 35,* 45.

Bloor, M., Frankland, J., Thomas, M., & Robson, K. (2001). *Focus groups in social research.* Thousand Oaks, CA: Sage.

Bouchard, T. J., Jr. (1976, February). Unobtrusive measures: An inventory of uses. *Sociological Methods & Research, 4,* 267–300.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports of data.* Cambridge: MIT Press.

Fedder, C. J. (1990, January 8). Some differences between consumer-based and business-to-business groups require a special touch. *Marketing News, 24,* 46.

Fern, E. (1981). Why do focus groups work? A review and integration of small group process theories. *Proceedings of the Association for Consumer Research Annual Conference,* pp. 441–451.

Garee, M. L., & Schori, T. R. (1996, September 23). Focus groups illuminate quantitative research. *Marketing News, 30,* 41.

Gibby, K., Holland, D., Lee, F., Lundell, C., Merrill, S., & Robison, J. (1995). *Positioning a credit union Visa.* (Field study conducted under the supervision of Professor Scott M. Smith.) Provo, UT: Brigham Young University.

Green, P. E., Tull, D. S., & Albaum, G. (1988). *Research for marketing decisions* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Greenbaum, T. (1993, January 4). Convention focus groups are a way to increase geographic representation. *Marketing News, 27,* FG-2.

Haire, M. (1950, April). Projective techniques in marketing research. *Journal of Marketing, 14,* 649–656.

Johnson, R. R. (1988, October 24). Focus groups are divided into 4 distinctive categories. *Marketing News, 22,* 21.

LaFlamme, T. A. (1988, August 29). Symphony strikes a note for research as it prepares to launch a new season. *Marketing News, 22,* 12.

Lawless, P. (1999, January 4). Empathic interviewing for insightful results. *Marketing News, 33,* 20

Mariampolski, H. (1988, October 24). Probing correctly uncovers truth behind answers in focus group. *Marketing News, 22,* 22ff

Newman, J. W. (1957). *Motivation research and marketing management.* Cambridge, MA: Harvard University Graduate School of Business Administration.

Olson, Jerry C., Kristin Waltersdorff and James Forr (2008), Incorporating Deep Customer Insights in the Innovation Process, in Hans H. Hinterhuber and K. Matzler (eds.), Kundenorientierte Unternehmensführung, 6th edition, Gabler Verlag: Wiesbaden, 2008.
http://www.olsonzaltman.com/downloads/2008 Insights in the Innovation Process.pdf

Jerry C. Olson, Deep Metaphors and Expressions in Conversation, Olson Zaltman Associates (www.olsonzaltman.com).

Reilly, M. D. (1984, September/October). Household refuse analysis and market research. *American Behavioral Science, 28,* 115–128.

Reynolds, T. J., & Gutman, J. (1988, February/ March). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research, 28,* 11–31.

Reynolds, T. J., Dethloff, C., & Westberg, S. (2001). Advancements in laddering. In T. J. Reynolds & J. C. Olson, Eds., *Understanding Consumer Decision Making: A Means-End Approach to Marketing and Advertising Strategy.* Mahwah, NJ: Lawrence Erlbaum.

Reynolds, T. J., & Whitlark, D. (1995, July/August). Applying laddering data to communications strategy and advertising practice. *Journal of Advertising Research, 35*(4), 9–17.

Rydholm, J. (2000, March). Preserving the preservationists. *Quirk's Marketing Research Review,* pp. 8–19ff.

Sechrest, L., Ed. (1979). *Unobtrusive measurement today.* San Francisco: Jossey-Bass

Simon, M. (1988, August 29). Focus groups by phone: Better way to research health care. *Marketing News, 22,* 47–48.

Webb, E., Campbell, D., Schwartz, R., & Sechrest, L. (1966). *Unobtrusive methods: Nonreactive research in the social sciences.* Chicago: Rand McNally.

Zaltman, G. (2003). *How consumers think.* Boston: Harvard Business School Press.

Chapter 7
SAMPLING PROCEDURES IN RESEARCH

Researchers must answer many questions when developing a project to collect information about a population of interest. Consider the following questions involving sampling:

- Should we take a census (complete canvas) or a sample?
- What kind of sample should be taken?
- What size should the sample be?

Answering these questions depends upon the application of statistical inference. This chapter first considers the selection of sample types and the process of sample planning. It then describes different kinds of samples. Finally, related to design, but a unique decision area in its own right, is the determination of the size of the sample.

PLANNING THE SAMPLE

Two broad objectives are fundamental to the use of samples in survey research projects:

- Estimation of information about a population based on a sample from the population,
- Hypothesis testing to compare the relationships between data items for some selected population groups.

Each involves making inferences about a population on the basis of information from a sample. The precision and accuracy of project results are affected by the manner in which the sample has been chosen. However, as Exhibit 7.1 illustrates, precision is only a reflection of sampling error and confidence limits and has nothing to do with accuracy.

Exhibit 7.1 Precision versus Accuracy in Sampling

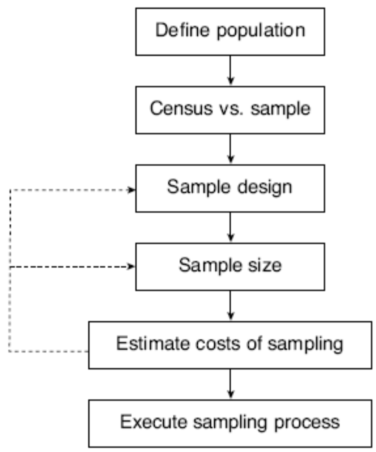
There often is confusion between precision and accuracy in marketing research. When a researcher speaks of sampling error and the size of the confidence limits placed on an estimate, that researcher is speaking of precision and not accuracy.

Accuracy is affected by nonresponse bias, memory error, misunderstanding of questions, problematic definition of terms, and processing errors, and not by sample size (Semon, 2000). The researcher must constantly ask: "Have we asked the right questions?" and "Have we measured what we want to measure properly and correctly?"

The difference between precision and accuracy is analogous to the difference between errors of degree and errors of kind. Accuracy, like errors of kind, are clearly more serious. If a researcher has made the correct measurements, precision will show the degree of error; but if the right measurements have not been taken, precision is probably meaningless (Semon, 2000). The quality of a sample depends upon the quality of design and execution of a research project at every stage of the process.

Consequently, strict attention must be paid to the planning of the sample. It must also be recognized that sample planning is only one part of planning the total research project. The process of selecting a sample follows the well-defined progression of steps shown in Figure 7.1, and will be discussed in turn.

Figure 7.1 Steps in Sample Planning



Defining the Population

The first step in sample panning is to define the population to be investigated. A population, also known as a universe, is defined as the totality of all units or elements (individuals, households, organizations, etc.) to which one desires to generalize study results. While seemingly an easy task, an imprecise research problem definition often leads to an imprecise population definition.

Specifying a population involves identifying which elements (in terms of kind) are included, as well as where and when. For example, a group medical practice that is considering expanding into sports medicine might acquire information from any or all of the distinct population groups listed in Table 7.1. The population element is the unit of analysis, and may be defined as an individual, household, institution, patient visit, and so on.

Table 7.1 Possible Population Choices for a Research Study by a Medical Practice

Group Members	*Where*	*When*
All patients	Designated group practice	Last 12 months
Patients who have had orthopedic work	Designated group practices	Last 12 months
All people	Specified geographic area	Last 12 months
All people who have had orthopedic work	Specified geographic area	Last 12 months

The second and third columns of Table 7.1 define population choices in terms of its location and timeframe. One useful approach is to first define the ideal population to meet study objectives, and then apply practical constraints to ultimately limit and define the study population. However over defining the population should be avoided unless it is absolutely necessary. Over defining can limit the extent to which findings can be generalized, and greatly increase the operational cost and difficulty of finding population elements (Sudman, 1976).

Census, or Sample?

Once the population has been defined, the investigator must decide whether to conduct the survey among all members of the population, or only a sample subset of the population. The desirability and advantages of using a sample rather than a census depend on a variety of factors such as geographic location, the absolute size of the population, and the sample size required for results sufficiently accurate and precise to achieve the required purposes.

Two major advantages of using a sample rather than a census are speed and timeliness. A survey based on a sample takes much less time to complete than one based on a census. Frequently, the use of a sample results is a notable economy of time, money and effort, especially when a census requires hiring, training and supervising many people.

In other situations, a sample is necessary because of the destructive nature of the measurement, such as in product testing. There is a related problem over surveying human populations when many different surveys need to be conducted on the same population within a relatively short period of time. Nonprobability sampling techniques explicitly protect against this problem.

In still other situations, a sample may control non-sampling errors. Samples (a smaller number of interviews compared to a census) may result in better interviewing, higher response rates through more call backs, and better measurement in general. The total amount of sampling and non-sampling error of a sample may actually be than the non-sampling error alone would be for a census. In Chapter 2, we emphasized the importance of minimizing total error.

However, under certain conditions a census may be preferable to a sample. When the population is small, the variance in the characteristic being measured is high, the cost of error is high, or the fixed costs of sampling are high, sampling may not be useful. In addition, if the characteristic or attribute of interest occurs rarely in the population, then a census of a tightly defined population might be desirable (for example, people with a rare genetic disorder). In this case, it would be necessary to sample a relatively large proportion of the general population to provide statistically reliable information. Obviously, the practicality of this depends upon the absolute size of the population and the occurrence rate for the characteristic of interest.

Sample Design

> *Sample design is "the theoretical basis and the practical means by which data are collected so that the characteristics of a population can be inferred with known estimates of error" (Survey Sampling, Inc.).*

Operationally, sample design is at the heart of sample planning. Sample design specification, including the method of selecting individual sample members, involves both theoretical and practical considerations (such as cost, time, labor involved, organization). The following checklist is suggested to obtain a sample that represents the target population (Fink, 2003):

1. Are the survey objectives stated precisely?
2. Are the eligibility criteria for survey respondents or experimental subjects clear and definite? Exclusion criteria rule out certain people.
3. Are rigorous sampling methods chosen? This involves selecting an appropriate probability or nonprobability sampling method.
4. Further questions to be answered in this section include:

- What type of sample should be used?
- What is the appropriate sampling unit?
- What is the appropriate frame (that is, list of sampling units from which the sample is to be drawn) for the particular design and unit decided upon?
- How are refusals and nonresponse to be handled?

Type of Sample

Much of the sampling in the behavioral sciences and in marketing research, by its nature, involves samples are selected by the judgment of the investigator, convenience, or other nonprobabilistic (nonrandom) processes. In contrast, probability samples offer the promise of bias-free selection of sample units and permits the measurement of sampling error. Nonprobability samples offer neither of these features. In nonprobability sampling one must rely on the expertise of the person taking the sample has selected respondents in an unbiased manner, whereas in probability sampling the sampling of respondents is independent of the investigator.

> **Example:** A dog food manufacturer tested consumer reactions to a new dog food by giving product samples to employees who own dogs and eliciting their responses about a week later. The employees' dogs liked the food and the pet food manufacturer actually introduced the new dog food product. However when it hit the market the product was a flop… dogs simply would not eat it. As managers scrambled to find out what went wrong, research showed that employees were so loyal to the company's products that their dogs had little variety and would eat anything for a change. In the broader market dogs were used to a greater variety of dog foods including table scraps and just did not like the taste of the new dog food. In this case, a biased sample was erroneously assumed to conform to the general population of dogs and dog owners.

A researcher choosing between probability and nonprobability sampling implicitly chooses the probability sample's relative size of sampling error against the nonprobability sample's combined sampling error and selection bias. For a given cost, one can usually select a larger nonprobability sample than probability sample, meaning that the sampling error should be lower in the nonprobability sample, but that the nonrandom process used for selecting the sample may have introduced a selection bias.

The Sampling Unit

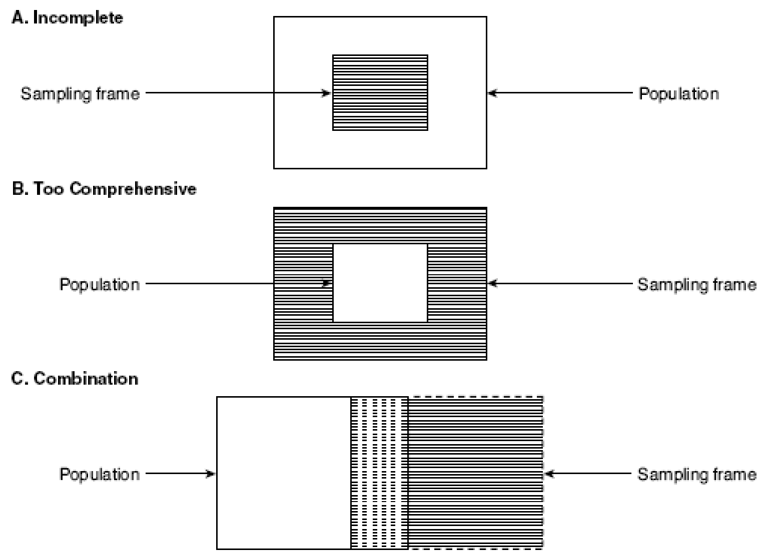
The sampling unit is the unit of the population to actually be chosen during the sampling process. The sampling unit may contain one or more elements describing the population. For instance, a group medical practice may be interested in surveying past patient behavior of the male wage earner *or* his entire household. In either case, it may be preferable to select a sample of households as sampling units.

The Sampling Frame

A sampling frame is a means of identifying, assessing and selecting the elements in the population. The sampling frame usually is a physical listing of the population elements. In those instances where such a listing is not available, the frame is a procedure producing a result equivalent to a physical listing. For instance, in a consumer survey where personal interviews are conducted with a mall intercept method is used to obtain data, the sample frame includes all those people who enter the mall during the study period.

Ideally, the sample frame should identify each population element once, but only once, and not include elements not in the defined population. Such a perfect frame is seldom available for marketing research purposes. As shown in Figure 7.2, a sampling frame may be incomplete, too comprehensive, or a combination of both. In addition, the frame may include individual population elements more than once. Any of these situations can lead to coverage error.

Figure 7.2 Sampling Frame – Population Relationship



Perhaps the most widely used frame in survey research for sampling human populations is the telephone directory. Use of such a frame, however, may lead to frame error from the exclusion of non-subscribers (no home phone or cell phone only), voluntarily unlisted subscribers, and involuntarily unlisted subscribers. For example, more than one in five American homes (20.2%) has only wireless telephones (National Center for Health Statistics, 2009). Combined with the fact that in many major cities more than 50% of home phones are unlisted, the potential for frame errors is huge. Similar frame errors can arise from such other sampling frames as city directories, maps, trade association membership lists, or any other incomplete or dated listing or representation of a population. The National Do Not Call Registry provisions do not cover calls from political organizations, charities, telephone surveyors, or companies with which a consumer has an existing business relationship, nonetheless, resident push back leads to further frame error.

Even though today's panel management software makes compiling and managing sample lists relatively easy, sampling is the single most important problem in e-mail and Internet-based surveys. Dillman (2000) suggests asking the following five questions about any potential sampling list:

- Does the list contain everyone in the survey population?
- Does the list include names of people who are not in the study population?
- How is the list maintained and updated?
- Are the same sample units included on the list more than once?
- Does the list contain other respondent information that can be used to customize the survey or direct survey logic?

Whether the researcher uses their own list, or one purchased from a commercial source, these questions are equally applicable.

Sample Size

Sample size, as part of sample planning is related to precision. Four traditional approaches can determine the appropriate sample size for any given research project:

1. Arbitrarily or judgmentally determined
2. Minimum cell size needed for analysis
3. Budget-based
4. Specifying desired precision in advance of sampling and then applying the appropriate standard error formula to determine the sample size.

For approaches 1-3, the precision could be measured after compiling the data and then applying the appropriate standard error formula or formulas if a probability design is being used.

Costs of Sampling

The sample plan must account for the estimated costs of sampling. Such costs include **overhead costs**, which are relatively fixed for a sampling procedure, and **variable costs**, which depend on the number of respondents contacted and interviewed in the study. In reality, costs from all aspects of data collection are usually considered together. The dashed line shown in Figure 7.1 when the sample design and sample size are considered, the estimated costs may be so great that other sample designs and/or smaller-sized samples may be considered.

Execution of the Sampling Process

The last step in sample planning is to execute the sampling process. A sample is chose that is thought to be both representative and to adequately mirror the various patterns and subclasses of the population (Exhibit 7.2). The sample should be of sufficient size to provide confidence in the stability of its characteristics.

Exhibit 7.2 Samples Are Not Always an Exact Match

Many researchers feel that the best way to assess the validity of a sample is to compare its demographic profile (i.e., distributions of the key demographic characteristics) with a national or otherwise known profile. This alone does not necessarily guarantee a good sample. At the same time, a poor fit does not necessarily mean that the obtained sample is bad.

When the fit is not good, some researchers tend to develop a weighting scheme. However many studies have shown that large differences in demographic characteristics may to translate into small differences in the variable of interest, whether it is a behavior or an attitudes, interests, or opinions (AIO) measure. Of course, there could be serious distortion if specific segments are of concern. But the differences generally have to be much greater than we would think for a significant effect.

This, in turn, requires a measure of precision, which necessitates using probability-based design. From this discussion, one might conclude that an ideal sample is obtained with a probability process. In general, this is preferable. However, recognize that it is more important to avoid distorted samples than to be able to measure sampling error. There may be a tendency to ignore the existence of potential bias when using probability designs.

This chapter is primarily concerned with probability sampling, but before discussing this topic it is useful to describe some of the procedures by which nonprobability samples are taken in marketing research.

NONPROBABILITY SAMPLING PROCEDURES

Nonprobability sampling is distinguished from probability sampling in that nonprobability sample elements do not have a known, nonzero chance of being selected for the sample. As such, sampling error is generally not able to be measured. Nonprobability samples are widely used in exploratory research, but are also valuable for non-exploratory research.

The Quota Sample

The quota sample is the most commonly employed nonprobability sampling procedure in marketing research. Quota samples are collected to reflect proportions in the various subclasses (or strata) of the population of interest. This might be, for example, the proportion of the adult population who fall into various age-by-gender-by-education groupings.

To prepare a quota sample, the subclass proportions are first estimated from some outside source, such as census data. Next, if an interviewer has a total number of, say, 600 interviews to obtain, the age-gender-education proportions in the population are applied to the 600 total interviews to determine the appropriate quotas.

However, in quota sampling the interviewer is *not* required to *randomly* select the respondents necessary to fill each quota category. The lack of random selection is the major distinction between quota sampling and stratified random sampling.

In quota sampling, the interviewer's judgment is relied upon to select actual respondents within each quota. Therefore, many sources of selection bias are potentially present. For example, the interviewer may not bother to call back if the first call results in a "not-at-home." Interviewers may go to selected areas where the chances are good that a particular type of respondent is available. Certain houses may be skipped because the interviewer does not like the appearance of the property, and certain people may be allowed to pass in a mall-intercept approach because they do not "look right." Still other interviewer habits and biases exist that can influence their selection of respondents within the quota.

The advantages of quota sampling are the lower costs and greater convenience provided to the interviewer when selecting respondents to fill each quota. Quota sampling is quite close to traditional probability sampling when selection is tightly controlled (Sudman, 1976).

Quota sampling is discussed rather extensively in Stephan and McCarthy (1958). A critique of the method is offered by Deming (1960). A comparison of quota and probability sampling was the subject of a seminar held in 1994 at the Survey Methods Centre in the United Kingdom (Survey Methods Centre, 1995).

The Judgment Sample

A somewhat representative sample may be provided through the use of purposive, or judgment, sampling. Sound judgment or expertise and an appropriate strategy, leads one to carefully and consciously choose the elements so as to develop a suitable sample. The intent is to select respondents representative of the population in such a way that errors of judgment in the selection will cancel each other out. The relative advantages of judgment sampling are that it is inexpensive, convenient to use, less time-consuming, and provides results as good as probability sampling. One weakness of this approach is that without an objective basis for making the judgments, there is no way of knowing whether the so-called typical cases are, in fact, typical.

The Convenience Sample

Convenience sampling is a generic term that covers a wide variety of ad hoc procedures for selecting respondents. Convenience sampling means that the sampling units are accessible, convenient and easy to measure, cooperative, or articulate. An illustration of convenience sampling is "intercept" interviews among shopping-mall customers or in other areas where large numbers of consumers may congregate. In this case, the researcher needs to select respondents with care. People frequent malls for different reasons and stay in malls different lengths of time, which leads to a biasing effect if a probability sample is attempted. A sample is said to be *length biased* if the probability of observing an individual at a particular site is dependent on the individual's length of stay at the site.

Firms may also authorize samples to be taken from such intact groups as Parent-Teacher Associations, church groups, philanthropic organizations, and so on. Again, the purpose is to obtain a relatively large number of interviews quickly from a cooperating group of respondents. Usually the sponsoring organization receives a donation from the interviewing firm for the help and cooperation of its members. However depending on the purpose of the research, many potential sources of selection bias may exist in that only certain members may respond, who may be disproportionately different on one of many demographic, attitudinal or behavioral dimensions.

Snowball Sampling

Snowball sampling (also known as multiplicity or chain-referral sampling) is the rather colorful name given to a procedure in which initial respondents are selected randomly, but where additional respondents are then obtained from referrals or other information provided by the initial respondents. One major advantage of snowball sampling is the ability to estimate various characteristics that are rare in the total population. A more complete discussion of this technique is given by Zinkhan, Burton, and Wallendorf (1983).

> **Example**: A study of international tourism, required researchers to interview respondents in the United Kingdom, France, and Germany who visited the United States in its bicentennial year. Given the likelihood of finding a qualified adult respondent was less than two percent, stratified probability methods were used to select initial respondents. A referral procedure (up to two referrals per qualified respondent) was then used to obtain a second group of qualified respondents. (However, this particular study did not obtain subsequent referrals from this second group of respondents.)

In other types of snowball sampling, referrals from referrals are obtained, and so on, thus leading to the "snowballing" effect. Even though a probability-based procedure may be used to select the initial group of respondents, the overall sample is a nonprobability sample.

Major advantages of this type of sampling over conventional methods are the substantially increased probability of finding the desired characteristic in the population, and lower sampling variance and costs.

PROBABILITY SAMPLING PROCEDURES

The best-known type of probability sample is no doubt the simple random sample. However, many occasions in marketing research require more specialized sampling procedures than those of simple random-sampling methods. Statisticians have developed a variety of specialized probability-sampling designs that, although derived from simple random-sampling principles, can be used to reduce sampling error and cost. Five major modifications can be made to the basic selection process, as shown in Table 7.2.

These techniques are discussed in turn, following a review of simple random sampling. Our purpose is to describe the major characteristics of each technique, rather than to present a detailed mathematical exposition of its procedures. Many excellent statistics and research books review the mathematical aspects of these sampling techniques.

The Simple Random Sample

In a simple random sample, each sample element has a known and equal probability of selection, and each possible sample of *n* elements has a known and equal probability of being the sample actually selected. It is drawn by a random procedure from a sample frame—a list containing an exclusive and exhaustive enumeration of all sample elements. One widely used process for generating a simple random sample is: upload the elements of the sample frame to a spreadsheet, number them, and then use the random-number generator function to select the sample members.

Simple random samples are not widely used in consumer research, for two reasons. First, it is often difficult to obtain a sampling frame that will permit a simple random sample to be drawn and second, one may not want to give all sample units an equal probability of being selected. Consumer research usually requires people, households, stores, or areas to be the basic sampling units. While a complete representation of areas is available through maps, there normally is no complete listing of persons, the households in which they live, or the stores available. When persons, households, or stores are to be sampled, some other sample design must be used.

In business-to-business (B2B) research, there is a greater opportunity to apply simple random sampling. In this case purchasing agents, companies, or areas are the usual sampling units and the population under study is often relatively small. One is therefore in a better position to develop a complete list of respondents or sample frame.

Table 7.2 Selection Methods for Probability Samples

Probability Samples	*Nonprobability Samples*
I. *Equal probability* for all elements a. Equal probabilities at all stages b. Equal overall probabilities for all elements obtained through compensating unequal probabilities at several stages	*Unequal probabilities* for different elements; ordinarily compensated with inverse weights a. Caused by irregularities in selection frames and procedures b. Disproportionate allocation designed for optimum allocation
II. *Element Sampling*: single stage, sampling	*Cluster Sampling*: sampling units are clusters of elements

unit contains only one element	a. One-stage cluster sampling b. Sub-sampling or multistage sampling c. Equal clusters d. Unequal clusters
III. *Unstratified Selection*: sampling units selected from entire population	*Stratified Sampling*: separated selections from partitions, or strata, of population
IV. *Random Selection* of individual sampling units from entire stratum or population	*Systematic Selection* of sampling units with selection interval applied to list
V. *One-Phase Sampling*: final sample selected directly from entire population	*Two-Phase (or Double) Sampling*: final sample selected from first-phase sample, which obtains information for stratification or estimation

SOURCE: Adapted from Kish, 1965, p. 20.

The Systematic Sample

Systematic sampling involves only a slight variation from simple random sampling. In a systematic sample, each sample element has a known and equal probability of selection. The permissible samples of size *n* that are possible to draw each have a known and equal probability of selection, while the remaining samples of size *n* have zero probability of being selected.

The mechanics of taking a systematic sample are rather simple. If the population contains *N* ordered elements and a sample size *n* is desired, one merely finds the ratio of *N*/*n* and rounds to the nearest integer to obtain the sampling interval. For example, if there are 600 members of the population and one desires a sample of 60, the sampling interval is 10. A random number is then selected between 1 and 10, inclusively; suppose the number turns out to be 4. The analyst then takes as the sample elements 4, 14, 24, and so on.

Essentially, systematic sampling assumes that population elements are ordered in some fashion—names in a telephone directory, a card index file, or the like. Some types of ordering, such as an alphabetical listing, will usually be uncorrelated with the characteristic (say, income level) being investigated. In other instances, the ordering may be related to the characteristic under study, as when a customer list is arranged in decreasing order of annual purchase volume.

If the arrangement of the elements of the sample is itself random with regard to the characteristic under study, systematic sampling will tend to give results close to those provided by simple random sampling. We say “close” because in systematic sampling, all combinations of the characteristic do not have the same chance of being included. For example, it is clear that in the preceding example, the fifth, sixth, and so on items have zero chance of being chosen in the particular sample after the first item has been determined.

Systematic sampling may increase the sample’s representativeness when items are ordered with regard to the characteristic of interest. For example, if the analyst is sampling a customer group ordered by decreasing purchase volume, a systematic sample will be sure to contain both high- and low-volume customers. On the other hand, the simple random sample may yield, say, only low-volume customers, and may thus be unrepresentative of the population being sampled if the characteristic of interest is related to purchase volume.

It is also possible that systematic sampling may decrease the representativeness of the sample in instances where the items are ordered in a cyclical pattern. For example, a sample interval of 7 for a systematic sampling of daily retail-store sales figures would reflect the same day of data for each week and would not reveal the day-of-the-week variations in sales.

The Stratified Sample

It is sometimes desirable to break the population into different strata based on one or more characteristics, such as the frequency of purchase of a product, type of purchase (e.g., credit card versus non-credit card), or the industry in which a company competes. In such cases, a separate sample is then taken from each stratum. Technically, a stratified random sample is one in which a simple random sample is taken from each stratum of interest in the population. In practice, however, systematic and other types of random samples are sometimes taken from each of the strata. In this case, the resulting design is still referred to as a stratified sample.

Stratified samples are generally conducted according to the following procedure:

- The entire population is first divided into an exclusive and exhaustive set of strata, using some external source, such as census data, to form the strata.
- A separate random sample is selected within each stratum.
- From each separate sample, some statistic (such as a mean) is computed and properly weighted to form an overall estimated mean for the whole population.
- Sample variances are also computed within each separate stratum and appropriately weighted to yield a combined estimate for the whole population.

The two basic varieties of stratified samples are proportionate and disproportionate. In proportionate stratified sampling, the sample drawn from each stratum is made proportionate in size to the relative size of that stratum in the total population. In disproportionate stratified sampling, one departs from the preceding proportionality by taking other circumstances, such as the relative size of stratum variances, into account.

> **Example**: A company is interested in estimating the average purchases of consumers of hot cereal. The researcher may be willing to assume that, although average consumption would vary markedly by family size, the variances around the means of the strata would be more or less equal among family sizes. If so, the researcher would make use of proportionate stratified sampling.

More generally, however, both means and variances will differ among strata. If this is the case, the researcher would make use of disproportionate stratified sampling. In this instance, the number of families included in each stratum would be proportionate to (the product of) the relative size of the different family-sized strata in the population and the standard deviation of each family class. This requires, of course, that the researcher be able to estimate (from past studies) the within-group standard deviation around the average purchase quantity of each purchasing stratum. Formulas for computing sampling errors in stratified samples are briefly discussed later in this chapter and can be found in standard texts on sampling.

As intuition would suggest, the increased efficiency of stratified sampling over simple random sampling depends on how different the means (or some other statistic) really are among strata, relative to the within-stratum variability. The greater the within-stratum homogeneity and among-stratum heterogeneity, the more efficient the stratified sampling is relative to simple random sampling.

Two final considerations need to be addressed regarding stratified sampling. First, a sample size needs to be calculated for each stratum or subgroup. Second, the process of sample selection can be time-consuming and costly to carry out if many subgroups are to be used.

Nevertheless, this method continues to be widely used due to the segmentation of markets that companies routinely engage in.

The Cluster Sample

The researcher will ordinarily be interested in the characteristics of some elementary element in the population such as individual family attitudes toward a new product. However, when larger primary sampling units are desired, cluster sampling may be used. For example, the researcher may choose to sample city blocks and interview all the individual families residing therein. The blocks, not the individual families, would be selected at random. Each block consists of a cluster of respondents. The main advantage of a cluster sample relative to simple random sampling is in lower interviewing costs rather than in greater reliability.

The Area Sample: Single Stage and Multistage

As the name suggests, area sampling pertains to primary sampling of geographical areas—for example, counties, townships, blocks and other area descriptions. A single-stage area sample occurs when only one level of sampling takes place (such as a sampling of blocks) before the basic elements are sampled (the households). If a hierarchy of samples within the larger area is taken before settling on the final clusters, the resulting design is usually referred to as a multistage area sample.

> Example: Consider the sample design used by the Gallup Organization for taking a nationwide poll. Gallup draws a random sample of locations as the first stage of the sampling process. Blocks or geographic segments are then randomly sampled from each of these locations in a second stage, followed by a systematic sampling of households within the blocks or segments. A total of about 1,500 persons are usually interviewed in the typical Gallup poll.

METHODS FOR DETERMINING SAMPLE SIZE

There are several ways to classify techniques for determining sample size. Two that are of primary importance are the following:

- Whether the technique deals with fixed or sequential sampling
- Whether its logic is based on traditional or Bayesian inferential methods

Other than the brief discussion of sequential sampling that follows, this chapter is concerned with the determination of a fixed sample size with emphasis on traditional inference, such as Neyman-Pearson, rather than Bayesian inference.[1]

Although the discussion will focus on the statistical aspects of setting sample size, it should be recognized that nonstatistical dimensions can affect the value of a research project. Such things as study objectives, the length of a questionnaire, budget, time schedule, and the requirements of data analysis procedures, all have a direct effect on sample size decisions.

Fixed Versus Sequential Sampling

As the name implies, in fixed-size sampling the number of items sampled is decided in advance. The size of the sample is chosen to achieve a balance between sample reliability and sample cost. In general, all sample observations are taken before the data are analyzed.

In sequential sampling, however, the number of items is not preselected. Rather, the analyst sets up in advance a decision rule that includes not only the alternative of stopping the sampling process (and taking appropriate action, based on the sample evidence already in hand) but also the possibility of collecting more information before making a final decision. Observations may be taken either singly or in groups, the chief novelty being that the data are analyzed as they are assembled and sample size is not predetermined.

In general, sequential sampling will lead to smaller sample sizes, on average, than those associated with fixed-size samples of a given reliability. The mathematics underlying sequential sampling are, however, more complex and time consuming. In addition, the problem may be such that it is less expensive to select and analyze a sample of many items at one time than to draw items one at a time (or in small groups) and analyze each item before selecting the next.

Sampling Basics: Terminology

Intuitively we would expect that when we increase the size of the sample, our estimate of the population parameter should get closer to the true value. Also, we would expect that the less dispersed the population's characteristics are, the closer our sample estimates should be to the true parameter. After all, the reason why we sample in the first place is to make some inference about the population. These inferences should be more reliable when the sample is larger and when there is less variability in the population variables measured. To see this in action, consider the example in Exhibit 7.3.

When you read about samples in newspapers or other documents, the researcher often reports the margin of error or confidence interval for the statistical findings reported in the study. The "**margin of error"** or "**confidence interval"** is the plus-or-minus figure that represents the accuracy of the reported. Consider another example:

> A Canadian national sample showed "Who Canadians spend their money on for Mother's Day." Eighty-two percent of Canadians expect to buy gifts for their mom, compared to 20 percent for their wife and 15 percent for their mother-in-law. In terms of spending, Canadians expect to spend $93 on their wife this Mother's Day versus $58 on their mother. The national findings are accurate, plus or minus 2.75 percent, 19 times out of 20.

In this example, if 82% of your sample indicates they will "buy a gift for mom" and you use a confidence interval of 2.75%, you can be "95% confident that for ALL CANADIANS, somewhere between 79.25% (82%-2.75%) and 84.75% (82%+2.75%) would have picked that answer.

The "**confidence level"** tell you how confident you are of this result. It is expressed as a percentage of times that different samples (if repeated samples were drawn) would produce this result. The 95% confidence level means that if 20 different samples were drawn, 19 times out of 20, the results would fall in this - + confidence interval. A 99% confidence level would mean that 99 out of 100 times, the results would fall into the stated -+ confidence interval. The 95% confidence level is the most commonly used.

When you put the confidence level and the confidence interval together, you can say that you are 95% (19 out of 20 times) sure that the true percentage of the Canadian population that will "buy a gift for mom" is between 79.25% and 84.75%.

Wider confidence intervals increase the certainty that the true answer is within the range specified. These wider confidence intervals are associated with smaller sample sizes and of course produce larger sampling errors. When the costs incurred from making an error are extremely high (you are betting your company, or a multi-million dollar decision is being made) the confidence interval should be kept small. This can be done by increasing the sample size to reduce the sampling error.

Exhibit 7.3 How Do Election Polls Work?

The following is an Edited version of "Inside the paper's election polls", an article by Elsa McDowell that appeared in The Charleston Post and Courier:

The beauty of... election polls is that they are straightforward. They use statistical formulae to estimate how many people will vote one way and how many will vote another. No spin. No qualifying clauses to muddy the picture. The difficulty of.. election polls is that they are not always straightforward. How else could you explain that a poll done by one candidate shows him in the lead and that a poll done by his opponent shows her in the lead? Statisticians say there are ways to twist questions or interpret answers to give one candidate an advantage over another.

One reader took issue with a recent poll results run in The Post and Courier. He questioned whether the methodology was described in enough detail, whether the sample size was adequate. He was right about one point. The story did not make clear who was polled. It said "voters" and failed to elaborate. It should have indicated that the people polled were registered and likely to vote in the November elections. His next point is debatable. He said the sample size of 625 likely voters was insufficient for a state with nearly 4 million residents and suggested at least 800 should have been polled.

Brad Coker, the researcher responsible for the study responded that "the standard sample size used by polling groups nationally is 625. It produces, as the story stated, a margin of error of plus-or-minus 4 percent. Increasing the sample size to 800 would have produced a margin of error of plus-or-minus 3.5 - more accurate, but not so much more accurate to justify the additional cost."

"Many people do not understand how sample sizes work. They believe that, the larger the pool, the larger the sample size needs to be. It's not like that. You can take a drop of blood from a 400-pound person and it will contain the same data you would get if you took it from a 100-pound person," he said.

The reader's next concern was that the margin of error of plus-or-minus 4 applies only to the group viewed in its entirety. "If 'minorities' constituted 27 percent of the total sample, then only 169 were sampled. The margin of error then skyrockets into double digits." Coker said the reader is right and wrong. The margin of error (also known as a confidence interval) does jump for subgroups, but does not reach double digits. In this case, it moves from plus-or-minus 4 to plus-or-minus 6 to 8.

Two days before The Post and Courier ran their poll, another short story was run about a poll commissioned by MSNBC. That poll indicated incumbent Gov. Jim Hodges with a 45-43 percent lead. (Our) poll indicated challenger Mark Sanford was ahead 45 to 41 percent. When the margin of error is considered, both polls show the race is still a toss-up.

Controlling the Size of the Confidence Interval

Sampling theory teaches us that the accuracy of a sample estimate is dependent on such factors as the dispersion and skewness of the population's responses, the sample size, and the size of the population. Controlling these variables contributes to the incidence (and elimination) of sampling error. Note that "non-sampling" errors, such as bad question design or selection of a "bad" sample frame are not controlled by sample size.

Sample Size

Larger sample sizes generally produce a more accurate picture of the true characteristics of the population. Larger samples tighten the size of the confidence interval, making your estimate much more accurate. This relationship is not linear as shown in Table 7.3. Increasing sample size from 500 to 1000 reduces the confidence interval from +- 4.38 to +- 3.1.

Dispersion

The accuracy of an estimate also depends on the dispersion and skewness of the population on the question being asked. A sample of individuals registered for the republican political party would likely give a less dispersed evaluation of former president George W. Bush than would a sample of democrats. Likewise, a sample of Catholic priests would have less variability on the issue of abortion than would a survey of the general population. Accuracy of the sample estimate increases as the dispersion in the population decreases. Depending on the method of sample size calculation, dispersion is expressed as sample variance or as a proportion holding alternative positions (favorable or unfavorable toward an issue).

When using a proportion to compute sample size or estimate confidence intervals, it is easy to cover all eventualities by assuming maximum variability (50-50 proportions). Likewise, once your data has been collected, the observed proportion and final sample size can be used to obtain a more accurate estimate of the actual confidence interval.

Population Size

The size of the population also influences the size of the confidence interval, but not as much as you might expect. For a sample of 1000 respondents from a population of 100,000, the confidence interval is +- 3.08%. However, if the population were instead 1 million, the confidence interval widens to only +- 3.1%. The confidence interval is far more sensitive to changes in the sample size than to the size of the total population.

Non-sampling errors cannot be compensated for by increased sample size. Often, larger samples accentuate non-sampling errors rather than reduce them. Non-sampling errors come from samples that are not truly random, bad scales, misleading questions, incomplete surveys, etc.

Table 7.3 Proportion Based Confidence Intervals Computed at the 95% Confidence Level

Sample Size	Variability Proportions					
	50/50%	40/60%	30/70%	20/80%	90/10%	95/5%
25	20	19.6	18.3	16	12	8.7
50	14.2	13.9	13	11.4	8.5	6.2
75	11.5	11.3	10.5	9.2	6.9	5
100	10	9.8	9.2	8	6	4.4
150	8.2	8	7.5	6.6	4.9	3.6
200	7.1	7	6.5	5.7	4.3	3.1
250	6.3	6.2	5.8	5	3.8	2.7
300	5.8	5.7	5.3	4.6	3.5	2.5
400	5	4.9	4.6	4	3	2.2
500	4.5	4.4	*4.1	3.6	2.7	2
600	4.1	4	3.8	3.3	2.5	1.8
800	3.5	3.4	3.2	2.8	2.1	1.5
1000	3.1	3.0	2.8	2.5	1.9	1.4
1500	2.5	2.5	2.3	2.0	1.5	1.1
2000	2.2	2.2	2.0	1.6	1.2	0.96
2500	2	1.9	1.8	1.6	1.2	0.85
5000	1.4	1.4	1.3	1.1	.83	0.6

*Example Interpretation: In a product usage study where the expected product usage incidence rate is 30%, a sample of 500 will yield a precision of +/- 4.1 percentage points at the 95% confidence level.

This table is compute using the following formula:

$$\frac{(\text{Number of Standard Errors})^2 * ((\text{proportion})*(1\text{-proportion})) / (\text{Accuracy})}{(1+((\text{Number of Standard Errors})^2 * ((\text{proportion})*(1\text{-proportion})) / (\text{Accuracy})\text{-}1) / (\text{the population size})}$$

This formula is easily entered into a spreadsheet, to compute a sample size determination table.

Sampling Basics: Sampling Distributions and Standard Errors

The reader will recall from elementary statistics the concept of a sampling distribution. For a specified sample statistic (e.g., the sample mean) the sampling distribution is the probability distribution for all possible random samples of a given size n drawn from the specified population. The standard error of the statistic is the standard deviation of the specified sampling distribution. We shall use the following symbols in our brief review of the elementary formulas for calculating the standard error of the mean and proportion (under simple random sampling):

μ = population mean
π = population proportion regarding some attribute
σ = standard deviation of the population
s = standard deviation of the sample, adjusted to serve as an estimate of the standard deviation of the population
$\overline{X}$ = arithmetic mean of a sample
p = sample proportion
n = number of items in the sample

We here identify eight important properties associated with sampling distributions:

1. The arithmetic mean of the sampling distribution of the mean ($\overline{X}$) or of the proportion (*p*) for any given size sample equals the corresponding parameter values μ and π , respectively.

2. The sampling distribution of the means of random samples will tend toward the *normal distribution* as sample size *n* increases, regardless of the original form of the population being sampled.

3. For large samples (e.g., $n \geq 100$ and for π fairly close to 0.5) the normal distribution also represents a reasonable approximation of the binomial distribution for sample proportions.

4. In the case of finite universes, where the sample size *n* is some appreciable fraction of the total number of items in the universe, *N*, the standard error formulas should be adjusted by multiplication by the following *finite multiplier*:
$$\sqrt{\frac{N-n}{N-1}}$$

   For practical purposes, however, use of the finite multiplier is not required unless the sample contains an appreciable fraction, say 10% or more, of the population being sampled. At a 10% sampling fraction, taking it into account will reduce the random sampling error by 5%. If the sampling fraction is 5%, 2%, or 1%, not ignoring it will reduce error very slightly—2.5%, 1.0%, and 0.5%, respectively.

5. Probabilities of normally distributed variables depend on the distance (expressed in multiples of the standard deviation) of the value of the variable from the distribution's mean. If we subtract a given population mean μ from a normally distributed variable X_i and divide this result by the original standard deviation σ , we get a standardized variable Z_i that is also normally distributed but with zero mean and unit standard deviation. In symbols this is as follows:
$$Z_i = \frac{X_i - \mu}{\sigma}$$

   Table A.1 in Appendix A at the end of this book presents the standardized normal distribution in tabular form. Note further that the original variate may be a sample mean, $\overline{X}$. If so, the denominator is the *standard error* (i.e., standard deviation of the sampling distribution). We can then define *Z* as some number of standard errors away from the mean of the sampling distribution
$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$
   where $\sigma_{\overline{X}}$ denotes the standard error of the mean. (A similar idea is involved in the case of the standard error of the proportion.)

6. The formulas for the standard error of the mean and proportion of simple random samples are, respectively, the following: Mean Proportion
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \qquad\qquad \sigma_{\overline{P}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

7. If the population standard deviation σ is not known, which is often the case, we can estimate it from the sample observations by use of the following formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_1 - \overline{X})^2}{n-1}}$$

We can consider *s* to be an *estimator* of the population standard deviation, σ. In small samples (e.g., less than 30), the *t* distribution of Table A.2 in Appendix A is appropriate for finding probability points. However, if the sample size exceeds 30 or so, the standardized normal distribution of Table A.1 is a good approximation of the *t* distribution.
In cases where σ is estimated by *s*, the standard error of the mean becomes

$$est.\sigma_{\overline{X}} = \frac{s}{\sqrt{n}}$$

where $est.\sigma_{\overline{X}}$ denotes the fact that σ is estimated from *s*, as defined in the preceding equation.

8. Analogously, in the case of the standard error of the proportion, we can use the sample proportion *p* as an estimate of π to obtain

$$est.\sigma_{\overline{P}} = \sqrt{\frac{p(1-p)}{n}}$$

as an estimated standard error of the proportion. Strictly speaking, *n*–1 should appear in the denominator. However, if *n* exceeds about 100 (which is typical of the samples obtained in marketing research), this adjustment makes little difference in the results.

Methods of Estimating Sample Size

In our discussion of sample planning we pointed out that there are four traditional approaches to determining sample size.

- First, the analyst can simply select a size either arbitrarily or on the basis of some judgmentally based criterion. Similarly, there may be instances where the size of sample represents all that was available at the time—such as when a sample is composed of members of some organization and data collection occurs during a meeting of the organization.

- Second, analysis considerations may be involved and the sample size is determined from the minimum cell size needed. For example, if the critical aspect of the analysis required a cross-tabulation on three categorical variables that created 12 cells (2 categories x 3 categories x 2 categories = 12 cells), and it was felt that there should be at least 30 observations in a cell, then the absolute minimum sample size needed would be 360.

- Third, the budget may determine the sample size. If, for example, the research design for an online survey of physicians, the cost of each interview was estimated to be $50, and the budget allotted to data collection was $10,000, then the sample size would be 200.

It may appear that these methods are for nonprobability samples. While this certainly is true, these methods are also applicable to probability samples and have occasionally been used for such samples. For probability samples, the precision must be determined after the fact.

- In a fourth approach to sample size determination, we specify the desired precision in advance and then applying the appropriate standard error formula to calculate the sample size. This approach using traditional inference (Neyman-Pearson), relies on either of two major classes of procedures available for estimating sample sizes. The first and better known of these is based on the idea of constructing confidence intervals around sample means or proportions. This can be called the confidence-interval approach. The second approach makes use of both type I (rejecting the true null hypothesis) and type II (accepting a false null hypothesis) error risks and can be called the hypothesis-testing approach. We discuss each of these approaches in turn.

Before doing this, however, two points must be made. First, as with the other approaches, the analyst must still calculate the standard error <u>after data collection</u> in order to know what it is for the actual sample that provided data. Second, the size of sample that results from traditional inference refers to the completed (or resulting) sample. Depending on the data collection method used, the original survey delivery may have to be much larger. For example, suppose that the size of the desired sample was 582. A mail survey is used for data collection and past experience has shown that the response rate would be around 25%. The original sample size in this case would have to be 2,328 so that 582 responses would be obtained.

The Confidence Interval Approach

It is not unusual to construct a confidence interval around some sample-based mean or proportion. The standard error formulas are employed for this purpose. For example, suppose a researcher sampled 100 student consumers and noted that their average per capita consumption of specialty fruit/energy drinks was 2.6 pints per week. Past studies indicate that the population standard deviation σ can be assumed to be 0.3 pint.

With this information, we can find a range around the sample mean level of 2.6 pints for which some prespecified probability statement can be made about the process underlying the construction of such confidence intervals.

For example, suppose that we wished to set up a 95% confidence interval around the sample mean of 2.6 pints. We would proceed by first computing the standard error of the mean:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.3}{\sqrt{100}} = 0.03$$

From Table A.1 in Appendix A we find that the central 95% of the normal distribution lies within ±1.96 *Z* variates (2.5% of the total area is in each tail of the normal curve).
With this information we can then set up the 95% confidence interval as

$$\overline{X} \pm 1.96\sigma_{\overline{X}} = 2.6 \pm 1.96(0.03)$$

and we note that the 95% confidence interval ranges from 2.54 to 2.66 pints.
Thus, the preassigned chance of finding the true population mean to be within 2.54 and 2.66 pints is 95%.

This basic idea can be adapted for finding the appropriate sample size that will lead to a certain desired confidence interval. To illustrate, let us now suppose that a researcher is interested in estimating annual per capita consumption of specialty fruit/energy drinks for adults living in a particular area of the United States. The researcher knows that it is possible to take a random sample of respondents in the area and compute a sample mean. However, what the researcher really wants to do is be able to state with, say, 95% confidence that the population mean falls within some allowable interval computed about the sample mean. The researcher wants to find a sample size that will permit this kind of statement.

The Case of the Sample Mean

Let us first assume that the allowable error is 0.5 gallon of energy drinks per capita and the level of confidence is 95%. With this in mind, we go through the following checklist:

1. *Specify the amount of error (E) that can be allowed*. This is the maximum allowable difference between the sample mean and the population mean. $\overline{X} \pm E$, therefore, defines the interval within which μ will lie with some prespecified level of confidence. In our example, the allowable error is set at *E*, or 0.5 gallon per year.

2. *Specify the desired level of confidence*. In our illustrative problem involving specialty fruit/energy drink consumption, the confidence level is set at 95%

3. *Determine the number of standard errors (Z) associated with the confidence level*. This is accomplished by use of a table of probabilities for a normal distribution. For a 95% confidence level, reference to Table A.1 indicates that the Z value that allows a 0.025 probability that the population mean will fall outside *one* end of the interval is Z = 1.96. Since we can allow a *total* probability of 0.05 that the population mean will lie outside *either* end of the interval, Z = 1.96 is the correct value for a 95% confidence level.

4. *Estimate the standard deviation of the population*. The standard deviation can be estimated by (a) judgment; (b) reference to other studies; or (c) by the use of a pilot sample. Suppose that the standard deviation of the area's population for specialty fruit/energy drink consumption is assumed to be 4.0 gallons per capita per year.

5. *Calculate the sample size using the formula for the standard error of the mean*. One standard error of the mean is to be set equal to the allowable error (*E* = 0.5) divided by the appropriate *Z* value of 1.96.

$$\sigma_{\overline{X}} = \frac{E}{Z} = \frac{0.5}{1.96} = 0.255$$

   This will assure us that the interval to be computed around the to-be-found sample mean will have a 95% preassigned chance of being ±0.5 gallon away from the population mean.

6. Neglecting the finite multiplier, we then solve for *n* in the formula

$$\sigma_{\overline{X}} = \frac{E}{Z} = \frac{\sigma}{\sqrt{n}} \qquad \text{or} \qquad \sigma_{\overline{X}} = 0.255 = \frac{4.0}{\sqrt{n}}$$

Hence, $n \cong 246$ (rounded)

7. In general, we can find n directly from the following formula:

$$n = \frac{\sigma^2 Z^2}{E^2} = \frac{16(1.96)^2}{(0.5)^2} \cong 246$$

If the resulting sample size represents a significant proportion of the population, say 10% or more, the finite population multiplier is required and the sample size must be recalculated using the following formula, where N is the size of the population:

$$n = \frac{N(\sigma^2 Z^{2)}}{NE^2 + \sigma^2 Z^2}$$

The Case of the Sample Proportion

Suppose that, in addition to estimating the mean number of gallons of specialty fruit/energy drinks consumed per capita per year, the researcher is also concerned with estimating the proportion of respondents using one or more specialty fruit/energy drinks in the past year. How should the sample size be determined in this case?

The procedures for determining sample size for interval estimates of proportions are very similar to those for interval estimates of means. In this case the following checklist would be used:

1. *Specify the amount of error that can be allowed*. Suppose that the desired reliability is such that an allowable interval of $p - \pi = \pm 0.05$ is set; that is, the allowable error E is 0.05, or 5 percentage points.
2. *Specify the desired level of confidence*. Suppose that the level of confidence here, as in the preceding problem, is set at 95%.
3. *Determine the number of standard errors* Z *associated with the confidence level*. This will be the same as for the preceding estimation; $Z = 1.96$.
4. *Estimate the population proportion,* (π). The population proportion can again be estimated by *judgment*, by *reference to other studies*, or by *the results of a pilot sample*. Suppose that π is assumed to be 0.4 in this case; that is, the researcher assumes that 40% of the population used one or more specialty fruit/energy drinks last year.
5. *Calculate the sample size using the formula for the standard error of the proportion*. One standard error of the proportion is to be set equal to the allowable error ($E = 0.05$) divided by the appropriate Z value of 1.96.

$$\sigma_{\bar{P}} = \frac{E}{Z} = \frac{0.05}{1.96} = 0.0255$$

6. Neglecting the finite multiplier, we then solve for n in the following formula:

$$\sigma_{\bar{P}} = \frac{E}{Z} = \sqrt{\frac{\pi(1-\pi)}{n}} = 0.0255 = \sqrt{\frac{0.4(0.6)}{n}}$$

Hence, $n \cong 369$ (rounded)

7. In general, we can find n directly from the formula

$$n = \frac{\pi(1-\pi)Z^2}{E^2} = \frac{0.4(1-0.4)(1.96)^2}{(0.05)^2} \cong 369$$

Once again, if the resulting n is 10% or more of the population size, the finite population multiplier is required and the sample size can be computed from the following:

$$n = \frac{N\pi(1-\pi)Z^2}{NE^2 + \pi(1-\pi)Z^2}$$

Determining Sample Size When More Than One Interval Estimate Is to Be Made from the Same Sample

The usual case when collecting sample data for estimation of various parameters is that more than one estimate is to be made. The sample size for each of the estimates will usually be different. Since only one sample is to be chosen, what size should it be?

A strict adherence to the allowable error and the confidence levels specified in the calculation of the sample sizes for the individual estimation problems leaves no choice but to take the largest sample size calculated. This will give more precision for the other estimates than was specified but will meet the specification for the estimate for which the size of sample was calculated. In the specialty fruit/energy drink consumption problem, for example, the sample size would be 369 (the sample size calculated for estimating the proportion of users) rather than 246 (the sample size calculated for estimating the mean amount used). Remember, the sample sizes determined in this manner are for obtained samples. In order to determine the size of the original sample the researcher must estimate the rate of response expected. For example, if a mail survey was to be conducted and it was believed that only a 20% response would be obtained, a desired obtained sample of 250 would need an original sample of 1,250.

Devices for Calculating Sample Size

In practice, manual devices (paper nomographs) and online sample size calculators can be used to easily compute a sample size helpful in rough-guide situations where the researcher is not sure of either allowable error levels or population standard deviations. One such calculator is located at http://marketing.byu.edu/samplesizecalculator.html .

The Hypothesis-Testing Approach

As indicated earlier, sample sizes can also be determined (within the apparatus of traditional statistical inference) by the hypothesis-testing approach. In this case the procedures are more are more elaborate. We shall need both an assumed probability of making a type I error—called the alpha risk—and an assumed probability of making a type II error—called the beta risk. These risks are, in turn, based on H_0: the null hypothesis, and H_1: the alternate hypothesis.

In hypothesis testing the sample results sometimes lead us to reject H_0 when it is true. This is a type I error. On other occasions the sample findings may lead us to accept H_0 when it is false. This is a type II error. The nature of these errors is shown in Table 7.4.

A numerical example should make this approach clearer. We first consider the case for means and then the case for proportions. Before doing this, however, a few words on the relationship between the Type I and Type II errors are in order. The relationship between these two errors is an inverse. The ability of a sample to protect against the type II error is called statistical power.

Table 7.3 Sample Size When Estimating Population Mean and Proportion (Selected Samples)

A. Mean

Population size, N	*Reliability, r*	*Z-value, Z*	*Standard deviation, s*	*Precision, d*	*Sample size, n*
400	95%	1.96	1.0	± .25	53
400	90%	1.645	1.0	± .25	39
400	90%	1.645	s	± .25s	39
400	95%	1.96	s	± 1/3s	32
400	95%	1.96	1.0	± .33	32
400	99.7%	3.0	s	± 1/3s	67
200	95%	1.96	s	± 1/3s	30
1600	95%	1.96	s	± 1/3s	34
$N \to \infty$	95%	1.96	s	± 1/3s	35
$N \to \infty$	90%	1.645	s	± .25s	43

B. Proportion

Population Size N	*Reliability r*	*Z-value Z*	*Standard Deviation s*	*Precision d*	*Sample Size n*
500	95%	1.96	0.5	± 10%	81
500	95%	1.96	.3 or .7	± 10%	69
$N \to \infty$	95%	1.96	0.5	± 10%	96
$N \to \infty$	95%	1.96	0.5	± 5%	384
$N \to \infty$	90%	1.645	0.5	± 10%	68
$N \to \infty$	99%	2.58	0.5	± 5%	666
$N \to \infty$	99.7%	3.00	0.5	± 5%	900
$N \to \infty$	99%	2.58	0.5	± 1%	16,641
$N \to \infty$	99.7%	3.00	0.5	± 1%	22,500
400	90%	1.645	.2 or .8*	± 10%	39
200	90%	1.645	.2 or .8*	± 10%	36

*As the difference between *p* and 0.5 increases, the sampling distribution for *p* becomes more skewed and may deviate from the normal approximation. Thus, the data should be interpreted with care for small samples as *p* approaches either 0.0 or 1.0. (The corresponding percentage would be 0.0 or 100.0.)

SOURCE: Tatham, 1979, p.b.

Table 7.4 Types of Error in Making a Wrong Decision

Action	**H_o is true**	**H_o is false**
Accept H_o	No error	Type II error (β)
Reject H_o	Type I error (α)	No error

When the hypothesis is one of difference, a type II error occurs when what is really chance variation is accepted as a real difference. Taking statistical power into account often indicates that larger (and, thus, more costly) samples are needed. Sample size is affected by the effect size—the needed or expected difference. When it is large, say more than 15 percentage points, statistical power is usually not a problem. Very small effect sizes (e.g., two percentage points) require such large sample sizes as to be impractical; surveys cannot really reliably measure such small differences or changes. As an illustration, if a researcher wants to discern an effect size of, say, 10 points at a 95 % confidence level and a desired power to detect the superiority of one alternative over another of 80%, the approximate sample size needed would be 310. If the desired power is raised to 95%, the sample size needed would be about 540 (Semon, 1994).

The researcher and the manager as well, must consider the relative business risks when setting appropriate levels of protection against the two types of errors (Semon, 1990).

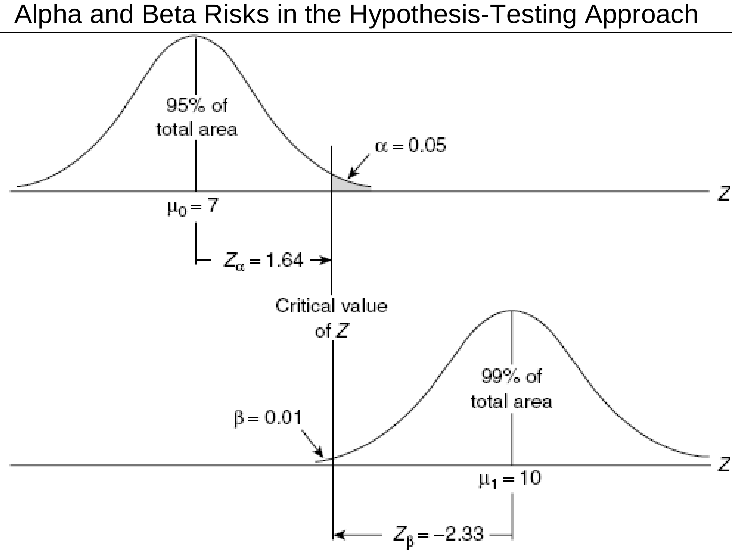
The Case Involving Means

As an illustrative example, let us assume that a store test of a new bleaching agent is to be conducted. It has been determined earlier that if the (population) sales per store average only 7 cases per week, the new product should not be marketed. On the other hand, a mean sales level of 10 cases per week would justify marketing the new product nationally. Using methods of traditional inference, how should the number of sample stores for the market test be determined? The procedures are similar to those for interval estimation problems but are somewhat more complicated. Specifically, we go through the following checklist:

1. Specify the values for the null (H_0) and the alternate (H_1) hypotheses to be tested in terms of population means μ_0 and μ_1, respectively. (By convention, the null hypothesis is the one that would result in no change being made, if accepted.) In the bleach market-introduction problem, the values are set at H_0: $\mu_0 = 7$ cases per week, and H_1: $\mu_1 = 10$ cases per week.

2. Specify the allowable probabilities (α and β, respectively) of type I and type II errors. The type I error is the error of rejecting a true null hypothesis. The type II error is made when the alternate hypothesis is rejected when it is true. α and β are the allowable probabilities of making those two types of errors, respectively. They are shown graphically in Figure 7.3, where we assume that in the bleach-introduction problem the allowable probabilities of error are assigned as $\alpha = 0.05$ and $\beta = 0.01$.

Figure 7.3 Alpha and Beta Risks in the Hypothesis-Testing Approach



3. Determine the number of standard errors associated with each of the error probabilities α and β. For a one-tailed test the Z values for the 0.05 and 0.01 risks, respectively, are found from Table A.1 in Appendix A to be $Z_\alpha = 1.64$ and $Z_\beta = 2.33$. These are shown in figure 7.3. Note that in the figure we affix a minus sign to the value of Z_β since the critical values lies to the left of $\mu_1 = 10$.

4. Estimate the population standard deviation σ. In the case of the new bleach the standard deviation of cases sold per store per week is assumed to be five cases.

5. *Calculate the sample size that will meet the α and β error requirements*. Because *two* sampling distributions are involved, a simultaneous solution of two equations is required to determine the sample size and critical value that will satisfy both equations. These equations are the following:

$$\text{critical value} = \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\text{critical value} = \mu_1 - Z_{\beta\alpha} \frac{\sigma}{\sqrt{n}}$$

6. Setting the right-hand side of these two equations equal and solving for *n* gives

$$n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_1 - \mu_0)^2}$$

In the bleach problem the desired sample size is

$$n = \frac{(1.64 + 2.33)^2 5^2}{(10-7)^2} \cong 44 \text{ stores (rounded)}$$

Having solved for *n*, the sample size, we can then go on to solve for the critical value for the mean number of cases by means of the following substitution:

$$\text{critical value} = \mu_1 - Z_\beta \frac{\sigma}{\sqrt{n}}$$

$$= 10 - (2.33)\frac{5}{\sqrt{44}} = 8.24 \text{ cases}$$

Alternatively, we could find the critical value from the first of the two equations:

$$\text{critical value} = 7 + (1.64)\ \frac{5}{\sqrt{44}} = 8.24$$

The decision rule then becomes the following: Take a sample of 44 stores for the controlled store test. If the mean number of cases of the new bleach sold per week in the sample stores is less than or equal to 8.24 cases, do not introduce the product. If the mean number of cases of bleach sold per week is greater than 8.24 cases, introduce the product.

The Case Involving Proportions

For sample-size determination involving proportions, the following analogous steps are required:

1. Specify the values of the null (H_0) and the alternate (H_1) hypotheses to be tested in terms of population proportions π_0 and π_1, respectively.

2. Specify the allowable probabilities (α and β, respectively) of type I and type II errors.

3. Determine the number of standard errors associated with each of these error probabilities (Z_α and Z_β).

4. Calculate the desired sample size *n* from the formula:

$$n = \left[\frac{Z_\alpha \sqrt{\pi_0(1-\pi_0)} + Z_\beta \sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0} \right]^2$$

This formula is appropriate for relatively large samples ($n \geq 100$), where the normal distribution is a good approximation to the binomial. To illustrate its application, suppose that a researcher is interested in the true proportion of residents in a large city who would be willing to pay over $400 for a portable refrigerator-bar combination if it were commercialized.

Assume that the marketing researcher would recommend commercialization of the firm's refrigerator-bar combination if the true proportion of consumers who would pay over $400 for this class of goods is 70%. If the proportion is only 60%, the researcher would not recommend commercialization. The hypotheses are:

$$H_0\text{: } \pi_0 = 0.6$$

$$H_1\text{: } \pi_1 = 0.7$$

The alpha risk associated with the null (status quo) hypothesis is selected by the researcher to be 0.05 if the true proportion π is equal to 0.6. Moreover, the researcher is willing to assume a beta risk of 0.1 if the true proportion is equal to 0.7. With these assumptions it is possible to obtain the approximate sample size by using the preceding formula:

$$n = \left[\frac{Z_\alpha \sqrt{\pi_0(1-\pi_0)} + Z_\beta \sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0} \right]^2$$

where $Z_\alpha = Z_{0.05} = 1.64$, $Z_\beta = Z_{0.1} = 1.28$, $\pi_0 = 0.6$, and $\pi_1 = 0.7$. The solution follows:

$$n = \left[\frac{1.64\sqrt{0.6(0.4)} + 1.28\sqrt{0.7(0.3)}}{0.7 - 0.6} \right]^2 \cong 193 \text{ (rounded)}$$

Accordingly, in this example the sample size to take is 193. The critical value can be found analogously as follows:

$$\text{critical value} = \pi_1 - Z_\beta \sqrt{\frac{\pi_1(1-\pi_1)}{n}}$$

$$= 0.7 - (1.28)\sqrt{\frac{0.7(0.3)}{193}} = 0.658$$

In this case the decision rule is the following: Take a sample of 193 residents. If the sample proportion who would pay over $400 is less than or equal to 0.658, do not commercialize the refrigerator-bar combination. If the sample proportion exceeds 0.658, commercialize the product.

DETERMINING SAMPLE SIZE FOR OTHER PROBABILITY–SAMPLE DESIGNS

Thus far we have discussed only the determination of *simple* random-sample sizes using the methods of traditional statistical inference. How are the sizes for other types of random-sample designs—systematic, stratified, cluster, area, and multistage—determined?

The answer to this question is that the same *general* procedures are used to determine the overall sample size, but the formulas for the standard errors differ. The formulas become more complex and difficult to estimate as one considers stratified sampling, cluster sampling, or the other more elaborate designs. This is because the standard error for these designs is partially a function of the standard deviation (or proportion) of each stratum or cluster included in the design. For a multistage sample consisting of several strata in one stage followed by clusters in another and systematic sampling in a third, the standard error formula can become very complex indeed. And once the overall sample size is determined; it must be apportioned among the strata and clusters, which also adds to the complexity.

Appropriate formulas for estimating standard errors and sample sizes for other random-sample designs are available elsewhere (Sudman, 1976; Kish, 1965). In general, as compared with the size of simple random samples, systematic samples may be the same (for purposes of calculating the standard error, the assumption is typically made that the systematic sample is a simple random sample). Stratified samples are usually smaller, and cluster samples will usually be larger in size to provide the same reliability as a simple random sample. If used properly, stratification usually results in a smaller sampling error than is given by a comparable-size simple random sample. Consequently, the advantages of a stratified sample design over a simple random sampling design are as follows (Sangren, 2000, p. 67):

- For the same level of precision, one would need a smaller sample size in total, and this leads to a *lower cost*
- For the same total sample size, one would gain a greater precision for any estimates that are to be made

EVALUATION OF THE TRADITIONAL APPROACH

If one were to devise the ideal method of determining sample size, at a minimum one would want it to meet the criteria of being:

1. Logically complete
2. Adaptable to a wide range of sampling situations
3. Simple to use

If the traditional (Neyman-Pearson) approach to sample-size determination were to be rated on these criteria, the rating would be low for logical completeness and high for both adaptability and simplicity.

The traditional approach is logically incomplete because sample size is specified as being a function only of the conditional probabilities of making errors. The conditional costs of wrong decisions, prior probabilities, nonsampling errors, and the cost of sampling are not considered in the model. More advanced texts, however, do consider traditional sample-size determination by means of formulas that include the costs of sampling (see Cochran, 1963, Chapter 4).

The fact that these variables are excluded implies that somehow they must be taken into account outside the model. However, the only way that accommodation can be made is through adjustment of either the specified confidence level or the assigned alpha and beta risks.

SUMMARY

One of the most difficult problems in research design is the one concerned with the size of sample to take. We discussed the determination of sample size from the standpoint of traditional inferential methods and all aspects of sample size determination were covered. Estimation and hypothesis-testing applications were discussed. We concluded the chapter with an evaluation of this traditional approach to determining sample size.

REFERENCES

Cochran, W. G. (1963). Sampling techniques (2nd ed.). New York: Wiley.

Deming, W. E. (1960). Sample designs in business research. New York: Wiley.

Dillman, D. A. (2000). Mail and internet surveys: The tailored design method (2nd ed.). New York:Wiley.

McDowell, Elisa (2002), "Inside the paper's election polls", The Charleston Post and Courier Accessed August 21, 2009 http://www.accessmylibrary.com/coms2/summary_0286-2181489_ITM

Kish, L. (1965). Survey sampling. New York: Wiley.

National Center for Health Statistics (2009), Data: Cell Phone-Only Households State-by-State, Associated Press, as reported on Fox News, March 11, 2009 http://www.foxnews.com/story/0,2933,508926,00.html;

http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200905.htm

http://marketing.byu.edu/samplesizecalculator.html

http://www.qualtrics.com/wiki

Sangren, S. (2000, April). Survey and sampling in an imperfect world: More second-best solutions. Quirk's Marketing Research Review, pp. 16ff.

Semon, T. T. (1990, January 8). Don't panic when sample isn't quite what you expected. Marketing News, 24, 31.

Semon, T. T. (1994, January 3). Save a few bucks on sample size, risk millions in opportunity loss. Marketing News, 28, 19.

Semon, T. T. (2000, June 9). Quality sampling, research means avoiding sloppiness. Marketing News, 34, 11.

Sudman, S. (1976). Applied sampling. New York:Academic Press.

Survey Methods Centre (1995). Quota versus probability sampling. Survey Methods Centre Newsletter, 15(1).

Tatham, E. (1979). A report on sample size selection. Shawnee Mission, KS: Customer Satisfaction Institute.

Unlisted phone numbers on the rise (1989) http://findarticles.com/p/articles/mi_m3065/is_n8_v18/ai_7812479/

Zinkhan, G., Burton, S., & Wallendorf, M. (1983). Marketing applications for snowball sampling. In W.R. Darden et al. (Eds.), Research methods and causal models in marketing. Chicago: American Marketing Association.

This page is intentionally blank

CHAPTER 8
EXPERIMENTATION

Experimentation is widely used in marketing research. Marketing experiments have been conducted in such diverse activities as evaluating new products, selecting advertising copy themes, determining the frequency of salespeople's calls, and evaluating all aspects of a movie (including ending, pacing, music, and even the story line). For example, as shown in Exhibit 8.1 the ending of the very successful movie Fatal Attraction was changed because test audiences did not like the original ending.

This chapter discusses the objectives of experimentation illustrates techniques for designing and analyzing marketing experiments including:

- the nature of experimentation
- ingredients of a marketing experiment
- sources of invalidity
- models of experimental design
- panels and experimental design
- difficulties in field experiments in marketing

This chapter further discusses experimental designs and logic for advanced surveys using the capabilities of Qualtrics.com, including the following:

- piping of text, graphics and experimental treatments
- simple branching and compound branching logic, looping and piping of answers
- question and treatment blocks
- quota fulfillment
- randomization of answer choices, questions, treatment blocks and alternative questionnaire forms.

The Nature of Experimentation

Two general types of experimental designs exist—natural and controlled. A <u>natural experiment</u> is one in which the investigator intervenes only to the extent required for measurement, and there is no deliberate manipulation of an assumed causal variable. “Nature” produces the changes. In contrast, in a <u>controlled experiment</u> two kinds of intervention are needed:

1. manipulation of at least one assumed causal variable
2. random assignment of subjects to experimental and control groups.

True experiments have both types of intervention, while *quasi-experiments* manipulate the variables but do not randomly assign the subjects. All true experiments have certain things in common—treatments (i.e., assumed causal variables), an outcome measure, units of assignment, and some comparison from which change can be inferred and, it is hoped, attributed to the treatment. Quasi-experiments, on the other hand, have treatment, outcome measures, and experimental units but do not randomly assign subjects to treatments. Rather, subjects already belong to groups that differ from each other in ways other than the presence of a treatment whose effects are being tested (Cook and Campbell, 1990).

Exhibit 8-1 Test Audiences have Profound Effect on Movies

What if E.T. hadn't made it home, or Richard Gere had failed to come back for his pretty woman, Julia Roberts? Believe it or not, in the original versions of these films, E.T. died on American soil and it was Roberts who rejected Gere at the end of "Pretty Woman."

So who changed these potential misses into hits? It wasn't the producers or the writers who prevailed for change – it was the moviegoers. In what may be Hollywood's last and most closely guarded secret, the test audience is having a profound effect on the movies you watch.

Scientific Findings?

In the original version of the hit "My Best Friend's Wedding," Rupert Everett had a minor role as Julia Roberts' gay best friend. But test audiences wanted more. So the ending was scrapped, the set rebuilt, and Everett's character came back for one final appearance. That's what can happen if test audiences love you. But what if they loathe you? In the 1987 thriller Fatal Attraction test audiences so despised Glenn Close's character that they became responsible for having her killed off in the end.

Director Ron Howard and his partner, producer Brian Grazer, are responsible for hits like Far and Away, Ransom, Apollo 13, and A Beautiful Mind. Howard says, "What I would hate to do is put the movie out there, find out that the audience is confused about something or upset about something that you could have fixed and realize I had no idea they'd respond that way." Grazer and Howard were dealt one surprise when they tested the 1989 film Parenthood. "The audience told us there was too much vulgarity in the movie," Grasier says. "We took out a lot of the vulgarity. The scores went up, made us feel better, the movie played better. It didn't offend anybody."

Source: Adapted from Bay (1998)

Objectives

The term experimentation is used in a variety of ways and for a variety of objectives. In the discussion of this chapter we shall use the term "experimentation" to describe an experiment may be conducted for the purpose of identifying relevant variables as well as the functional form of the model that links these variables with the criterion variable under study.

Perhaps the characteristic that best distinguishes experimentation from observational studies (which are also employed in measurement and estimation) is that experimentation denotes some researcher intervention and control over the factors affecting the response variable of interest to the researcher.

Experimentation permits the establishment of causal relationships. In contrast, correlation analysis (a useful technique in observational studies) permits the analyst to measure the degree to which changes in two or more variables are associated with each other.

An Industry Example

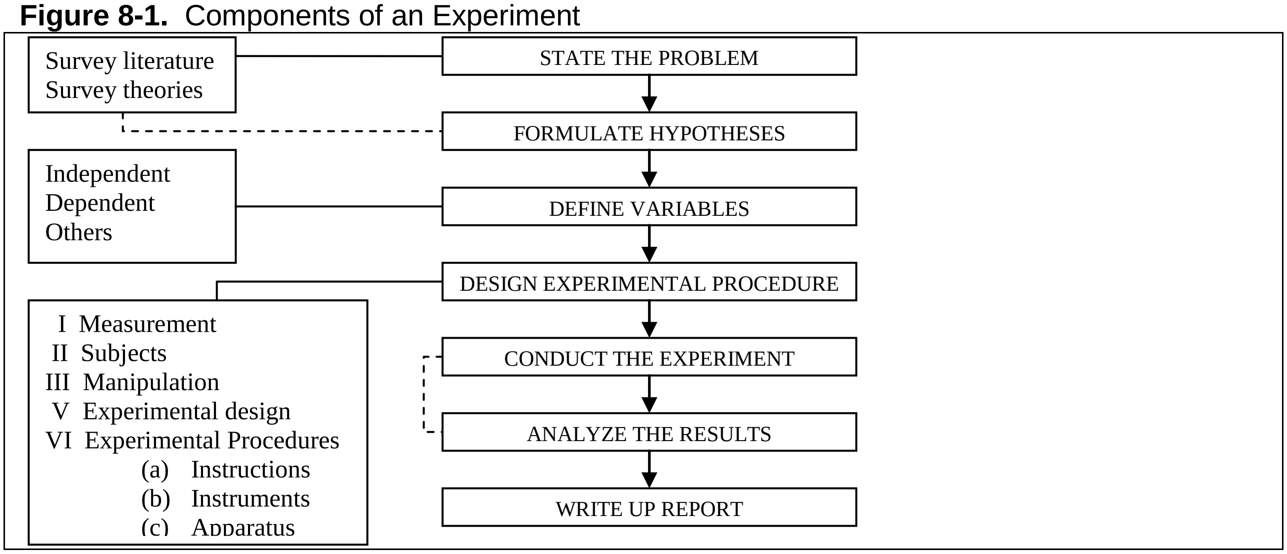
A national producer of packaged candies was interested in children's preference for various formulations of one of its well-known candy bars. Type of chocolate, quantity of peanuts, and amount of caramel were independently varied in a factorial design (described later in the chapter) of 2 types of chocolate by 3 quantities of peanuts by 3 amounts of caramel. Paired combinations (i.e., two combinations at a time) comparisons involving the 18 combinations were made up and evaluated by various school children between 8 and 12 years of age. Interestingly enough, the company found that preferences for type of chocolate varied with the amount of caramel. In addition, while children preferred more peanuts to fewer peanuts, the intermediate level of caramel was the most preferred. The company modified its formulation to match the most preferred test combination.

Many other experiments have been carried out involving taste testing, package design, advertising type and quality, price sensitivity, and other marketing variables.

Ingredients of a Marketing Experiment

An experiment involves a series of interrelated steps, as shown in Figure 8-1 (note the similarity to Figure 2-1). Our concern in this chapter is primarily with defining variables, designing the experimental procedure, and conducting the experiment. The other steps are discussed elsewhere in this book, and for experimentation they do not differ from general concepts.

Figure 8-1. Components of an Experiment



All experiments involve three types of variables. First, there is the "treatment" variable whose effect upon some other variable the experiment is designed to measure. This is the variable that is manipulated, and presumed to be the cause. It is often referred to as the independent variable. Marketing experiments often involve more than one treatment variable. When this is the case, the researcher may be interested in observing the effects of combinations of treatment variables, in addition to the effect of each one individually. In short, there may be interaction effects. Interaction refers to the situation like the candy bar example cited above, where the response to changes in the levels of one treatment variable (type of chocolate) is dependent on the level of some other treatment variable(s) (amount of caramel) in the experiment.

The second broad type of variable in an experiment is the outcome or dependent variable. In the preceding candy bar example, the dependent variable was product preference.

The last category of variables consists of those other than the manipulated independent variables that could influence the observed effects (i.e., dependent variable). These are known as extraneous variables, and unless controlled adequately they are the source of errors in an experiment. These will be discussed in a later section of this chapter.

Test Objects

The terms test units, test objects, and subjects are used interchangeably in discussing experimentation. All are used to refer to the units whose responses to the experimental treatment are being studied. In marketing research the experimenter has a choice of three possible universes of test units—people, stores, and market areas. Which is most appropriate depends on the problem forming the basis of the experiment (Banks, 1965, pg. 13-15).

The experimenter must contend with differences among the inherent properties of the test objects. For example, if a researcher is interested in the effect of shelf height on the sales of a packaged consumer product, it is to be expected that stores will vary in their amount of shopping traffic, placement of shelving units, and the like. If the experimenter is interested in the effect of various shelf heights on product sales over a variety of store sizes, several stores will have to be used in the analysis. If so, he or she may use a technique called covariance analysis, in which responses to the controlled variables (shelf height) are adjusted for inherent differences in the test objects (stores) through measurement of these characteristics before (or during) the experiment.

Measurement, Manipulation, and Experimental Procedures

A critical aspect of all experiments, indeed of all marketing research, is measurement. Our concern at this point is with the operational problems of measurement. The concepts, levels, and techniques of measurement, and scaling are covered in Chapter 9.

In a marketing experiment, it is the outcome or dependent variable that is measured. Generally, the operational measures used can be classified as physiological or psychological measures. Physiological measures include those obtained from devices that measure such things as eye movements or electrical conductivity of the skin (psychogalvanometer) during advertising readership. Such devices are used in laboratory experiments. Psychological measures include verbal measures include spoken and written responses, including responses provided interactively with a personal computer. They may additionally include direct measures such as dollar amounts or units of a product that are sold or consumed, and actual behavior (or assumed actual behavior) of people under the conditions of the experiment. Such measures are obtained by self-report, observation, or electronic means.

Turning now to manipulation, an experimental treatment must be capable of variation. There are at least three ways in which variation in the independent variable can be achieved. First, the type of variable can be manipulated. For example, a company interested in the effect of image on sales could conduct an experiment by running a series of advertisements, each of which was designed to convey a different image of its product. Variation is generated in the type of image conveyed.

Second, there is the presence versus absence technique. For instance, one group of people could be shown a new advertisement and their responses to an attitude measurement could be compared with the response from a group that did not see the advertisement.

Finally, the amount of a variable can be manipulated; different amounts are administered to different groups. This technique is used in such experiments as those where different prices for a product are "tested" and the outcome measured as "units sold."

As with any approach to marketing research, all phases of an experiment should be carefully planned in advance. After decisions have been made concerning measurement, research subjects, experimental design, control techniques, and manipulation, there is the need to plan everything that will take place in the actual experiment itself through to the end of data collection. This includes the setting of the experiment, physical arrangements, apparatus that will be used, data collection forms, instructions, recording of the dependent variable, and so forth.

Difficulty of Control

In any experiment, as previously mentioned, extraneous factors could potentially affect the dependent variable (response). Those factors that probably would not influence the dependent variable can be ignored. On the other hand, those extraneous variables that might reasonably influence the dependent variable must be controlled in some way in order that error, and thus threats to valid inference, can be minimized.

In any experiment, control over all possible variables affecting the response is rarely possible. Even in the laboratory it is not possible to control all variables that could conceivably affect the outcome. But compared with the laboratory situation, the researcher who is working in the marketplace has a really difficult control job to do. The marketing researcher must try to design the experiment so that the effects of uncontrolled variables do not obscure and bias the nature of the response to the treatment variables that are being controlled.

There are five general techniques for controlling extraneous variables.

1. **Elimination of extraneous variables**, including controlling the situation in which an experiment is being conducted so as to keep out extraneous forces. Used in this way, control is much easier to achieve in a laboratory environment than in a field setting.

2. **Constancy of conditions,** including the ability to determine which subjects or test units receive a particular treatment at a particular time. Control of the treatment variable helps separate out the effects attributable to irrelevancies that are correlated with a treatment.

3. **Matching (sometimes called balancing)**. Marketing researchers most often match by equating subjects or by holding variables constant. In equating subjects, each group is controlled for all variables except the treatment variable. For example, if it were felt that gender and age of subject would influence taste test results, each group would have the same gender-and-age proportions. Matching by holding variables constant involves creating constancy for all groups. Again referring to a taste-testing experiment, the gender variable could be controlled by using only all male or all female subjects. If time of day is an important extraneous variable that could affect one’s “taste buds,” then the experiment should be conducted at approximately the same time on successive days. Or, if time of day affects subjects’ purchasing behavior (purchase occurs after a certain time), then the experiment should not be conducted before that time; McDonald’s tested pizza for the dinner menu and required that tests be conducted after 4:00 pm. Similarly, using the same facilitator in all versions of the study may control experimenter effects.

4. **Counterbalancing** is a technique used to control confounding, or the tangling effects of two or more levels of a treatment variable (for two or more treatment variables), used in

experimentation. An illustration should make this point clearer. Suppose that a marketing researcher is interested in conducting a series of taste-testing experiments for a new soft drink. Subjective interpretations of, say, "sweetness" may well vary from subject to subject. A control procedure called counterbalancing would have each subject taste each of two drinks on the assumption that ratings will be expressed in terms of differences in sweetness over each subject. To avoid "ordering" effects on responses, the new and the control drinking would be presented in randomized order or one-half of the group would follow the sequence "established-new" while the other one-half would use the sequence "new-established." To reduce carryover tendencies, the subject would be asked to take a sip of water between testing trials.

5. **Randomization** provides assurance that known and unknown extraneous factors will not cause systematic bias. Statisticians have made a major contribution to experimental design in the development of statistical models that feature randomization over uncontrolled variables. Consequently, it is assumed that the effects of the extraneous variables will affect all groups in an experiment to the same extent. In theory, randomization is supposed to accomplish the conversion of all irrelevant sources of possibly systematic variability into unsystematic variability, i.e, into random error (Brown and Melamed, 1990, p. 3).

The above discussed techniques for control of extraneous variables are a key part of all experiments, however the fact remains that confounding effects can never be entirely eliminated.

Sources of Invalidity

In Chapter 3, we briefly introduced experimental errors as the extraneous forces that can affect the outcome of an experiment. Each extraneous force potentially has a bearing on the validity of an experiment and, consequently may threaten the validity of the results.

In the context of experimentation, the term validity refers to the extent to which we really observe (or measure) what we say we observe (or measure). Four distinct types of validity have been identified (Cook and Campbell, 1990, chap. 2):

- Statistical conclusion
- Internal
- Construct
- External

A necessary condition for inferring causation is that there be covariation between the independent variables. **Statistical conclusion validity** involves the specific question as to whether the presumed independent variable, X, and the presumed dependent variable, Y, are indeed related (Rosenthal and Rosnow, 1991, chap. 3). After it has been determined that the variables covary, the question arises as to whether they are causally related. This is the essence of **internal validity**. A given experiment is internally valid when the observed effect is due solely to the experimental treatments and not due to some extraneous variables. In short, internal validity is concerned with how good the experiment is, as an experiment. The third type of validity, **construct validity**, is essentially a measurement issue. The issue revolves around the extent to which generalizations can be made about higher-order constructs from research operations and is applicable to causes and effects. Because construct validity is concerned with generalization, it is a special aspect of external validity. **External validity** however, refers to the ability to generalize a relationship beyond the circumstances under which it is observed. That is, external validity is concerned with how good an experiment is: to what degree can conclusions of an experiment be applied to and across populations of persons, settings, times, and so on.

To a large extent, the four kinds of validity are not independent of each other. That is, ways of increasing one kind may decrease another kind. Consequently, in planning an experiment it is essential that validity types be prioritized, and this varies with the kind of research being done. For applied marketing research, it has been observed that the priority ordering is internal, external, construct of the effect, statistical conclusion, and construct of the cause (Cook and Campbell, 1990). Accordingly, we now examine those extraneous factors that affect internal and external validity. Construct validity is discussed in Chapter 9, and factors affecting statistical conclusion validity are covered throughout the later sections of this book.

Internal Validity

Internal validity is concerned with whether or not the observed effect is due solely to the experimental treatments or due to some other extraneous variables. The kind of evidence that is required to support the inference that independent variables other than the one(s) used in an experiment could have caused the observed effect(s) varies, depending on the independent variable(s) being investigated. However, there are some general classes of variables affecting designs that deserve mention. The following factors affect internal validity:

1. *History.* An extraneous event that takes place between the pre-measurement and post-measurement of the dependent variable has an impact on the results.
2. *Maturation*. The results of an experiment are contaminated by changes within the participants with the passage of time.
3. *Testing*. A prior measurement can have an effect on a later measurement.
4. *Instrumentation*. Changes in the measuring instrument or process, including interviewers' instructions, over time, affects the results.
5. *Selection*. Different kinds of research subjects have been selected for at least one experimental group than have been selected for other groups.
6. *Mortality*. Different types of persons drop out from experimental groups during the course of an experiment.
7. *Statistical Regression*. This error may arise when experimental groups have been selected on the basis of extreme pretest scores or correlates of pretest scores.
8. *Interactions with Selection*. Independent effects of the interaction of selection with history and maturation have an impact on experimental results.

Example: To illustrate each of these factors, suppose that a controlled experiment is setup for salesperson retraining. Assume one group of salespeople had taken a retraining course during a three-month period, while another group had not been retrained. The brand manager of a particular brand of detergent wants to determine if retraining is a producer (i.e., a cause) of sales performance. During the three-month period after retraining, sales of the detergent showed an unusually large increase.

History deals with events outside the design that affect the dependent variable. History is therefore comprised of the producers that are extraneous to the design. For salesperson-retraining, the level of competitive promotion and advertising, the overall level of demand, or any one of many other producers may have changed substantially in some territories. Clearly, the longer the time period involved, the greater the probability that history will significantly affect the results.

Maturation is concerned with the changes that occur with the passage of time in the people involved in the design. For example, as time passes, salespeople gain more experience in selling and hence know their customers better, and the customers become better acquainted with the product. Similarly, some biological and psychological changes within salespeople and customers over time can affect the performance of either.

Testing effect has to do with the effect of a first measurement on the scores of a second measurement. Familiarity with a measurement (i.e., a test) can sometimes enhance performance because items and error responses are more likely to be remembered at later measurement sessions. Suppose that the attitudes of all salespeople toward their job, performance, the company, and so on, had been measured prior to and after the retraining program. There is the possibility that some responses obtained after the retraining program were due to the salespeople remembering responses given on the first measurement.

Instrument effect refers to the changes in the measuring instrument or process that may affect the measurements obtained. This can occur in many ways, including changes in cost-value (a drop in home prices) over time, the learning process on the part of the investigators, a change in investigators, or simply boredom or fatigue may affect the measurements and thus the interpretation of results. A related issue arises when the research subjects themselves self report or record responses. In this case, there exists the possibility of a confounding of the instrument with the experimental treatment because the subjects' understanding or awareness of the variable being measured may have changed.

Selection is concerned with the effect of the selection procedure for the test and control groups on the results of the study. If the selection procedure is randomized, the effect will be a measurable random variation. However, if self-selection or some other nonrandom (purposive) procedure occurs, the results will be affected in a non-measurable manner. Sizable systematic errors may well result. The concern, of course, is that the resulting groups may differ on important characteristics, and these differences may be influencing the dependent variable.

Statistical regression dictates that under conditions where measures are unreliable, high pretest scorers will score relatively lower at the posttest and low pretest scorers will score higher. There is a tendency with repeated measures for scores to regress to the population mean of the group (Cook and Campbell, 1990).

Some of the foregoing error sources affecting internal validity can interact with selection to produce forces that might appear to be treatment effects. Selection-maturation results when experimental groups mature at different speeds. Selection-history can occur when the experimental groups come from different settings. For example, the salespeople receiving the retraining all come from one region of the country, whereas those not retrained come from some other region. In this situation, each group may have a unique local history that might affect outcome variables.

External Validity

External validity is concerned with whether or not conclusions are externally valid, that is, whether conclusion apply to and across populations of persons, settings, times, and so on. The following are sources of error in external validity:

1. *Reactive Effect of Testing (Interaction)*. Pre-measurement may have an impact on the experimental subject's sensitivity or responsiveness to the treatment variable.
2. *Reactive Effects of Experimental Situation*. Experimental subjects react to the experimental situation (such as setting, arrangements, and experimenter).
3. *Interaction of History and Treatment*. Measuring the dependent variable at a point in time that does not reflect the actual effect of the independent variables can skew results.
4. *Interaction of Selection and Treatment*. The method of selection affects the extent to which the measured effect can be generalized to the population of interest.

A frequent problem in research design is that a "before" measurement is desired, but it is recognized that making such a measurement may alert the subjects that they are participating in a study. If an experiment is being conducted on the impact of advertising, a pre-measurement may sensitize an experimental subject to pay particular attention to a company's advertising, product(s), or both. Although this potential source of error may seem the same as that from testing, it is distinctly different that it is an interaction of testing and treatment. Despite the plausibility, the empirical evidence of a pretest treatment interaction bearing on attitude change research is, in general, meager.

Experimental situations refers to "all aspects of the experiment which cause the subject to perceive, interpret, and act upon what he believes is expected or desired of him by the experimenter (Sawyer, 1975, p. 20). This includes the setting within which the experiment is conducted, the arrangements made for the experiment such as the apparatus used, and the presence and behavior of an experimenter. These reactive effects involve both experimenter and subject effects (Christensen, 2004, Chap. 7).

To a large extent, the effects of demand characteristic depend on the roles and expectations adopted by the experimenter and the subjects, particularly when subjects become aware of, or believe they know, the experimental hypothesis. Ideally, the investigator would seek subjects who are faithful or naïve (i.e., do not know the hypothesis).

Unfortunately, you can never be absolutely certain that demand characteristics will not be present in the chosen experimental procedures. Research design, measurement of dependent variables, and use of procedures (e.g., deception and natural environments) will reduce such demand characteristics (Sawyer, 1975, pp. 25-8). In addition, manipulation and confounding checks can be used to investigate the plausibility of demand characteristics (Perdue and Summers, 1986, p. 38).

History-treatment interaction can affect external validity when the dependent variable measurement and the treatment variable measurement are not time generalized. For example, if an experiment were conducted on the effectiveness of advertising and this experiment was run at a time close to any holiday, then a question would remain as to whether the same cause-effect relationship would exist at some other time during the year. Moreover, with some independent variables their effects may be long-run in nature, but typically dependent variable measurement occurs at one point in time.

Selection-treatment interaction asks if the observed effects can be generalized beyond the groups used to establish the relationships? Even when research subjects belong to the target group of interest, the recruitment approach used may limit any generalizations to only those who participated in a given experiment. Experiments, including laboratory and Internet based experiments, rely heavily upon volunteers, paid or otherwise. There is some evidence that volunteer experimental subjects differ from non-volunteer subjects in more or less systematic ways: they are better educated, higher social class, higher need for social approval, more sociable, more intelligent (Rosenthal and Rosnow, 1975, pp. 88-90).

Models of Experimental Design

A number of experimental designs have been developed to overcome and reduce the various sources of invalidity. Experimental designs can be categorized into two broad groups—classical and statistical. Classical designs consider the impact of only one dependent variable at a time, whereas statistical designs allow for examining the impact of two or more independent variables.

Classical Designs

The major types of classical designs were mentioned earlier in this chapter. Pre-experimental designs are so called because there is such a total absence of control and are of value in exploratory research, but are of minimal value in establishing causality. Quasi-experimental designs involve control but lack random assignment of subjects, as required for true experiments. Where any given design fits in this classification spectrum will depend on whether the treatment variable has been deliberately manipulated, the nature of control, and whether there has been random assignment of subjects to experimental groups.

The following notational system will be used in the discussion of classical experimental designs:

- *X* represents the exposure of test groups to an experimental treatment of a producer or event whose effect is to be observed and/or measured.
- *O* refers to a measurement or observation taken.
- *R* indicates that individuals have been assigned at random to differing treatments.
- Movement from left to right indicates a sequence of events. When *O's* and *X's* are found in a given row, they are to be interpreted as having occurred in sequence and to the same specific individual or group. Vertical arrangement of symbols is to be interpreted as the simultaneous occurrence of the events that they denote.

Three classes of designs seem to fit for pre- and quasi-experiments:

- Time-series and trend designs
- Cross-sectional designs
- Combinations of the two previous classes

Time-Series and Trend Designs

Time-series and trend designs are similar in concept, yet their differences in implementation and analytic procedures warrant a brief discussion. A time-series design involves obtaining data from the same sample (or population) for successive points in time. The common method of gathering primary data of this kind is to collect current data at successive intervals through the use of a continuous panel. One may, however, collect current and retrospective data from respondents during a single interview. If the latter technique issued, respondent recall must be relied on to reconstruct quasi-historical data. An alternative method of obtaining data for past periods is to use secondary sources, when available.

Trend data differ from time-series data in that they are obtained from statistically matched samples drawn from the same population over time. Current data are gathered from each successive sample.

Both time-series and trend data are used to investigate the existence and nature of causal relationships based on associative-variation and sequence-of-events types of evidence. While individuals or households are the most commonly used sample units, data are also obtained from retail stores, wholesalers, manufacturers, and other units.

Time-series and trend designs involve at least one treatment and a subsequent measurement and can involve a large number of measurements with several interspersed treatments over a long period of time. We now describe and discuss four types of time-series and trend designs.

After-Only without Control Group

This design is often termed a "tryout" or "one-shot case study." It is the simplest of all designs, as it involves only one nonrandomly selected group, one treatment, and one measurement. Symbolically, it may be diagrammed as follows:

$$X \qquad O \tag{1}$$

The many weaknesses of this design may be illustrated by applying it to the salesperson-retraining problem. Assume that no prior measurement of sales volume of the salespeople to be retrained had been made. A group of salespeople are selected by a nonrandom method and retrained (X); a measurement (O) is made after the retraining.

Since no prior measurement of sales volume was made for each of the retrained salespeople, there is no method, short of making assumptions as to what would have happened in the absence of retraining, of estimating what the effect of retraining was. The effects of history, maturation, and selection are all potentially substantial and non-measurable. Because of these limitations the use of this design is to be avoided if at all possible.

Before-After without Control Group

This design is the same as (1) with the addition of a "before" measurement. In its simplest form it is shown as

$$O_1 \qquad X \qquad O_2 \tag{2}$$

More advanced forms of this design include multiple measurements (O's) before and after the single measurement.

Although design (2) is relatively weak, it is frequently used. It is a decided improvement over (1) in that the apparent effect of the treatment, $O_2 – O_1$, is measured. In terms of the salesperson-retraining illustration, a measurement (O_1) of sales volume of the retrained salesperson is made for the quarter of the preceding year.

Multiple Time Series

In using a time-series design the possibility of establishing a control group should always be investigated. It may be possible to find a comparable, if not equivalent, group to serve as a control against which to compare the results of the group that underwent the treatment involved. This design may be diagrammed as

$$\begin{array}{ccccccccc} O_1 & O_2 & O_3 & O_4 & X & O_5 & O_6 & O_7 & O_8 \\ O'_1 & O'_2 & O'_3 & O'_4 & & O'_5 & O'_6 & O'_7 & O'_8 \end{array} \qquad (3)$$

where the primed *O's* represent measurement of the control group. Note that the individuals constituting the groups were not selected at random. It may be possible, however, to select at random the group that will receive the treatment.

This design can easily be adapted to the sales-retraining evaluation problem. If it is assumed that the sales volume of each of the salespeople is measured during each period as a matter of course anyway, a group could be selected for retraining after period 4. Either a comparable group or all the rest of the salespeople could be selected as the control group. After the sales of both groups in the periods after the training had been measured, the apparent effect of the training would be shown by comparing the differences in average sales volume for the two groups before and after treatment.

Cross-Sectional Designs

Cross-sectional designs involve measuring the dependent variable of interest for several groups at the same time, the groups having been exposed to different levels of treatments of the treatment variable whose effect is being studied. Cross-sectional designs may be viewed diagrammatically as follows:

$$\begin{array}{cc} X_1 & O_1 \\ X_2 & O_2 \\ X_3 & O_3 \\ . & . \\ X_n & O_n \end{array} \qquad (4)$$

Examples of frequent applications of this design are studies of the effect of such variables as price, package design, or level of advertising on sales in different geographic areas. This design can be used when direct manipulation of the producer involved is not possible or practical. When this is the case, the design is being used in a natural experiment. The effect of the different levels of treatment is measured by determining the degree of association between producer and product. The techniques that can be employed are discussed in later chapters.

A variation of this design is the static-group comparison. This is a design in which a group exposed to a treatment is compared with one that was not:

$$\begin{array}{cc} X & O_1 \\ & O_2 \end{array} \qquad (4a)$$

History may play a critically important role in cross-sectional designs. There could be a sizable differential effect of extraneous variables between the groups being measured.

Combination Cross-Sectional, Time-Series Designs

A number of designs employing a combination of time-series and cross-sectional treatment and measurement may be used in observational studies. The multiple-time-series design (3) could be considered a combination of the two types, as it involves measurements of a product for different groups at the same time as well as for the same group over time.

Combination designs are well adapted for use with consumer-panel data. One commonly used design is the ex post facto test-control group. In this design the test and control group are not known until after the treatment has been administered. This design is illustrated as follows:

$$\begin{array}{lll} O_1 & X & O_3 \\ O_2 & & O_4 \end{array} \qquad (5)$$

This design is widely used in connection with testing the sales effectiveness of price changes, "deals," and advertising. Data on the sales of the brand of interest are reported regularly by the members of a continuous consumer panel. After a given advertisement is run (X), panel members may be questioned to determine whether or not they saw it. Those who saw it are part of the test group, since they have had exposure to the treatment involved. Those who did not see the advertisement become a part of the control group. The apparent effect is determined by comparing the difference in test and control-group purchases before with that after the advertising was run.

True Experimental Designs

As previously mentioned, for true experiments two kinds of investigator intervention are required:

- Manipulation of at least one assumed causal variable
- Random assignment of subjects to experimental and control groups

Through use of a random selection procedure, systematic errors due to selection are eliminated and the effects of the many extraneous variables tend to be equalized between the experimental and the control groups as the size of these groups increases. Random selection permits the use of inferential statistical techniques for analyzing the experimental results.

Three single-variable experimental designs are described below.

After-Only with Control Group

The simplest of all experimental designs is the after-only with control group. It requires only one treatment and an "after" measurement of both the experimental and the control group. Yet it has the essential requirements of the true experiment: manipulation of at least one variable and randomly selected test and control groups. It is illustrated as follows:

$$\begin{array}{lll} R & X & O_1 \\ R & & O_2 \end{array} \qquad (6)$$

The absence of a "before" measurement (or "pre-test") is a feature that concerns many researchers about this design. Such a measure is not actually essential to true experimental designs because respondents are randomly assigned, controlling for extraneous variable effects. This design is of major interest, therefore, when "before" measurements are impractical or impossible to obtain and/or when the testing and instrument effects are likely to be serious.

A common application of this design is in the testing of direct-mail advertising. Random-sampling procedures are used to select an experimental and a control group. Direct-mail pieces are sent to the experimental group and withheld from the control group. "After" measurements of sales to each group are made and the differential is determined ($O_1 - O_2$).

Before-After with One Control Group

If "before" measurements are added to design (6), we arrive at the following configuration:

$$\begin{array}{llll} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \end{array} \qquad (7)$$

This design is very similar to that of (5) but with an important difference: the experimental and control groups are randomly selected, rather than self-selected. Most of the sources of systematic error are controlled in this design.

This design offers three ways to evaluate the effect of the treatments: $O_2 - O_1$, $O_2 - O_4$, and $(O_2 - O_1) - (O_4 - O_3)$. If the results of each of these evaluations are consistent, the strength of our inferences about the effect of the experimental treatment is substantially increased.

An example of the use of this design is in advertising tests that use a dual cable television system with two consumer purchase panels (one from the subscribers to each cable). "Before" measurements can be made on the test and control panels, an experimental advertising treatment introduced on the test cable, and "after" measurements made for both panels.

There are many classical designs. In this section, we have discussed the more common designs. A useful way to summarize what we have discussed is to look at each in terms of sources of potential error and whether such sources are controlled for, as shown in Table 8-1. More detailed discussions are found in Cook and Campbell (1990), Campbell and Stanley (1966), and Shadish, Cook and Campbell (2002).

Statistical Designs

For the most part, statistical designs are "after-only" designs (6) in which there are at least two treatment levels. In addition, such designs can examine the effects of more than one independent variable. Two principal aspects of statistical designs are the following:

- the experimental layouts by which treatment levels are assigned to test objects
- the techniques that are used to analyze the results of the experiment

We will now briefly discuss now the major types of layouts used to obtain data. These are discussed in more depth in Brown and Melamed (1990) and other specialized experimental design texts such as Field and Hole (2003) and Shadish, Cook and Campbell (2002). The analysis techniques, known generically as analysis of variance and covariance, are discussed in later chapters.

Completely Randomized Design

The completely randomized design is the simplest type of statistical design. In this design, the experimental treatments are assigned to test units on a random basis. Any number of treatments can be assigned by a random process to any number of test units.

> *As an illustration, suppose that a marketer is interested in the effect of three versions of an ad on customer preferences. The marketer has been able to obtain an email list of customers and has corporate approval to run an experiment involving the three ads.*

This simple design would entail randomly showing one of the ads to each survey respondent and measuring their degree of liking of the ad they viewed. This could be described as classification by a single factor. This design is most applicable when it is believed that extraneous variables will have about the same effect on all test units, and when the marketer is interested in only one independent variable. Moreover, there is no absolute guarantee that randomization will keep extraneous influences in check. Suppose that our marketing researcher were interested in the effect of other variables such as the introductory message or gender of the salesperson appearing in the ads. Or, suppose that the researcher would like to generalize the results of the experiment to other sizes of ads or products. It may be preferable to ask many rather than few questions of nature if the researcher would like to establish the most general conditions under which the findings are expected to hold. That is, not only may single-factor manipulation be difficult to do in practice, but it may be inefficient as well. We now discuss somewhat more specialized experimental designs, all of which are characterized by two or more variables of classification.

Table 8-1 Sources of Invalidity and Selected Experimental Designs

	Internal								*External*		
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Others	Interaction of Testing and *X*	Interaction of Selection and *X*	Reactive Arrangements
One-Shot Case Study *X O*	-	-				-	-			-	
One-Group Pretest-Posttest Design *O X O*	-	-	-	-	?	+	+	-	-	-	?
Time Series *O O O X O O O*	-	+	+	?	+	+	+	+	-	?	?
Multiple Time-Series <u>*O O O X O O O*</u> *O O O O O O*	+	+	+	+	+	+	+	+	-	-	?
Static-Group Comparison *X O* *O*	+	?	+	+	+	-	-	-		-	
Nonequivalent Control Group Design *O X O* *O O*	+	+	+	+	?	+	+	-	-	?	?
Posttest-Only control Group Design *R X O* *R O*	+	+	+	+	+	+	+	+	+	?	?
Separate-Sample Pretest-Posttest Design *R O (X)* *R X O*	-	-	+	?	+	+	-	-	+	+	+
Pretest-Posttest Control Group Design *R O X O* *R O O*	+	+	+	+	+	+	+	+	-	?	?
Solomon Four-Group Design *R O X O* *R O O* *R X O* *R O*	+	+	+	+	+	+	+	+	+	?	?

Note: In the tables, a minus (-) indicates that the factor is not controlled, a plus (+) indicates that the factor is controlled, a question mark (?) indicates a possible source of concern, and a blank indicates that the factor is not relevant.

SOURCE: Adapted from Campbell, D. T., &Stanley, J.C., Experimental and Quasi-Experimental Designs for Research. Copyright ©1963 by Houghton Mifflin Company. Adapted by permission.

Factorial Designs

A factorial experiment is one in which an equal number of observations is made of all variable combinations. The variables must have at least two levels or categories each. In essence, the factorial design is one that has combined two or more completely randomized designs into a single experiment. This type of experiment enables the researcher to study possible interactions among the variables of interest. Suppose we return to our Ad illustration but now assume that the researcher is interested in studying the effects of two variables of interest: Ad Spokesperson type (at three levels – Male, Female, Cartoon Character) and Message Type (at two levels – Benefits Message and Humorous Message). The design is shown in Table 8-2. Note that each combination of M_iS_j occurs only once in the design. While the plan still is to use a single sample for the experiment, the researcher intends to randomize the presentation of each combination among the various respondents (each receiving one Ad).

In the factorial experiment we can test for all main effects (i.e., spokesperson types, messages), and in this case, where we have replicated each combination, for the interaction of the variables as well. If the interaction term is significant, ordinarily the calculation of main effects is superfluous, since the experimenter will customarily be interested in the best combination of variables. That is, in market experimentation the researcher is typically interested in the combination of controlled variables that leads to the best payoff in terms of preference, sales, market share, cash flow, or some other measure of effectiveness.

Table 8-2 Factorial Design—Advertising Experiment

	Ad Spokesperson Gender		
Message	Male	*Female*	*Cartoon Character*
Benefit Statements	M_1S_1	M_1S_2	M_1S_3
Humorous Statements	M_2S_1	M_2S_2	M_2S_3

Latin Square

Latin-square designs are multivariable designs that are used to reduce the number of observations that would be required in a full factorial design. In using Latin-square designs the researcher is usually assuming that interaction effects are negligible; in so doing, all main effects can be estimated by this procedure.

As an illustration of a Latin-square design, suppose that the researcher is interested in three variables (each at four levels) on store sales. For example, the researcher may be interested in the effect of shelf placement on the sale of energy drinks and be able to manipulate the following variables for product placement:

A: shelf height—four levels—knee level, waist level, eye level, reach level
B: shelf facings—four levels—25%, 50%, 75%, and 100% of total width of gondola
C: shelf fullness—four levels—25%, 50%, 75%, and 100% of total height of gondola section

If the researcher were to run a full factorial experiment, with one replication only, there would be (4)3 = 64 observations required. By using a Latin-square design only 16 observations are required (with estimation of main effects only).

Table 8-3 Latin-Square Design—Energy Drink Shelf Placement Experiment

	Variable B—Shelf Facing			
Variable A—Shelf Height	*B_1*	*B_2*	*B_3*	*B_4*
A_1	C_1	C_2	C_3	C_4
A_2	C_4	C_1	C_2	C_3
A_3	C_3	C_4	C_1	C_2
A_4	C_2	C_3	C_4	C_1

Table 8-3 shows one possible Latin-square design for this experiment. Notice in Table 8-3 that each level of treatment C (shelf fullness) appears once in each row and each column. See Edwards, 1968, pp. 175-77 for more advanced Latin-Square designs.

Randomized Block Design

Randomized-block designs, also known as treatments by blocks designs, represent a frequently used experimental framework for dealing with multivariable classifications. These designs are typically used when the experimenter desires to eliminate a possible source of uncontrolled variation (a nuisance variable) from the error term in order that the effects due to treatments will not be masked by a larger-than-necessary error term. For example, suppose that our researcher were interested only in the effect of ad spokesperson effect on product performance, but had designed the experiment so that more than one product was used in the study. The effect of product type could influence preference, and the experimenter might wish to remove this effect from the error term by "blocking" on product types. That is, each product would be considered a test unit and each level of spokesperson would be tested in each product.

To illustrate, suppose researchers were interested in examining three levels of spokesperson for each of four products. Respondents could be randomly assigned to the respective treatments, or even shown multiple treatments if measurement and learning effects (distracters of internal validity) were ignored. Results could be summarized in the form shown in Table 8-4, where we are dealing with a two-variable classification and can separate the block effect from the error term. Thus, if genuine treatment effects are present, this type of design will be more likely to detect them than a single-variable classification in which the block effect would become part of the error term.

Table 8-4 Randomized-Block Design—Spokesperson Experiment

	Treatments—Spokesperson		
BLOCKS—Products	*Level 1 Male*	*Level 2 Female*	*Level 3 Cartoon*
1	X_{11}	X_{12}	X_{13}
2	X_{21}	X_{22}	X_{23}
3	X_{31}	X_{32}	X_{33}
4	X_{41}	X_{42}	X_{43}

Covariance Design

Covariance designs are appropriate in situations where some variable affects response but is not subject to control during the experiment. For example, if test units consist of human subjects and the response variable is the number of correct identification of trademarks that are shown on a projection screen (where such factors as length of exposure and clarity of focus are varied), it may be that response is affected by the general intelligence level of the viewing subject. Suppose that it is too costly to screen subjects, and only those with approximately the same intelligence quotient are selected. We shall assume that the researcher is able to measure each subject's IQ.

In this type of situation, the researcher may use covariance analysis. Roughly speaking, the computational procedure is similar to a regression problem. The researcher, in effect, determines the effect on response resulting from differences (in IQ) among test units and removes this influence so that the effect of the controlled variables can be determined independently of the effect of test differences on response.

Recapitulation

As noted earlier, the study of experimental design is basically the study of two things:

1. various experimental layouts, such as single factor, factorial, Latin-square, and randomized block designs
2. analysis of variance and covariance techniques for testing whether the various treatment effects are significant

Many other kinds of design layouts, including fractional factorial designs, balanced incomplete blocks, hierarchical designs, split-plot designs, and partially balanced incomplete blocks are available. We have only described briefly the characteristics of some of the basic statistical designs.

We have not covered the issue of how to select an experimental design. The design chosen must be congruent with the research question and capable of implementation with resources that are available. Such issues as the number of independent variables, the sources and number of extraneous variables, the nature of the dependent variable, the number of subjects available for participation in the experiment and other methodological issues will have a bearing on what might be an appropriate design to use. Increasingly, marketing researchers use personal computer software to aid in this selection.

The Panel as a Natural Experimental Design

The normal course of operation of a consumer panel generates a continuing set of natural experimental data. Customer responses to changes in any of the controllable or environmental variables affecting purchase decisions are recorded in the normal process of conducting the panel. Audience and deal panels provide similar response measurements.

Time-series, cross-sectional, and combination cross-sectional, time-series designs are all inherent in panel data. To illustrate their application, suppose that we have increased the price of our product in selected territories. We can analyze the price-increase effect, at either the aggregated or individual household level, using the data from those territories in which price was increased with either the after-only without control group or the before-after without control group designs [classical designs (1) and (2)]. A cross-sectional analysis may be made by comparing, for a given period after the increase, the purchase data for the territories in which the price was raised with those in which no change was made [classical design (4)]. A preferable approach here would be to use a combination cross-sectional, time-series design and compare the change in purchases before and after the price increase in the territories in which price was not changed (control group). Such a study could employ either classical design (3) or (5).

The limitations of each of these designs discussed earlier apply when they are used with panel data as well. A major difficulty, or course, is in sorting out the effect of the price increase from the extraneous producers affecting purchases over time and among territories. In this illustration selective price increases by territory would only have been made in response to differing conditions among the sales territories (a price increase by competitors, higher levels of demand, etc.). History variables must therefore be analyzed carefully in using panel data.

Controlled Experimental Designs Using Panels

The controlled experimental design in conjunction with a panel is most often applied to market tests of prospective new products, different levels of promotion, new campaign themes, price changes, and combinations of two or more of these variables. Consider, for example, a market test of a general price increase. The requirement of random selection of test and control groups can be met by selecting territories at random in which to raise prices. The remaining territories automatically constitute the control group. Depending on the kinds of information desired, an after-only with control group, or a before-after with one control group [classical design (7)] may be used.

The general advantages and limitations of these designs were discussed earlier. We must, however, consider the limitations that arise from the use of the panel for measurement, applicable to both natural and controlled experimental designs.

The Limitations of Continuous Panels

Although panels can provide highly useful marketing information that is difficult to obtain by alternative research methods, there are some important limitations. The first of these limitations involves selection and stems from the difficulty of obtaining cooperation from the families or firms selected in the sample and the resulting effect on the degree of representativeness of the panel. To be most useful for drawing inferences about the population being studied, the sample should be drawn by a random process. The sample of families to comprise a consumer purchase panel may be chosen randomly, but the typical panel has experienced a high refusal rate during the period of establishment and a high attrition rate, once in operation. Panels require a great deal of effort to recruit and maintain.

Evidence indicates that the characteristics of both those families who refuse to participate and those who later drop from the panel are different from those who agree to participate and remain. To reduce the bias introduced by such nonrandom attrition, replacements are typically chosen from families with the same demographic and usage characteristics as those lost from refusals and dropouts.

An additional source of bias is in testing effect arising from continued participation on the panel. Since the individual is undoubtedly conditioned to some extent by the fact that data on purchases are reported, panel members may become atypical in their purchase behavior as a result of being a part of a panel. In short, being a panel member may influence what products and brands are purchased as the panel member is sensitized.

Panel data may also be systematically biased through instrument effects. The majority of panels use diaries for reporting. These are self-administered, structured questionnaires. If filled out properly and submitted on schedule, the information is relatively accurate and inexpensive to obtain. However, in those cases where there is incomplete data possible biases are introduced and the total amount of data is decreased.

Despite these limitations, the use of data from panels has become widespread, particularly in online research. If the panel is administered carefully, the resulting data are important additions to the information required for making sound marketing decisions.

DEVELOPING EXPERIMENTAL DESIGNS IN ONLINE SURVEYS

The Qualtrics.com survey software provides advanced survey capabilities and logic, including experimental designs and data analysis. This advanced technology now provides a simple and efficient way to design experiments into online surveys. In addition, integrated panel management capabilities make possible accurate tracking, profiling and monitoring of response history, response timing and response progress for each respondent. We discuss the capacities of Qualtrics survey software to solve research problems through management of question flow, logic, and randomization. These features provide fundamental tools for increasing control of field experiments. Once we have provided a view of the experimental design methods and examples for data collection, we proceed to a discussion of the integration of database information into experimental designs.

Structuring Advanced Questionnaires

Advanced online questionnaires have at their foundation five essential building blocks. The researcher who understands these building blocks will produce questionnaires that are better organized and capable of superior functionality. This functionality can include the development of and presentation of experiments. These building blocks require a clearly identified and stated set of research objectives for the questionnaire and the experiment:

1. Identify the populations to be sampled, along with experimental and control groups
2. Map the logic and flow of the experiment, including randomization
3. Identify the organization of blocks of questions if needed for survey flow, branching or experimentation
4. Specify the questions (and question types) within each block
5. Define sample size quotas that are required for appropriate analyses

Note that we have not yet considered the important issues of measurement, including construct validity, the wording of questions, or controlling survey content. These topics are discussed in the next chapter.

Survey Realism, Interest and Respondent Involvement

Experimental design is all about presenting specific treatments to the respondent and measuring the effects. With online experimental designs the boundaries of what is possible for treatments has been greatly expanded through multi-media content.

Enhancing Survey Content: Graphics, Audio, Video and Interactive Experiences

Graphics increase interest and provide realism to a survey. The advantage of graphic visualizations over text descriptions cannot be questioned. A corporate logo at the top of the survey may provide identity and authenticity to the questionnaire. Similarly, graphics can provide an experiential view of product and brand treatments for concept tests.

Prior to online surveys, the use of audio creatives (songs, jingles, messages with varying content, etc.) in research was limited to telephone and personal interviews, or focus group settings. Online technology facilitates the use of songs, commercials, legal arguments, and other verbal, melody, or tone-based treatments in online concept tests and experiments.

Video files (MOV, MP4 and FLASH) provide rich, realistic portrayals of products and can be added easily to online experiments for testing (Figure 8-4). The increased adoption of high-speed Internet continues to expand the type and amount of information that can be tested.

Figure 8-4 A Nokia Flash Commercial in an Evaluation Question

While interactive technology is limited in use within mainstream market research, it is used on many entertainment Web sites and is applicable to highly technical and youth market segments who are accustomed to this high-involvement graphics approach. Qualtrics accepts browser interfacing programs that may be used to provide interactive Web Services content (programmed using Flash, Java Script, PHP or Visual Basic). These tools give researchers the option of creating and adding to a survey, custom programs that control their own experimental treatments and measurements.

As examples, these programs could control treatments for sophisticated experiments using Latin-square or d-optimal Federov designs that present options in choice based conjoint analysis. Control could be developed for unique ways of displaying, timing and rotating text or graphics. Similarly, a researcher could for each survey respondent randomly select, say, 5 open ended text responses from other respondents who have already completed the survey, and then ask the current respondent to rank the 5 extracted text statements in order of preference. Qualtrics provides through these API and Web Services tools, open access to multiple databases in real-time and an interface for programming tools. The researcher is thereby unconstrained, except in their own creativity.

Creating Dynamic Questions through "Piping"

Piping is one method of customizing surveys. Piping generally refers to the movement of information from one location to another within the survey. Information may include question text, answer text, graphics, or database information about the respondent. It may also include the text of answer options selected, or those not selected. The point of application is that the respondent's questionnaire flow may be customized to reflect their individual choices or options they did not choose, so that additional topic specific in-depth information can be collected.

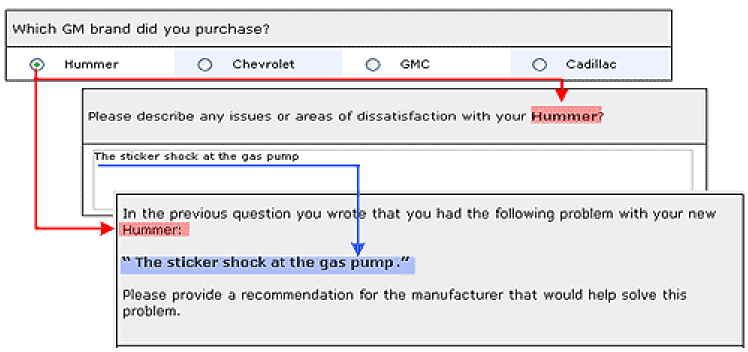
> *Travelocity customers receive a survey a few days after they return from a trip booked online. The invitation letter and questionnaire are not of the "Dear Traveler" variety. Instead, the customer is addressed by name and is asked specific questions about their flight to Dallas on Delta Airlines. If they booked a rental car, the block of questions about satisfaction with the service of their rental car company (identified by name) is included in the survey flow. Question blocks are also customized for hotels and cruises booked for the trip. In short, the online customer database is interfaced with the survey database through a simple API (application program interface) and the questionnaire is totally customized to ask about specific aspects of their personal travel experience.*

Panel information linked from external databases can be used in a variety of ways. For example, if the respondent is evaluating a given brand, factual data about a product category, information about the category leaders, or even information about the respondent's past purchases could be piped from a database to allow the respondent to make more accurate relative comparisons given their purchase history.

As a simple example of piping, consider pieces of information that are piped based on the respondent's answers to previous questions. The respondent first chooses the brand Hummer as the GM car they had previously purchased. The text "Hummer," piped to an open ended text question about dissatisfaction the car. Here, the respondent typed, "The sticker shock at the gas pump", a phrase that is provide to provide context and detail in the final question that asks the customer to make recommendations to General Motors.

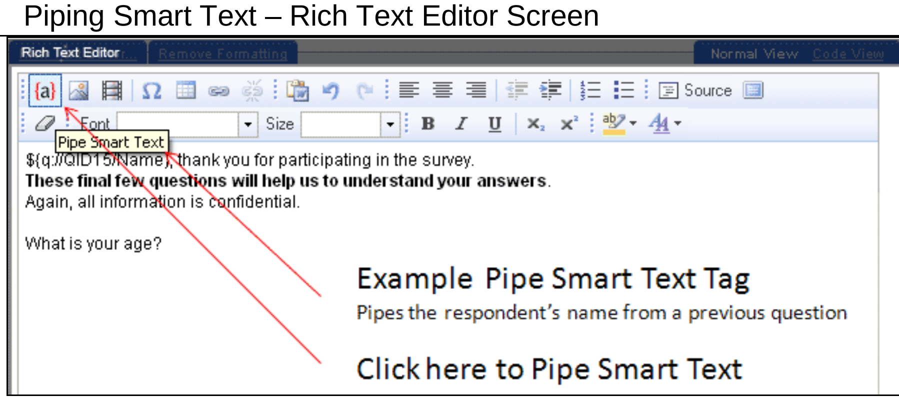
Other uses of piping might include concept or logo tests where respondents select a company logo or other concept-related graphic. Similar to the above example, the selected graphic would follow the respondent as needed in the survey, either as part of the question text or as a possible answer choice. At the advanced end of the spectrum, personal data for a given respondent may be read from a CRM (customer relationship management) database, such as in the Travelocity example cited above, and then piped into questions or answers within the questionnaire.

Figure 8-2 Text Piping Example



The Pipe Smart Text {a} option is accessed by clicking on the text of the question and selecting {a} in the Rich Text Editor (Figure 8-3).

Figure 8-3 Piping Smart Text – Rich Text Editor Screen



Creating Dynamic Surveys with Skip Logic, Branching and Looping and Control

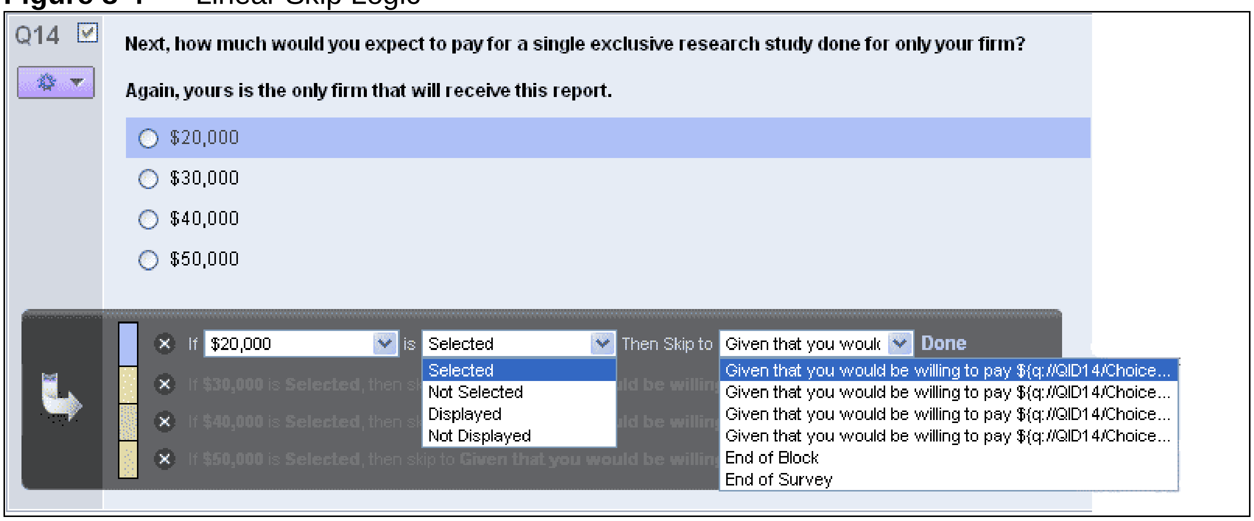
Logic to control question order based on respondent answers can be easily added to online surveys. Survey flow logic exists in several levels of complexity, each of which will be discussed in this section. We will begin with an introduction to simple skip logic and progress to more advanced logic used in experimental designs.

Linear Skip Logic

Linear skip patterns move the respondent from one question to another, based on the respondent's answer to a single question. For example, if a respondent indicates that she is a non-consumer of energy drinks, she is skipped past energy drink product and brand specific questions, again entering the linear flow of survey questions that ask about general consumption of casual beverages. In this example the survey flow moves linearly down through the questionnaire based on answers to an individual question. Multiple instances of skip logic could be entered where appropriate.

Skip logic is added in Qualtrics after the question has been entered. Select "Add Skip Logic" from the purple dropdown button to the left of the question. The logic bar will appear wherein you can specify the answer option to skip on, the logical condition (selected, not selected, displayed, not displayed), and the skip to destination (a specific question, end of block or end of survey). Multiple skip logic statements are supported for each question (Figure 8-4).

Figure 8-4 Linear Skip Logic



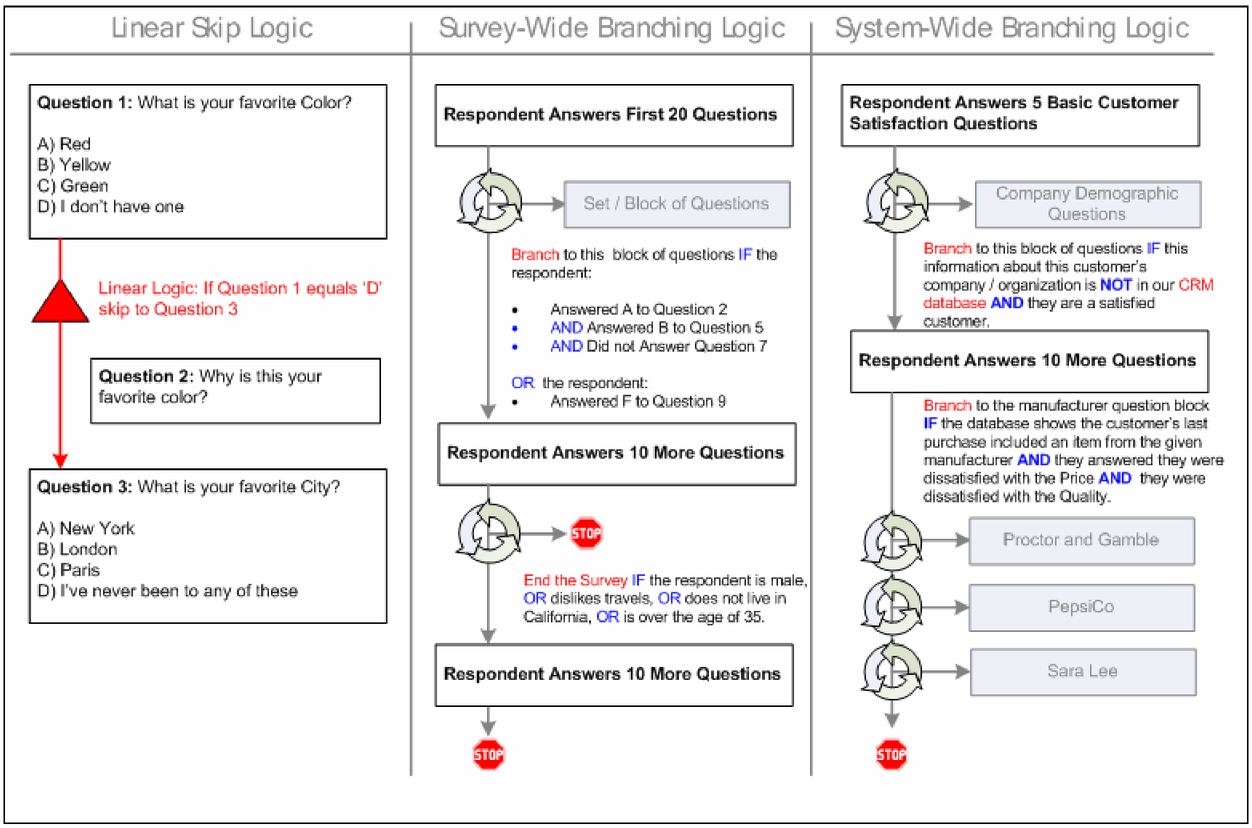
In the above example, when the answer "$20,000" is selected, the branch is taken and the text "$20,000" is piped from Q14 to Q33. To Pipe Smart Text from Q14 to Q33, just go to Q33 and edit the question. Select {a} in the Rich Text Editor at the point in the question or answer where you would like the piped text to appear. The specific piped text is specified using the menu options. Remember that linear skip logic uses positive logic conditions setting the criteria for the people who enter the branch, not for those who don't—as a default people will NOT enter the branch.

Survey- and System-Wide Boolean Logic for Compound Branching

Branching is more sophisticated than linear skip logic and is used to skip questions based on a complex sequence of conditions. Suppose that a respondent meets the multiple conditions required to be defined as a member of our target demographic segment. She is “female,” “between 18 and 34,” “married,” and has “one or more children.” Branching then is based on multiple questions or points of information provided by the respondent, or found in the panel data. With branching, the intersection of these identifying characteristics is the point where we change survey flow and question logic.

The concept of system-wide Boolean logic expands the locus of control for branching. This approach enables a survey to branch based on (a) answers from multiple questions within the survey; (b) outside variables nonspecific to the respondent, such as a randomly assigned experimental group defined in either the panel information or by Web Services content; (c) panel and database information about the respondent (such as CRM systems or past surveys the respondent has completed). Figure 8-5 contrasts these approaches.

Figure 8-5 Linear Skip Logic versus Boolean Logic and Compound



Looping with Piping

Looping with piping is a combination of features designed to reduce the length and complexity of a survey building process when a standard set of evaluation questions are repeated over and over. The following example shows how the use of looping and piping shortens the questionnaire and logic building process.

> *As an example, suppose that a respondent is asked to select from a list of 150 educational programs, those that were used by his or her organization. Next, the person responds to three evaluation questions about each program selected. Rather than building a linear survey of 600*

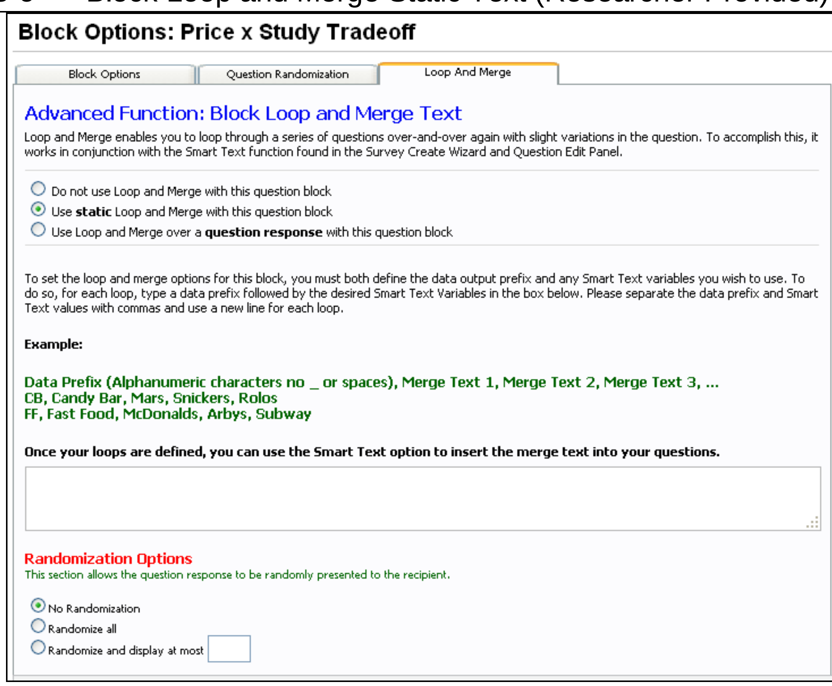
questions (do you use program x? – yes, no; if yes, then evaluate, if no, ask about next program), the researcher would build only 4 questions (check all that apply in the list of 150 programs and the three evaluation questions). The question "Select from the 150 programs those you have used" would appear in the main program block and then a second block would contain the 3 evaluation questions. The pipe and loop feature would loop through only those programs selected, identifying them by name as the three evaluation questions were repeated.

.

Note that looping based on a "Question Response" can be based on all choices presented to the respondent in a previous question, all choices selected by the respondent, or those choices not selected (for asking questions of the type: "why didn't you consider this option in your decision"), or those choices that appear or do not appear, where appearance occurs as a function of some previously defined condition. In other words, the choices of the respondent can dictate the flow of this type of question.

In contrast to control by "Question Response", the experimenter can use "Static" control with a predefined set of loops and piping. For example, a researcher has predefined a set of product brands for evaluation. In this case, the predefined product category and brand names "fast food, McDonalds, Arbys, Subway" could be piped and merged into the sequence of three questions that is automatically repeated for each of the three fast food brands evaluated. In experimental settings, this approach could be used to present treatments, or evaluate a defined set of brands, usage occasions, employee, or similar list of interest. This predefined list is defined in the Block Options / Loop and Merge dialogue box as shown in figure 8-6. Once the text is defined, the Smart Text Piping option is used to specify placement of the merge text within a question.

Figure 8-6 Block Loop and Merge Static Text (Researcher Provided)



With increasing time demands on potential respondents, they often review questionnaires to estimate the time and effort required. Streamlining questionnaires using looping and piping has the potential of reducing the physical length of the survey and associated respondent fatigue, thereby increasing data quality and completion rates.

Carry Answers Forward

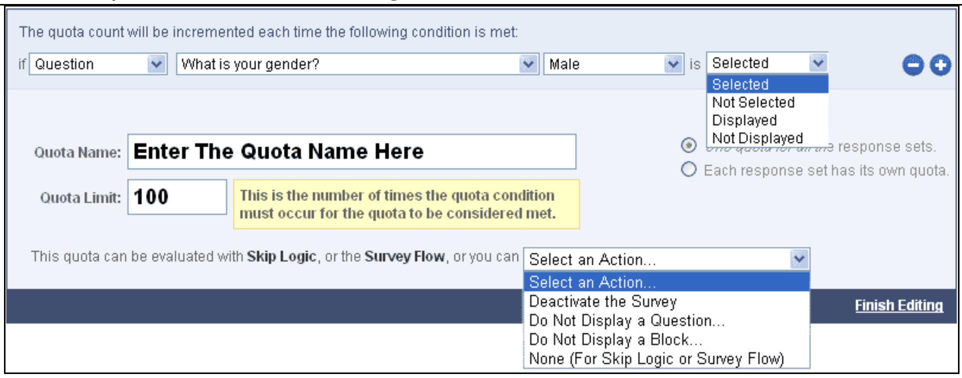
Closely related to the streamlining effects of loop and pipe feature is the ability to carry forward any or all respondent answer choices from one question to a later question. This is particularly useful for experimental design where additional follow-on information is desired.

All possible answers or only the *k of n* items that were selected can be moved to a later question to ask further questions about those items. For example, sports actively participated in or the names of brands previously purchased could be moved to a next question or block of questions for further drill down questioning. The order of item presentation can be maintained or randomized. The ability to carry answers forward is also useful where the respondent selects an "other (please explain)" category and provides the associated text response. This text response can then be carried forward to gain additional information.

Quota Fulfillment

Quota fulfillment is an integral part of experimental designs. Quotas work by testing the size of the sample groups (defined by answer categories within a given question) against target numbers (Figure 8-7). Where experimental designs are used, quotas may be created to avoid over sampling of some groups while under sampling other groups. Given that many studies pay respondents for completed surveys, quotas, if used early in the questionnaire can screen and eliminate unneeded respondents, thereby reducing total research costs. The quota specification menu is found in the Advanced Options dropdown link in the main Edit Survey page.

Figure 8-7 Sample Size Quota Setting



This section has provided a brief overview of the embedded logic and complex information flows that can be adapted to meet the rigors of an experimental design. Unlike paper- or interview-based questionnaires, these "smart surveys" perform on-the-fly analysis and incorporate recorded answers, embedded or panel data into questions. In addition, experiments and statistically sound quota sampling are facilitated in dynamic sequences of questions that are displayed based on a respondent's answers and/or match to a group or quota. The next section expands on this logical base to discuss the use of question blocks, flow, and randomization in experimental designs.

Using Survey Flow to Create Experiments

Qualtrics contains powerful, but easy to use tools for customizing question and block presentation order in experimental designs. The complex branching previously discussed can be used to direct the respondent through blocks of questions, advanced question types (conjoint analysis), and even complex randomization in the survey flow of an experiment.

Question Blocks

A question block is a cohesive group of questions that receive similar treatment. Simple questionnaires using only linear skip patterns or no skips at all will often use only a single question block (a single default question block is created automatically in Qualtrics). More advanced questionnaires may split questions into multiple question blocks based on topic or logical flow.

A question block may be controlled independently of the rest of the survey and for example, be given special instructions that apply only to questions within the block. Questions within blocks may be arranged in a fixed order, random order, or partial random order that holds some questions in a fixed position, while randomizing the order of other questions. The presentation order of multiple blocks, should they be present in the survey flow, may also be manipulated:

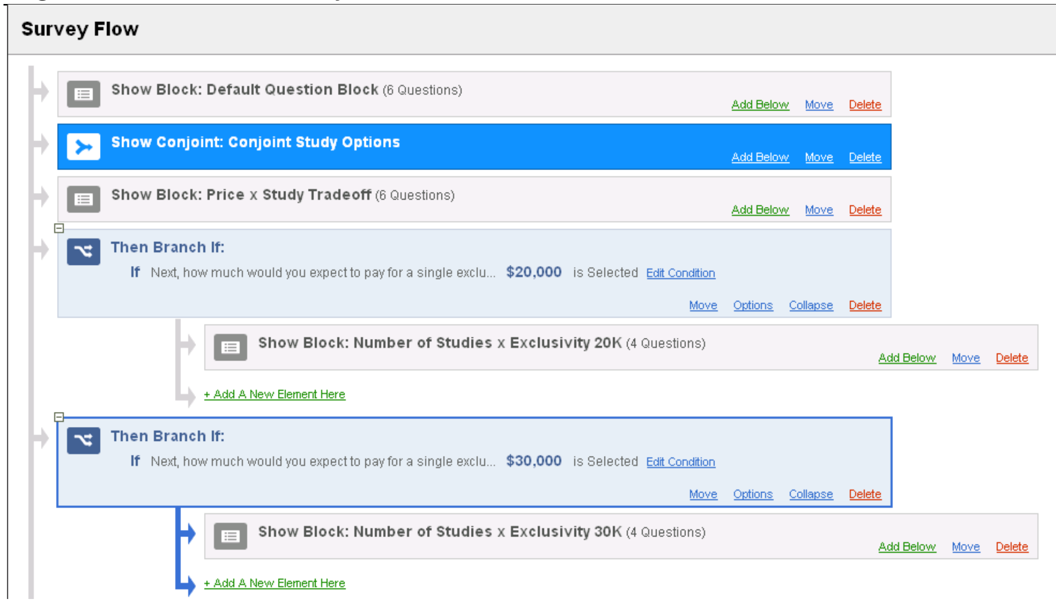
a) A block of questions can be displayed based on choice previously made by a respondent (evaluate a block of brand-specific questions if the respondents indicated they were familiar with the brand),
b) The blocks can be randomly presented or held in a fixed position while randomizing the order of other blocks (each block is a treatment or control in an experiment)
c) Blocks can be repeated in a looping pattern with piped text that describes the purchase situation or usage occasion under which the block of questions is to be evaluated (as in the loop and pipe example presented above).

Any element in the survey flow can be dragged and dropped into any location in the flow. In this way, nesting of treatments within the logical flow of an experiment is easily accomplished.

Randomizing Questions and Blocks of Questions

Randomization is used when it is not important to maintain the question order or context of the questions. Blocks may be used to assist in maintaining context and at the same time achieving a level of randomization. For example, context often requires that questions be presented in a given sequential order, but blocks of questions (each block, for example, containing questions about different brands) can be randomized. In many concept testing and other experimental situations, randomization of questions within a block or randomization of blocks of questions is a valuable option.

Figure 8-8 The Survey Flow Chart



Once the survey building has been completed and question blocks are created, the initial survey flow is created. In Qualtrics, question block and element presentation can be visualized as a color coded sequence organized the way the survey "flows" (Figure 8-8). The survey follows the main flow line on the left, and presents questions in the order shown from top to bottom. The secondary flows branch to the right, where each color coded "element" shows the branching logic statement (if used) and the questions shown to the respondent. Survey flow Elements may be added at any point in the flow.

The researcher adds survey flow “Elements” into the flow control sequence. The two survey flow elements most often used are “Block" (question blocks), and "Branch." Other elements include “Embedded Data,” “Randomizer,” “Web Service,” “End of Survey,” and the “Conjoint” analysis module (Figure 8-9).

Figure 8-9 Survey Flow Elements

Adding Block and Branch Elements to the Survey Flow

Block

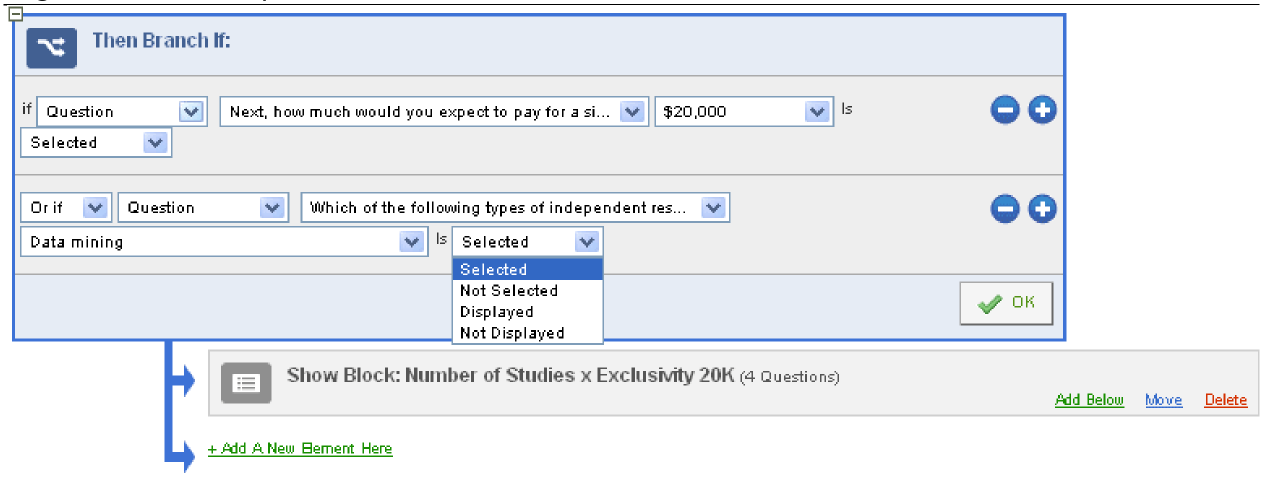
Blocks of questions, once created, are automatically included in the questionnaire flow. Blocks can be added at any location in a Branch or in the Main Flow. Once a block is inserted, it will appear as a survey flow element on its own secondary line. As shown in Figure 8-7, respondents will first see the Default Question Block, and then conditions branching to other blocks of questions (perhaps identifying treatments) that were developed by the researcher.

Branch Logic

Branches function differently than Blocks. A Branch acts as a filter which, based on answers to previous questions or other information, directs certain people to an area of a survey if the specified condition is met. Those not meeting the condition proceed to the next condition in the survey flow.

All branch logic clauses operate on responses received to a previously answered Question, Embedded Data (generally panel data), or Quotas. The logic section can also be triggered by choices that were Selected, Not Selected, Displayed or Not Displayed. Selected and Non-Selected are self explanatory and can be used to identify respondents who indicate these are, for example, purchasers or non-purchasers of a brand. Displayed and Not Displayed options are useful when text is piped or answers are carried forward from one question to another. Displayed is synonymous with the respondent receiving the treatment or not… and is useful in construction of if so/not, then show/don't show a treatment of further questions. Options are specific to the question type. For example, a text input question includes the options: Empty, Not Empty, Displayed, Not Displayed, etc.

Figure 8-10 Multiple Then Branch If Statements



Note: Because experimental designs often require extensive logic sequences, it is worth noting that within the chain of logic statements, quotas have first priority, followed by the "And If" clause, which has a higher priority than the "OR IF" clause (i.e. with a chain of clauses such as 1 AND 2 AND 3 OR 4, the system will read it as (1 AND 2 AND 3) (OR 4), meaning IF 1 AND 2 AND 3 are all fulfilled, OR if 4 is fulfilled, then the respondent will be branched.

Enhancing Surveys by Adding Embedded Data into the Survey Flow

Embedded Data is "behind the scenes" respondent data. Generally, embedded data will be uploaded as part of panel data. Once this data is uploaded, operations such as skipping and branching can be performed using the embedded data as a key selection variable.

For example, if we were to access the person's home state as an embedded data field in a panel mailing list, we would specify the location, variable, logic and condition as:
"If: Embedded Data : Home State Equals Colorado."

Embedded data can also be static codes marking events or conditions that are referenced by logic conditions at other points in the survey.

Randomization of Answer Choices, Branches, Question Blocks and Surveys

Randomization techniques are used to control answer presentation order bias, as well as the effects of extraneous variables not controlled by an experimental design. In new product concept tests and other advanced measurement situations, the researcher may randomize and control the presentation order of the following in online research and experimental designs:

- Complete randomization of all choice options
- Complete randomization of choice options and non-randomization of "other" or text input options
- Partial randomization of choice options, where one or more specific answers (such as the item of interest) are held in a fixed position, and all other answers are randomly ordered
- Randomization of blocks containing treatments
- Partial randomization of blocks, where one or more of the blocks are held in fixed presentation order, or are presented to all respondents, while other blocks are randomly or intermittently presented (randomly present 2 of the 4 other treatment blocks)
- Separate questionnaires containing separate treatments are presented randomly.

The randomization of choice options is the most basic type of randomization, and deserves little special attention other than to suggest that answer choice sets without a fixed order and matrix type questions (displaying lists of questions, each evaluated on the same scale) are best presented in random order. Randomization at the question level is found the individual question options.

Figure 8-9 Block Randomizer Screen

Randomization of treatment blocks works by inserting the element "Block Randomizer" into the survey flow. Do this by clicking on the "Add New Element" link, choosing "Block Randomizer," and then choosing the Blocks you wish to place it in. The randomized blocks will appear in the survey in the place where you insert the block randomizer.

For experimental research treatments can be randomized with all, or random subsets of the treatments presented. To select a subset of blocks, make sure all blocks to be randomized have been added into the Block Randomizer. The number in the dropdown reflects the number of blocks you want the respondents to see. You may set it to show all of the blocks in the randomizer, or a subset of those blocks (i.e., If 4 blocks are added to the randomizer and each respondent is to randomly respond to two of those blocks, put a "2" in the box.

Randomization using Multiple Questionnaires

Occasionally, experimental treatments are so extensive that a completely different questionnaire is required for each treatment. In such cases, the questionnaires are prepared and randomly assigned to the respondents. Such a design may be achieved by either using a “portal survey” where the links to the other surveys are randomly presented via questions or block randomization, or by inserting a separate java script into a question. The java script would be similar to the one that follows to randomly assign respondents to a survey.

Sample Experimental Designs and Associated Survey Flows

Script to Open Survey in the Same Window, or in a Separate Window:

```
<SCRIPT LANGUAGE="JAVASCRIPT" TYPE="TEXT/JAVASCRIPT">
Event.observe(window, 'load', function() {
        <!-- Javascript to redirect user to a randomly selected survey //-->
        // Array structure - as many links as you need
        var seedcount = 7
        var index = 3;
        var linkArray = new Array(index);
        // Load array with available survey links here:
        linkArray[0] = "http:// new.qualtrics.com/SE?SID=SV_6LR6QjjYiHkSe21&SVID=Prod";
        linkArray[1] = "http:// new.qualtrics.com/SE?SID=SV_6LR6QjjYiHkSe22&SVID=Prod";
        linkArray[2] = "http:// new.qualtrics.com/SE?SID=SV_6LR6QjjYiHkSe23&SVID=Prod";
        // Generate random number that will correspond to an index in the above array.
        // The integer in the equation is the total number of links you have minus 1
        now = new Date();
        var seed = now.getSeconds();
        randNum = Math.round(1000000*Math.random(seed))
        for (loopcount = 0; loopcount < seedcount; loopcount++) {
          randNum += Math.round(1000000*Math.random(randNum * seed));
        }
        randNum = Math.round(1000000*Math.random(randNum)) % index;
        // Use one of the two statements below by changing the comment
        // Redirect user to appropriate random survey page in the current window
         window.location.replace(linkArray[randNum]);
        // Or open the appropriate random survey page in a new window
        // window.open(linkArray[randNum]);
});
</script>
```

Field Experimentation in Marketing

Unfortunately, experimental methods are not always the methods of choice in the study of marketing phenomena because depending on the mode of field experimentation, it may be quite expensive, subject to large amounts of uncontrolled variation, or may produce results that are difficult to generalize to other products, market areas, or time periods

There is little question that field experimentation in marketing can be costly undertaking. Consider the case of a sales manager who wishes to determine the effects of varying amounts of sales effort on product sales of a nationally distributed brand. Sales in a given time period could be affected by point-of-purchase advertising, personal sales effort, broadcast promotion, competitors' selling efforts, relative prices, seasonal effects, past promotional expenditures, and so on. However the capabilities of online surveys software like Qualtrics.com has provided one possible approach to reduce costs while maintaining the precision associated with experimental research.

Experimentation represents another way of information gathering and should be approached in terms of its cost versus value (compared with other techniques) for decision-making purposes, with regard to both present and anticipated problems. Unfortunately, it is much easier to state this objective than to design suitable techniques for implementing it.

SUMMARY

In this chapter our primary objectives were (1) to introduce the conceptual bases underlying marketing experimentation, and (2) to discuss basic models of experimental design. The statistical techniques for analyzing experimental data will be discussed in later chapters.

The first section of the chapter covered the nature of experimentation. The critical aspects of this discussion were the various ingredients of an experiment. We next turned to the potential sources of invalidity, internal and external, associated with an experiment. These are the experimental errors that relate to extraneous factors that can affect an experimental outcome.

We next described some of the major classical and statistical experimental designs. Some of these designs—single-factor, factorial, randomized blocks, Latin squares—were illustrated.

The Qualtrics research suite was then presented as a powerful and integrative tool for researchers who need to conduct advanced field studies, including experimentation. We have introduced an array of advanced online survey functionalities and tools that allow the researcher to create more visually interesting, involving, and realistic approaches to conducting research. We have also discussed advanced logic and flow capabilities that support the creation of experiments.

We concluded the chapter with a discussion of some of the problems involved in marketing experimentation and the relationship of this means of data collection to the cost and value of information.

REFERENCES

Banks, S. (1965). *Experimentation in Marketing*. New York: McGraw-Hill

Bay, W. (1998). NewsStand: CNN & Entertainment Weekly Report, September 28.

Brown, S.R. and Melamed, L.E. (1990). *Experimental Design and Analysis*. Thousand Oaks, CA: Sage Publications,Inc.

Campbell, D.T. and Stanley, J.C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally

Christensen, L.B. (2004). *Experimental Methodology*. Boston: Allyn & Bacon, Ninth Edition

Cook, T.D. and Campbell, D.T. (1990). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin Co.

Edwards, A.E. (1968). *Experimental Design in Psychological Research*. New York: Holt, Rinehart & Winston, Third Edition

Field, A. and Hole, G. (2003). *How to Design and Report Experiments*. Thousand Oaks, CA: Sage Publications

Perdue, B.C. and Summers, J.O. (1986). "Checking the Success of Manipulations in Marketing Experiments," *Journal of Marketing Research*, 23 (November), 317-26

Rosenthal, R. and Rosnow, R.L. (1975). *The Volunteer Subject*. New York: John Wiley & Sons

Rosenthal, R. and Rosnow, R.L. (1991). *Essentials of Behavioral Research: Methods and Data Analysis*. New York; McGraw-Hill, Second Edition

Sawyer, A. G. (1975). "Demand Artifacts in Laboratory Experiments in Consumer Research," *Journal of Consumer Research*, 1 (March)

Shadish, W., Cook, T.D., and Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Causal Inference*. Boston: Houghton Mifflin Co.

Appendix 8.1
Sample Experimental Designs

Case 1: Trademark Confusion Testing [1]

"Schema theory suggests that when consumers are offered a junior mark in a product category that is different from the senior mark's category (e.g., a CADILLAC brand notebook computer), the likelihood that consumers will categorize the product as a subtype of the senior mark (i.e., the likelihood that consumers will suffer confusion) is related to the product's incongruity with the senior mark's schema.

With the goal of creating stimuli to test the hypotheses, a well known automotive brand (CADILLAC) was chosen as the senior mark. Notebook computers were chosen as a product category for the junior mark because notebook computers were anticipated to be relevant to online participants. A fictitious junior mark (CADILLAC brand notebook computers) as then created, along with an advertisement that described an upscale, luxurious notebook computer that could be offered by a firm attempting to infringe on the CADILLAC trademark. The following advertising text was introduced with the claim that it would "appear in the next edition of a popular computer-sales catalog."

> Cadillac brand notebook computer! This luxury notebook has an upgraded keyboard for better feel and absolutely silent typing. A dark mahogany inlay surrounds the screen. Matching dark mahogany accents are found on the wireless mouse and the leather carrying case. Intel Core 2 Quad Q6700, 4GB RAM, 200GB Hard Drive. *Price available upon request.*

H1 suggests that viewers of this advertisement will be more likely to suffer confusion by believing CADILLAC MOTORS to be the source of the notebook computer if another maker of luxury automobiles has already extended its brand into the luxury notebook computer product category.

To test this hypothesis, three stimuli were created. The first stimulus (see Figure 1) was an experimental control designed to represent a typical advertisement for a notebook computer at the time of the study. The advertisement emphasized computing power and did not attempt to create an upscale positioning with luxurious product attributes.
The second stimulus informed participants of the launch of an upscale notebook computer (see Figure 2). Luxurious product attributes such as a titanium case were described in support of an upscale positioning. This second stimulus was another experimental control because the upscale notebook computer was described as being offered by a well-known computer company (Dell) rather than by an automotive company.

The third stimulus (see Figure 3) informed participants of the launch of an upscale notebook computer as a brand extension of a well-known automotive brand (MERCEDES-BENZ). To encourage participants to subtype the notebook computer in the MERCEDES-BENZ schema, the ad included a photo of a MERCEDES-BENZ car and a close-up of a MERCEDES-BENZ hood ornament. By comparing viewers of this stimulus with viewers of the two control stimuli, H1 can be tested.

Participants were randomly assigned to view either the typical DELL computer advertisement (condition 1, n = 167), the luxury DELL computer advertisement (condition 2, n = 165), or the luxury MERCEDES-BENZ computer advertisement (condition 3, n = 163). The experimental method was used to provide a rigorous test of the causality claimed in H1. Depending on the assigned experimental condition, one of the three stimuli was shown as part of an online questionnaire instrument. The stimulus was shown in full-color, and the time of exposure was self-paced. Participants then completed a filler task to mask the purposes of the study. They were then shown the advertisement for the CADILLAC notebook computer and asked to complete the source-confusion battery. The study's independent variables were then measured."

[1] Source: Lee, T. R., Derosia, E. D. and Christensen, G. L., Sophistication, Bridging The Gap, And The Likelihood Of Confusion: An Empirical And Theoretical Analysis, The Trademark Reporter, Vol. 98, No. 4, July-August 2008. Used by permission of Professor Glen Christensen, Brigham Young University.

Figure 1 (Condition 1: Control)

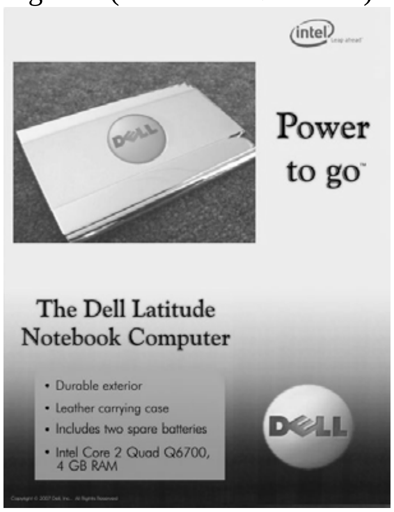


Figure 2 (Condition 2 (Control)

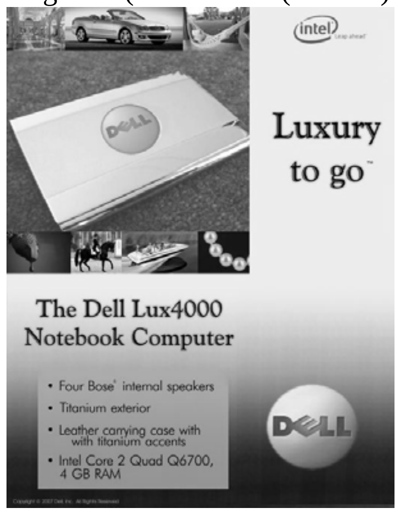
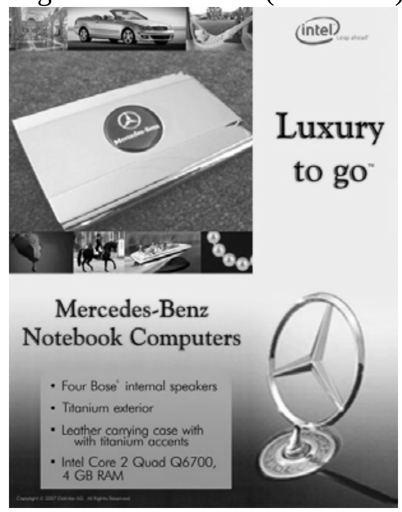


Figure 3 Condition 3 (Treatment)



The following is the flow chart shows the above three group experimental design with control

Flow runs vertically from left to right

Default question block of Introductory questions, followed by screening questions

- Block: Default Question Block(FL_1)
- Block: Screener Questions(FL_4)

Branch Screener using IF and OR IF questions to qualify the respondent and send to treatments or end survey.

- Branch: Screener End(FL_21)
  - End Survey: Screener End(FL_22)
- Block: Control (FL_23)

Stimulus Condition One (Control)

- Block: (Treatment Construct One) Introduction Questions(FL_6)
- Branch: (Treatment Construct One) Product Recognition (FL_11)
  - Block: (Treatment Construct One) Product Knowledge Questions (FL_12)

Stimulus Condition Two (Control)

- Block: (Treatment Construct Two) Introduction Questions (FL_18)
- Branch: (Treatment Construct Two) Product Recognition (FL_19)
  - Block: (Treatment Construct Two) Product Knowledge Questions (FL_20)

Stimulus Condition Three (Treatment)

- Block: (Treatment Construct Three) Introduction Questions (FL_8)
- Branch: (Treatment Construct Two) Product Recognition (FL_15)
  - Block: (Treatment Construct Two) Product Knowledge Questions (FL_16)

End of Treatments

- Block: Demographics(FL_17)
- Block: Debrief, Final Questions(FL_9)

End Survey

Case 2: Determining Price Sensitivity

A major investment firm is interested in determining the price sensitivity of fund managers to specially prepared research studies about selected new equity offerings. The research design is to include a conjoint analysis. This analysis is followed by an experimental design (unrelated to the conjoint analysis) that measures the fund manager's sensitivity to the report price given their current payment expectations for a single research study that is to be prepared exclusively for their firm. In the design, report price must vary at fixed levels to reflect volume discounts (more reports per year) and report exclusivity (proprietary verses syndicated).

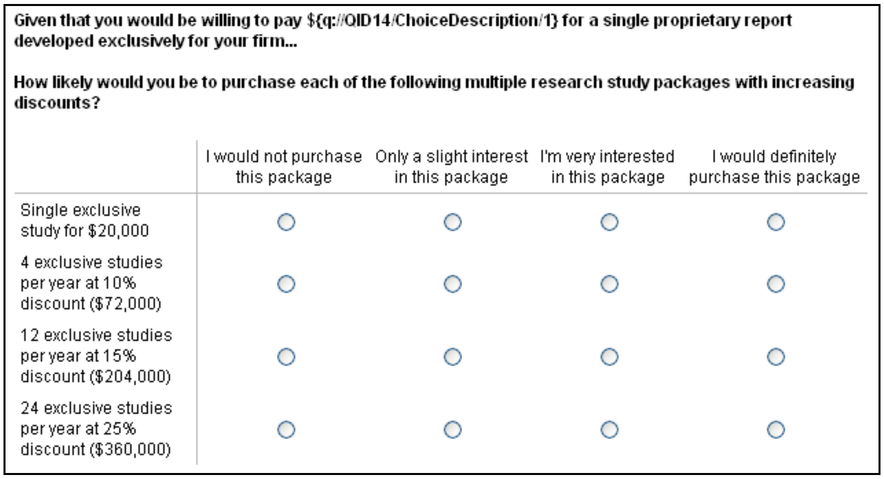
The Experimental design in this example would split respondents based on the price they currently expect to pay for a single exclusive research report ($20k, $30k, $40k, $50k). From this base point, price acceptability is measured using a based on two other factors: (1) The number of studies purchased per year (4, 12, 24), and (2) The exclusivity of the research reports (The report will be distributed to 1, 4, 10, 25 or 100 firms). The prices tested will vary as shown in Table 1.

Table 1: Design and Price Matrix

Expected Price	20K	30K	40K	50K
Studies to Purchase	4 Studies (1 per quarter)	4 Studies (1 per quarter)	4 Studies (1 per quarter)	4 Studies (1 per quarter)
Exclusivity	1 @$204k 4@$163.2K 10@$81.6K 25@$40.8K 100@$20.4K	1 @$108k 4@$86.4K 10@$43.2K 25@$21.6K 100@$10.8K	1 @$144k 4@$115.2K 10@$57.6K 25@$28.8K 100@$14.4K	1 @$180k 4@$144K 10@$72K 25@$36K 100@$18K
Studies to Purchase	12 Studies (about 1/month)	12 Studies (about 1/month)	12 Studies (about 1/month)	12 Studies (about 1/month)
Exclusivity	1 @$360k 4@$288K 10@$144K 25@$72K 100@$36K	1 @$306k 4@$244.8K 10@$122.4K 25@$61.2K 100@$30.6K	1 @$408k 4@$326.4K 10@$163.2K 25@$81.6K 100@$40.8K	1 @$510k 4@$408K 10@$204K 25@$102K 100@$51K
Studies to Purchase	24 Studies (about 2/month)	24 Studies (about 2/month)	24 Studies (about 2/month)	24 Studies (about 2/month)
Exclusivity	1 @$72k 4@$57.6K 10@$28.8K 25@$14.4K 100@$7.2K	1 @$540k 4@$432K 10@$216K 25@$108K 100@$54	1 @$720k 4@$576K 10@$288K 25@$144K 100@$72K	1 @$900k 4@$720K 10@$360K 25@$180K 100@$90K

The survey flow next presents a default question block of classification and qualification questions, and is followed by the conjoint analysis module. The flow next branches based on the respondent's stated acceptable price, to a matrix evaluating exclusivity x volume discount. This matrix uses piped smart text to reference the appropriate acceptable price level specified by the respondent (Figure 1). Finally, Branch elements direct the respondent to the appropriate pricing sensitivity section.

Figure 1 Piped Text to Branched Matrix Question


Given that you would be willing to pay ${q://QID14/ChoiceDescription/1} for a single proprietary report developed exclusively for your firm...

How likely would you be to purchase each of the following multiple research study packages with increasing discounts?

	I would not purchase this package	Only a slight interest in this package	I'm very interested in this package	I would definitely purchase this package
Single exclusive study for $20,000	○	○	○	○
4 exclusive studies per year at 10% discount ($72,000)	○	○	○	○
12 exclusive studies per year at 15% discount ($204,000)	○	○	○	○
24 exclusive studies per year at 25% discount ($360,000)	○	○	○	○

Figure 2 The Survey Flow

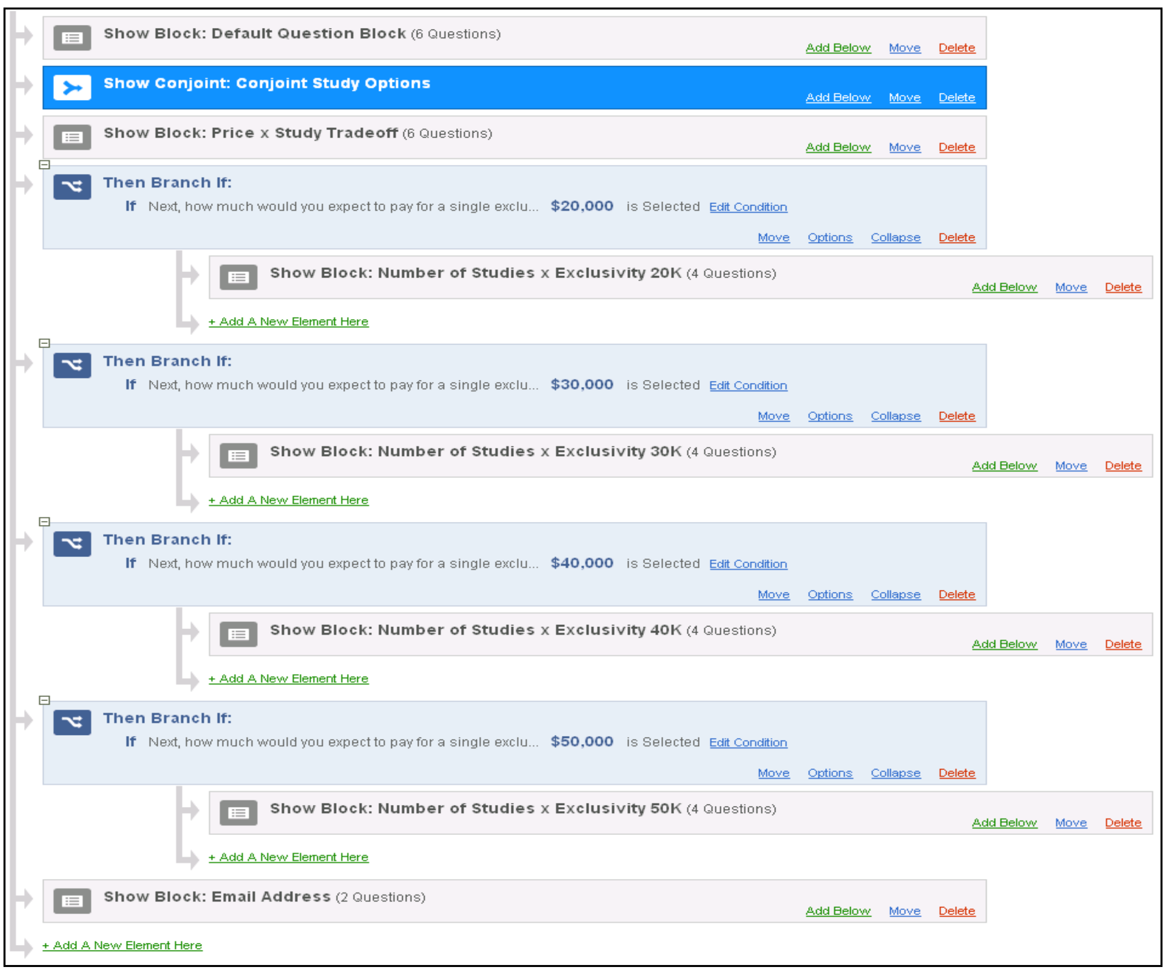
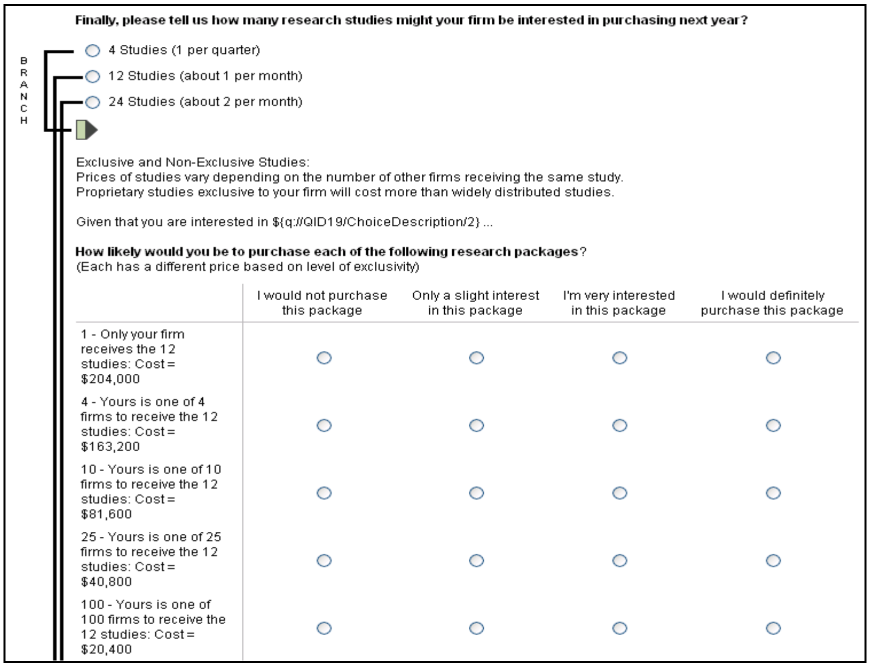


Figure 3 Number of Studies x Exclusivity for Branch (Cell 1-1 of the Design – Price Matrix)



Finally, please tell us how many research studies might your firm be interested in purchasing next year?

B R A N C H

- 4 Studies (1 per quarter)
- 12 Studies (about 1 per month)
- 24 Studies (about 2 per month)

Exclusive and Non-Exclusive Studies:
Prices of studies vary depending on the number of other firms receiving the same study.
Proprietary studies exclusive to your firm will cost more than widely distributed studies.

Given that you are interested in ${q://QID19/ChoiceDescription/2} ...

How likely would you be to purchase each of the following research packages?
(Each has a different price based on level of exclusivity)

	I would not purchase this package	Only a slight interest in this package	I'm very interested in this package	I would definitely purchase this package
1 - Only your firm receives the 12 studies: Cost = $204,000	○	○	○	○
4 - Yours is one of 4 firms to receive the 12 studies: Cost = $163,200	○	○	○	○
10 - Yours is one of 10 firms to receive the 12 studies: Cost = $81,600	○	○	○	○
25 - Yours is one of 25 firms to receive the 12 studies: Cost = $40,800	○	○	○	○
100 - Yours is one of 100 firms to receive the 12 studies: Cost = $20,400	○	○	○	○

Chapter 9
MEASURING RESPONDENT INFORMATION:
ATTITUDES, SATISFACTION, LOYALTY AND BEHAVIOR

INFORMATION FROM RESPONDENTS

All marketing decisions involve recognizing alternatives to, and making predictions of, the behavior of market participants. Marketing decisions ultimately hinge, in whole or in part, on a prediction of the behavior of consumers, industrial users, marketing intermediaries, competitors, and, at times, the government. Whether the decision is to introduce a particular new product, raise the price of an existing product, change distribution channels, or determine an advertising budget, the solution involves forecasting the behavior of one or more groups of market participants.

We now consider the most frequently studied variables used to predict behavior: demographics, attitudes, intentions, satisfaction and loyalty. The information used to predict behavior is grouped into behavioral correlates (past or intended behavior) and nonbehavioral correlates (demographics, attitudes, lifestyles, values).

Behavioral Correlates

Behavioral correlates address the question of what past behaviors or intended behaviors predict future behavior.

Past Behavior

Past behavior is widely used as a predictor of future behavior. Each of us relies heavily upon our knowledge of others' past behavior in our everyday relationships with our family, friends, and associates. When we state that we "know" someone well, we are implicitly saying that we believe we are able to predict that person's behavior in a wide range of social and interpersonal situations. In economics applications, we examine past trends, seasonal averages, and cyclical patterns for forecasting.

Regardless of the nature of the variable or variables to be forecasted, a basic premise of using past behavior to predict future behavior is that there is a relationship between the two that, to some extent, is stable. Recognizing that the degree of stability is sometimes difficult to determine, and that we do not always understand the underlying causal relationships, we nonetheless must believe that there is some continuity and stability in the behavior patterns of people.

A typical consumer brand purchase study would concern itself in part with determining such facts as what brands have been used, the last brand bought, where and with what frequency purchases are made, what the exposure to company advertising has been, and similar aspects of past behavior.

Information about past behavior toward products is often classified into three categories: acquisition, use, and possession. Within each of these behavioral areas, we collect detailed information about who, what, when, where, how much, and in what situation the behavior occurs. This data produces detail useful in understanding product consumption patterns. The particular study's requirements will dictate which of these types of information will be most useful. Table 9.1 shows the requirements for a study on tomato juice to determine, among other things, whether a new type of container should be developed. Often, such information comes from secondary sources, as previously discussed.

Intended Behavior

Intentions may be defined as presently planned actions to be taken in a specified future period of time. What more logical method of predicting the future behavior of respondents could be used than determining their intentions? After all, intentions are self-predictions of behavior, and thus, if obtained from people whose behavior we want to predict, would seemingly be the most direct and reliable method of prediction.

Intentions are relevant and commonly sought as predictors of behavior. However, consideration of our own experiences in terms of what we have planned to do vis-à-vis what we have actually done later should raise some questions concerning the reliability of intentions as a predictive tool. The question “What will you do?” must always be answered conditionally.

TABLE 9.1 Past Behaviors: A Study of Tomato Juice Usage Patterns

	Acquisition	*Use*	*Possession*
Who	Who in your family usually does the shopping	Who in your family drinks tomato juice	
What	What brand of tomato juice did you buy last time?	What dishes do you cook or prepare with tomato juice?	What brands of tomato juice do you now have on hand?
When	How long has it been since you last bought tomato juice?		
Where	Do you usually do your shopping at a particular store or supermarket? (if Yes, where?)		
sssHow Much	What size can of tomato juice do you usually buy? About how often do you buy tomato juice? About how many cans do you buy at a time?	About how much tomato juice does your family drink in a week? For which purpose does your family use more juice?	Do you now have any unopened cans of tomato juice on hand? (If Yes, about how many cans?)
Usage Situation		How does your family use tomato juice? Beverage □ Cooking □ Both □ Beverage with friends □	How do you store tomato juice after it is opened? Can □ Bottle □ Plastic □ Other □

The degree of assurance that planned actions will be translated into actual actions varies widely depending on circumstances and future happenings, many of which are outside the respondent’s control.

The results of a study of expected and purchase rates of a few products and services are shown in Table 9.2. Researchers collected intentions data from a consumer panel sample using a 0 to 10 scale to measure purchase probabilities. Verbal definitions were assigned to each point on the scale. A 10 was defined as "absolutely certain of buying" and a 0 as "absolutely no chance of buying." The definition of a 5 was given as "five chances out of ten of buying," and the other points between 1 and 9 inclusively were similarly defined.

TABLE 9.2 Intentions Fulfilled and Not Fulfilled During a 60-Day Period

Product/Service	Intentions-Based Expected Purchase Rate %	Purchase %	Difference
Ride local public transportation	22.5	21.7	-0.8
Purchase tax-sheltered investment	11.4	7.2	-4.2
Purchase stereo system	17.6	15.6	-2.0
Take a trip on cruise ship	4.2	3.7	-0.5
Purchase new automobile	14.3	14.1	-0.2

Expected purchase rates were calculated as the average purchase probability for each item. The actual rate was determined by reinterviewing the panel members 60 days later to find out what purchases they had actually made.

Intentions to buy are often conditioned by judgments and expectations of future events or future situations, as well as by our past experiences. Such variables as expected change in financial status, price expectations, general business forecasts, and predictions of need all contribute to the final intention decision. Since each of these is subject to variation, it seems plausible to suppose that the intender views them as such and that his or her stated intention is based on a subjective probability of purchase. This supposition is supported by the fact that intentions data with assigned probabilities have generally proven to be more accurate than those expressed in "either/or" form.

Past experiences can often be inferred from a customer's level of satisfaction. Increasingly, customer satisfaction surveys fill a critical role in many firms' customer relationship management (CRM) systems (James, 2002). Customer satisfaction relates to intentions as it can significantly affect the repurchase decision of consumers.

Nonbehavioral Correlates

So far we have discussed how people's past behaviors and their intentions are correlates of what they will do. We now need to examine the nonbehavioral correlates that are useful for predicting their future behavior.

Socioeconomic Characteristics

How is information on the social and economic characteristics of respondents useful for forecasting what people will do? The answer can be readily suggested by an illustration. The Radio Corporation of America (RCA), when introducing color television in the 1950s, was very much interested in the age, income, educational, and occupational composition of the market. They judged that the initial market willing to pay for a premium priced color television set would be families proportionally higher in income and educational levels and older, than either the black-and-white set owners or the population as a whole. These judgments were subsequently confirmed by a study of early purchasers of color sets. This information was useful for both pricing and promotional

EXHIBIT 9.1 Measuring Intentions

1. Measuring intentions takes on various forms. Hotel chains typically place self-report satisfaction questionnaires in each room. A typical intentions question would be:

Do you plan to return to this hotel on your next visit?

- ◯ Yes
- ◯ No (If not, Why?) [ ]

2. Automobile dealers also conduct customer satisfaction surveys of car owners. The following question was asked on such a survey for a Volvo dealer:

If you had to do it over again, would you still buy a Volvo?

No	Not sure, probably wouldn't	Not sure, probably would	Yes
◯	◯	◯	◯

3. A somewhat different approach would ask consumers about their plans on a number of behaviors. A survey of consumers attending a theater activity during a special festival asked the following question:

What other activities have you done or plan to do while visiting the area? Check all that apply.

☐ Drive up the McKenzie River	☐ Golf
☐ Attend a sports event	☐ Sailing
☐ Fish	☐ River running
☐ Shop	☐ Visit the University
☐ Antiquing	☐ Visit wineries
☐ Sight-seeing	☐ Other [ ]

4a. As a final example, consider a study of the demand for newly constructed housing in a community of 110,000 people. This study included the following three questions:

When would you be most likely to purchase a new home?

- ◯ Would never be interested in purchasing a new home
- ◯ More than 3 years
- ◯ Within 2 or 3 years
- ◯ Within 1 year
- ◯ Within 6 months
- ◯ Immediately

4b. How important might each of the following reasons be in perhaps preventing you from purchasing a new home with the next two or three years?

	Not at all Important	Very Unimportant	Neither Important nor Unimportant	Very Important	Extremely Important
Might move away	◯	◯	◯	◯	◯
Happy with present home	◯	◯	◯	◯	◯
Could not afford down payment	◯	◯	◯	◯	◯
Could not afford higher monthly payment	◯	◯	◯	◯	◯
Haven't seen any new homes that I like	◯	◯	◯	◯	◯
Cheaper to rent	◯	◯	◯	◯	◯
Other: [ ]	◯	◯	◯	◯	◯

4c. How would you assess your chances of buying a newly constructed home in the area during the next two or three years, assuming you found one that met your quality and price requirements?

Zero No chance	Less than 50–50	About 50–50	Better than 50–50	100% Certain to buy
◯	◯	◯	◯	◯

decisions, since an association was found to exist between families with these characteristics and the purchase of color televisions sets.

In studies of consumers where there is a basis for believing that such associations might exist, researchers obtain information on one or more socioeconomic characteristics; those most frequently obtained are income, occupation, level of education, age, sex, marital status, and size of family. While socioeconomic characteristics are by far the most widely used bases for classification of consumers, other bases exist. Among these are attitudes, preferences, personality traits, perceived risk, and such measures of actual buying behavior as amount purchased and brand loyalty. It may be interesting to know, for example, that owners of SUVs show different personality traits than owners of other vehicles; such knowledge will be useful in marketing automobiles, however, only if it can be used to develop and evaluate appeals for each type of buyer. Doing so can enhance segmentation, positioning, and market targeting.

In general, the identification of consumer segments is useful in marketing so long as the following four statements apply:

1. **Substantial**: The value in terms of potentially increased sales makes it worthwhile to do so.
2. **Differentiable**: There are practical means of differentiating purchase behavior among market segments. There is homogeneity within and heterogeneity between segments.
3. **Operational**: There is a cost effective means of reaching the targeted market segment
4. **Responsive**: The differentiated market segments respond differentially to marketing offerings tailored to meet their needs.

Two commonly used and widely accepted classifications of consumers are by stage of the life cycle and by lifestyle. One classification identifies the household life-cycle groups as the following:

1. Young unmarrieds
2. Young marrieds, no children
3. Young marrieds, with children, youngest child under six
4. Older marrieds, with children, youngest child six or older
5. Older marrieds, with children maintaining separate households
6. Solitary survivors or older single people

Some writers have expanded the number of stages by distinguishing in the last two stages whether a person is in the labor force or retired. See Wells and Gubar (1966) and Wagner and Hanna (1983) for more detailed explanations of the life-cycle concept and marketing research.

The life-cycle stage has obvious implications with respect to purchases associated with family formation (furniture, appliances, household effects, and housing) and addition of children (food, clothing, toys, expanded housing). Other, less obvious relationships exist as well. New-car buying reaches its peak among the older married couples whose children have passed the age of six. A second stage of furniture buying takes place when children begin to date and have parties at home. Dental work, travel, and purchases of insurance are examples of service purchases associated with the life cycle.

Lifestyle has a close association with membership in a social class. It is a basis for segmenting customers by values, activities, interests and opinions, as well as by income. These differences tend to be expressed through the products bought and stores patronized, as well as the area in which one lives, club membership, religious affiliation, and other means. The media used for expression are often either consciously or subconsciously, symbolic representations of the class to which the person perceives he or she belongs (or would like to belong). When used with personality traits, lifestyle variables form the basis of psychographic classification, as illustrated in Exhibit 9.2. An illustration of psychographic questions is shown in Table 9.3.

As an example, let us consider the life styles of the Harley Owners Group (HOG). By examining a group of questions used in a segmentation study of values and motorcycle use, we find a divergent group of lifestyles that have embraced the mystique of owning a Harley.

EXHIBIT 9.2 Harley Owners Group (HOG) Classification by Psychographics

Psychographic research has suggested many different segmentation schemes. Such schemas represent interesting demographic and product markets, and provide a much more colorful description of the group as a whole as well as the diversity within.

Research by William Swinyard (1994a, 1994b) suggests that Harley-Davidson owners are a diverse group consisting of six distinct segments with very different motorcycling lifestyles:

- **Tour Glides** find the appeal of motorcycling in long distance touring. They like riding long distances, use their bike both for touring and everyday transportation, are more interested in the comfort of their motorcycle than its speed, prefer riding with a passenger, and wear a helmet.

  More than the average Harley rider, Tour Glides are religiously traditional, have somewhat old-fashioned tastes and habits, are disciplinarians with their children, like reading, and feel they live a full and interesting life. They are less ambitious than others, and are distinctively unattracted by social gatherings and danger.

- **Vanilla Dream Riders.** The Vanilla Dream Riders are more interested in the dream of motorcycling than in motorcycling itself, and are otherwise just plain vanilla—a relatively undistinguished group.

  This is the largest, oldest, wealthiest, and among the best educated segment of Harley owners, who have the newest motorcycles yet ride them least, and spend little in accessorizing them. You see the Dream Riders taking riding on short trips around town (often by themselves), wearing a helmet, and riding a stock bike. They are distinctively unaffiliated with the “live to ride” ethic, and receive relatively little psychic satisfaction from riding. Their motorcycle is merely a possession, having no real place as a “family member.” They are conservative in their moral values, marital roles, and daily behavior.

- **The Hard Core** segment is on the fringe of society, and identifies with the stereotypical biker subculture.

  They are the youngest, next-to-least well-educated, and certainly the poorest, yet spend nearly 50 percent more than any other segment on accessorizing their motorcycles. Virtually all are blue-collar workers. In relative terms, Hard Core members are much more likely than others to feel like an outlaw, and believe people would call them and their friends "dirty bikers." Note, however, that they still only "slightly agree" that these lifestyles describe them well. More than others, the Hard Core likes to be outrageous, enjoys danger, favors legalizing marijuana, and embraces the ethic of "eat, drink, and be merry, for tomorrow we die."

- **Hog Heaven** finds great psychic and spiritual satisfaction in owning and riding a Harley. More than others, these riders feel like an "old wild west cowboy" and closer to nature when they ride. They have many motorcycle friends, and when group riding, they feel the group becomes "one." They do not like helmets, and feel cars are like a "cage."

  This segment is distinctively mechanically inclined, likes to work on their motorcycles, but spends little on accessories. They have old-fashioned tastes and habits, and read relatively little. They are less likely than others to believe in a life after death, but often think about how short life really is.

- **Zen Riders** too find solace and spiritual satisfaction, but find it in motorcycling itself, and escape life's stresses in doing so.

  They include the highest percentage of married riders, but otherwise are typical of Harley owners in most demographic characteristics. More than others, Zen Riders find motorcycling fulfilling in many dimensions: their motorcycle seems alive and they like all motorcycles, from dirt bikes to four-cylinder Japanese motorcycles.

  Zen Riders are more impulsive and are willing to take chances and to run risks. They believe they are more ambitious than other segments. They like to party, and have trouble relaxing in everyday life, but they are "modern" husbands and are opposed to legalizing marijuana.

- **Live to Ride** rides more than any other segment, and motorcycles represent a total lifestyle to them. They "ride to live and live to ride."

  Members of this, the smallest segment, are most likely to have bought their motorcycle new, and ride it the most by a wide margin. They simply love riding. More than other Harley owners, they use their bike for everyday transportation, enjoy riding long distances, and use their bike for touring. They find riding to be a magical experience, and motorcycling is a total lifestyle to them.
  If they did not have family, members of this segment would quit their jobs and take off. They agree with an "eat, drink, and be merry" premise, like to create a stir, like danger, and get lots of satisfaction from their hobbies. They care little about their appearance and tend not to believe in a life after death.

Life style segmentation offers significant implications for marketing and advertising to these segments in terms of not only better communicating with each group, but in creating the product configuration and accessories that appeal to each group.

A common designation of social classes is the one originally used by Warner and others (1960) that designates the familiar *upper*, *middle*, and *lower* class designations, each divided into upper and lower segments. Thus, the Warnerian classification results in six classes, ranging from the UU (upper upper) down through the LL (lower lower). Somewhat newer is the value and lifestyles (VALS) schema, which classifies the American public into: survivors, sustainers, belongers, emulators, achievers, I-am-me's, experientials, societally conscious, and integrateds (Mitchell, 1983).

EXHIBIT 9.3 Psychographic Measurement Questions for Three Motorcycle Owner Segments: Most and Least Frequent Motorcycle Lifestyle Descriptors

	Most Frequently Agreed With	Least Frequently Agreed With
TOUR GLIDERS	I like long-distance touring bikes. I use my bike for touring. My bike is made more for comfort than for speed. I love to ride long distances . . . to me, 500 miles is a short trip. I like bikes with plastic farings and engine covers. I like good bikes no matter where they are made. I usually ride with someone on the back of my bike. I like it best when someone is on my bike with me. When I ride I wear leather boots. I use my bike for everyday transportation.	My bike is really quick. I only wave at other riders on bikes like mine. I like to ride aggressively. I have spent a lot of money modifying my bike. I like to have my bike look really different. I don't pay much attention to what I wear when I ride. Most of the time, my motorcycle is just parked. I have spent lots on speed modifications for my bike. I get excited about motocross or scrambling. I like dirt bikes.
DREAM RIDERS	Most of the time, my motorcycle is just parked. I like wearing a helmet when I ride. I don't know many other people that ride motorcycles. My bike is pretty much stock. I mainly use my bike for short trips around town. To me, a motorcycle is just transportation. I don't pay much attention to what I wear when I ride. All things considered, I think Japanese bikes are the best. Hot 4-cylinder bikes sound fantastic. I like to ride alone.	It's true that "I live to ride and ride to live." Riding, to me, is often a magical experience. To me, motorcycles are a symbol of freedom. Motorcycles are a total lifestyle to me. My bike is everything to me. When I am riding in a group, the group almost becomes one. When I'm on my bike it's sometimes a spiritual experience. When I'm on my bike, people seem to be admiring me. I spend most of my free time with my bike buddies. I like to have my bike look really different.
HARD CORE	Some people would call me and my friends "outlaws." I have spent lots on speed modifications for my bike. Sometimes I feel like an "outlaw." Some people would call me a "dirty biker." I think it's true that "real men wear black." My bike is everything to me. I have spent a lot of money modifying my bike. I spend most of my free time with my bike buddies. Motorcycles are a total lifestyle to me. I like tattoos.	My bike is pretty much stock. Most of the time, my motorcycle is just parked. I like wearing a helmet when I ride. I don't know many other people that ride motorcycles. I like bikes with plastic farings and engine covers. I like good bikes no matter where they are made. I like the spacecraft look of some bikes today. Hot 4-cylinder bikes sound fantastic. My bike is made more for comfort than for speed. I mainly use my bike for short trips around town.

Although less direct and more subtle than life-cycle stage in its effect on overt buying behavior, there can be little question that an upper-middle-class household will show more similarity in purchasing and consumption patterns of food, clothing, furniture, and housing to another upper-middle-class household than it will to a blue-collar, upper-lower-class household. The media to which the managerial-professional, upper-middle-class family is exposed, and the appeals to which it responds, are also likely to be closer to those of other managerial-professional families than to those of the blue-collar family. Similarly, on the basis of VALS, a processor and packager of tofu may find that because "experientials" have a greater appreciation for natural things, they are heavier users of tofu. The marketer can then direct the product at this lifestyle group.

Another approach to values and lifestyles is the List of Values (LOV) developed by Kahle (1983). These values fit well into the qualitative research discussed in Chapter 6, (means-end laddering to produce hierarchical value maps and ZMET analysis):

- Self-respect
- Security
- Warm relationships with others
- Sense of accomplishment
- Self-fulfillment
- Being well-respected
- Sense of belonging
- Fun and enjoyment in life

Classification of consumers is vital if we are to learn more about consumer behavior and utilize this information to develop more efficient marketing techniques. But caution is needed to ensure that managers do not segment too finely and use categorizations when they should not. Given the new software available that simplifies using such advanced techniques as neural networks, latent-class models, fuzzy or overlapping clustering, and even occasion based segmentation, there will be a tendency to use a backhoe when a shovel is more appropriate. The technology in this area is far ahead of other aspects of the marketing process, and ahead of most managers' needs (Freeman, 2001). Advanced statistical techniques for consumer classification purposes are discussed in later chapters.

Although the discussion thus far has focused on consumers, similar classification requirements exist and are used in studies of industrial users and marketing intermediaries. Comparable characteristics of these firms include sales volume, number of employees, and the type of products manufactured or handled.

Extent of Knowledge

Prediction of what actions respondents will take is often aided by knowing "how much they know." This is especially so when making advertising budget and media allocation decisions, where consumers' choices are strongly affected by their levels of awareness and extent of knowledge of the product and its attributes. To illustrate questions measuring consumer awareness, a study of homeowners examined their knowledge of conventional and variable-rate mortgages. One question sought knowledge of interest rate differences:

If you were offered both types of mortgages, indicate the difference, if any, between the interest rate for the fixed-rate plan and the initial interest rate for the variable-rate plan.

Fixed rate was higher	No difference	Variable rate was higher	Cannot recall	Did not inquire
○	○	○	○	○

Attitudes and Opinions

Investigators in the behavioral science fields of psychology, sociology, and political science have made extensive studies of attitudes and opinions over a wide range of subject areas. The study of people's behavior in business and economic contexts is also a behavioral science. As such, it has been a natural consequence that marketing research has adopted, adapted and applied many techniques, including attitude-opinion studies to obtain information applicable to the solution of marketing problems.

The terms *attitude* and *opinion* have frequently been differentiated in psychological and sociological investigations. A commonly drawn distinction has been to view an attitude as a predisposition to act in a certain way, and an opinion as a verbalization of an attitude. Thus, a statement by a respondent that he or she prefers viewing Blue-Ray DVD's to HDTV color television programs would be an opinion expressing (one aspect of) the respondent's attitude toward high-definition television.

When used to predict actions that the respondent will take, this distinction between attitude and opinion rapidly becomes blurred. Since the major purpose of attitude-opinion research in marketing is to predict behavior, this differentiation is, at best, of limited usefulness. We shall therefore use the terms interchangeably. Attitude research in marketing has been conducted with the use of both qualitative and quantitative techniques. In either form, researchers encounter problems that are more severe than those involved in obtaining any other type of descriptive information discussed. Despite these problems, which we will discuss in later chapters in some detail, attitude-opinion research has been widely used to provide information for choosing among alternatives. Its greatest use has been in the areas of product design (including packaging and branding) and advertising. Other uses have been in selecting store locations, developing service policies, and choosing company and trade names. In fact, attitudes and opinions are central in customer satisfaction studies.

Measuring Attitudes

Expectancy Value Measures of Behavioral Intention (BI) and Attitudes (A)

Expectancy value models were first developed in the 1960's as a method of predicting behavioral intentions (a precursor of actual behavior). Expectancy value models use attitudes to predict behavioral intention (intention to try, purchase, recommend, or re-purchase a product or service). This methodology has become a mainstay of marketing research and is found to perform well in predicting both consumer behavior and consumer satisfaction/dissatisfaction.

The Expectancy value model uses attitudes and beliefs in a mathematical formulation that is read as follows:

- *Behavior (purchase of brand X) is approximated by Behavioral Intention (intention to purchase brand X), which in turn is approximated by the Overall Attitude toward brand X.*
- *The Overall Attitude toward brand X equals the sum of all salient attitudes about brand X, weighted by how important each attitude is in the purchase decision process.*
- *Normative beliefs are added to this compensatory evaluative process by considering the norms surrounding an attitude (My friends insist I should buy an environmentally friendly car that gets 50 mpg), and applying weights reflecting the buyer's motivation to comply with the norm (Will they praise me for buying a 50 mpg car? – or Do I really care about getting 50 mpg?).*

$$B \approx BI \approx A_o = w_1 \sum_{i=1}^{k} a_i * b_i + w_2 \left(\sum_{i=1}^{k} nb_i * mc_i \right)$$

where:

B = Behavior

BI = Behavioral Intention, which is measured using a question such as "how likely are you to buy a Toyota Prius sometime during the next year?" with a five or seven-point Likert or semantic differential scale labeled "definitely will purchase" and "definitely will not purchase" at the endpoints.

A_o = Overall Attitude toward the object (brand, product, service or company)

w_1, w_2 = Weights that indicate the relative influence of the overall attitude toward the object and the normative influence to purchase the product

$\sum a_i * b_i$ = the sum of individual attitudes toward the object, weighted by importance. The overall attitude is formed by the multiplicative product of a_i (the person's affective evaluation - liking of the attribute *i*), and b_i (here defined as the importance of attribute *i* in the purchase decision). The sum is taken over the *k* attributes that are defined as salient in the purchase decision.

$\sum nb_i * mc_i$ = The normative component of the decision process that asks "should the norms surrounding this attribute be important to me?" This is computed as the multiplicative product of nb_i (the norms governing attitude *i*), and mc_i (the motivation of the respondent to comply with those norms).

Attitude (a_i*b_i)

a_i is the affective (liking) component of the evaluation of attribute *i*. A representative measure of a_i would be using the context "In terms of buying a Toyota Prius", to evaluate "Gets 50 miles/gallon". The evaluation would use a five or seven point scale with endpoints ranging from "Poor" to "Excellent", or "Not at all Desirable" to "Very Desirable". This is found on the right side of Figure 9.1.

b_i - the importance of attribute I in the context of behavior B. This is sometimes measured as the probability of attribute i being associated with brand X. At other times it is measured as the importance of attribute i in achieving behavior B. Using the context of purchasing a Toyota Prius, the attribute would be "Gets 50 miles per gallon" could be rated on a seven point scale with endpoints labeled "Very Unlikely" and "Very Likely", or as in the example below, is measured as "Not at all Important" to "Very Important". This is found on the left side of Figure 9.1.

Figure 9.1 Expectancy Value Importance-Performance Rating Scales

Assume you are purchasing a Toyota Prius. Please evaluate the following attributes using the Importance scale on the left and the performance scale on the right.

	Importance				Desirability			
	Not at all Important	Somewhat Unimportant	Somewhat Important	Very Important	Not at all Desirable	Somewhat Undesirable	Somewhat Desirable	Very Desirable
Gets 50 miles/gallon	○	○	○	○	○	○	○	○
Has a 6 year/60,000 Mile drive train warranty	○	○	○	○	○	○	○	○

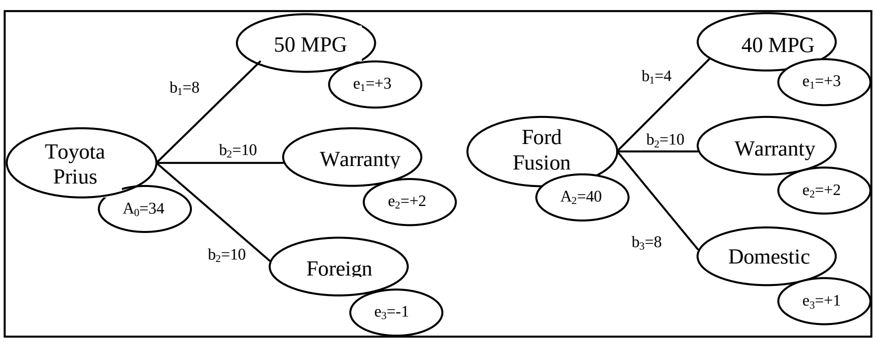
$\Sigma\, nb_i * mc_i =$ The overall normative component of the decision process.
This is computed as the multiplicative product of nb_i (the norms governing belief i), and mc_i (the motivation of the respondent to comply with those norms) see Figure 9.2.

Figure 9.2 Normative Rating Scales

Assume you are purchasing a Toyota Prius. How much influence would your friends and family have on the attributes of you car purchase decision? Please evaluate the following attributes using the scale on the left and the right.

	In buying a Toyota Prius, my friends and family are very favorable toward the following attributes...				I will follow my friends and family advice in buying a Toyota Prius that...			
	Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree
Gets 50 miles/gallon	○	○	○	○	○	○	○	○
Has a 6 year/60,000 Mile drive train warranty	○	○	○	○	○	○	○	○

The normative component is often considered to be absorbed in the overall attitude component and depending on the application, is often ignored.



$$Attitude_{Prius} = \sum_{i=1}^{3} b_i e_i$$
$$A_e = (8)(3) + (10)(2) + (10)(-1)$$
$$A_e = 24 + 20 - 10$$
$$A_e = 34$$

$$Attitude_{Fusion} = \sum_{i=1}^{3} b_i e_i$$
$$A_e = (4)(3) + (10)(2) = (8)(1)$$
$$A_e = 12 + 20 + 8$$
$$A_e 40$$

The expectancy value model is an aggregation of attitudes and beliefs about the most important attributes that predict behavior. This compensatory type of decision model results in a powerful approach to measuring customer attitudes and predicting customer behavior, and is well used in research. It is not surprising that many popular methodologies, including conjoint analysis, are theoretically based on this measurement model.

Measuring Satisfaction

Customer satisfaction is the most common of all marketing surveys and is part of the "big three" research studies in marketing that include market segmentation and concept testing. Customer satisfaction has been defined as the state of mind that customers have about a company and its products/services when their expectations have been met or exceeded over the lifetime of product or service use (Cacioppo, 2000, p. 51). The positive attitudes that lead to customer satisfaction usually, in turn, lead to customer loyalty and product repurchase. But measuring satisfaction is not measuring loyalty. The following are typical measures of overall satisfaction:

1. Overall, how satisfied are you with (brand name)?
2. Would you recommend (brand name)?
3. Do you intend to repurchase (brand name)?

According to William Neal (2000), these questions are usually measuring the same thing—satisfaction with the product or service. Satisfaction is a condition that is necessary, but not sufficient to predict loyalty. Customer satisfaction can be measured by traditional surveys, by using comment cards, and for business-to-business situations by field or call reports, to name just some of the methods (Cochran, 2001). Many companies routinely measure satisfaction in consumer tracking studies (see Exhibit 9.4).

The attitudes and opinions of prospective buyers will clearly affect expectations and the resulting satisfaction in a purchase decision. Consequently, the marketing manager should be as well informed as possible about both the nature of their relevant attitudes and opinions and the intensity with which they are held. Subaru of America, for example, has a program that includes a Purchase Experience Survey and a Service Experience Survey that goes out to all customers who have purchased a Subaru or had it serviced. Findings from these surveys are reported back to the dealer, who then acts on them. In addition, Subaru sends out a Product Survey to a sample of new Subaru owners every year. This survey examines the quality of the product and whether new owners are satisfied with the performance, fit, and finish of their new vehicle.

Measuring satisfaction and building a satisfaction survey requires at least a basic knowledge of the satisfaction measurement literature, combined with knowledge and experiences with the company's customers. We will first introduce the theoretical and methodological underpinnings of satisfaction research by first defining the concept of customer satisfaction and how satisfaction is used in business. Next, different satisfaction survey measures are discussed and presented. Finally, the components of a satisfaction survey are presented, along with sample satisfaction survey questions. The Qualtrics.com survey library contains sample satisfaction questionnaires, questions and scales. This discussion of satisfaction research provides the basis for understanding how to measure satisfaction and why the suggested should be used.

Exhibit 9.4 Common Ingredients of a Customer Satisfaction Survey

Product Use
Frequency of product use
Primary use location
Primary precipitating events or situations for product use or need
Usage rates and trends

Product Familiarity
Degree of actual product use familiarity
Knowledge (read product information, read product label, etc.)
Knowledge and Involvement with product and the purchase process
Awareness of other brands
Reasons for original product purchase (selection reasons)
Primary benefits sought from the product

Product Evaluation
Attribute evaluation matrix: (quality, price, trust, importance, performance, value)
Perceived benefit associations matrix
Importance, performance
Identification of primary benefits sought
Comparison to other brands (better, worse)
What is the best thing about the brand, what could be done better

Message and Package Evaluation
Packaging size, design
Advertising Promise, message fulfillment evaluation

Value Analysis
Expectation of price
Expectation of relative price (full price, on sale)
Current price paid

Satisfaction Measurements
Overall Satisfaction
Reasons for Satisfaction Evaluation
Satisfaction with attributes, features, benefits
Satisfaction with use
Expected and Ideal Satisfaction-Performance Measures

Likelihood of recommending
Likelihood of repurchasing

What Is Customer Satisfaction?

Customer satisfaction measures indicate how well a company's products or services meet or exceed customer expectations. These expectations will reflect many aspects of the company's business activities including the actual product, service and company. Customer satisfaction measures will tap the customer's lifetime of product and service experience.

Why is Customer Satisfaction So Important?

Effective marketing focuses on two activities: retaining existing customers and adding new customers. Customer satisfaction measures are critical to any company, be it consumer or B2B, because customer satisfaction is a strong predictor of customer retention, customer loyalty and product repurchase.

When to Conduct Customer Satisfaction Surveys

When to measure customer satisfaction depends on the kind of product or service provided, the kinds of customers served, the number of customers served, the longevity and frequency of customer/supplier interactions, and the intended use of the results.

Three very different approaches may each produce meaningful and useful results that are appropriate for specific situations, uses and needs:

- **Post Purchase Evaluations** reflect the satisfaction of the individual customer at the time of product or service delivery (or shortly afterwards). This type of satisfaction survey is typically used as part of a CRM (Customer Relationship Management System) and focuses on securing a long term relationship with the individual customer

- **Periodic Satisfaction Surveys** provide an occasional snapshot of customer experiences and expectations, and are conducted for specific groups of consumers on a periodic basis.

- **Continuous Satisfaction Tracking** is often part of a management initiative to assure quality is at high levels over time, and can involve post purchase evaluations or a succession of periodic satisfaction surveys conducted on a regular basis (daily, quarterly or monthly basis). Satisfaction feedback is obtained from the individual customer at the time of product or service delivery (or shortly afterwards). Satisfaction tracking surveys are.

Satisfaction surveys are developed to provide an understanding of customers' expectations and satisfaction. Satisfaction surveys typically require multiple questions that address satisfaction with different dimensions of the product or service concept. That is, satisfaction measurement includes measures of overall satisfaction, satisfaction with individual product and service attributes, and satisfaction with the benefits recorded as a result of purchase. Satisfaction measurement is like peeling away layers of an onion-each layer reveals yet another deeper layer, closer to the core.

Each of the three methods of conducting satisfaction surveys are helpful in obtaining customer feedback for assessing overall accomplishments, degree of success, and areas for improvement. Ulwick (2005) advocates the combination of satisfaction and importance to identify unfulfilled opportunities in the marketplace. In this case, performance gaps not being adequately served by the products in the market are identified, evaluated for feasibility, and targeted for development.

Customer Satisfaction Survey Measures

Customer satisfaction surveys will often include several different multiple measures of satisfaction, including:

- Overall measures of customer satisfaction
- Affective measures of customer satisfaction
- Cognitive measures of customer satisfaction
- Behavioral measures of customer satisfaction
- Expectancy value measures of customer satisfaction

We will now explain each in detail. Because general measures of customer satisfaction usually involve product fulfillment, we will discuss product use scenarios focusing on where and how the product is used.

Satisfaction Measurement: Overall Measures of Satisfaction

Like when we have a great food experience at a favorite local restaurant, elevated levels of customer satisfaction usually leads to customer loyalty and product repurchase. But measures of satisfaction are different than measures of loyalty. Satisfaction measurement questions typically include items like those found in Exhibit 9.5 and like the following:

1. An overall satisfaction measure (emotional):
   Overall, how satisfied are you with "Yoni fresh yogurt"?

   This question reflects the overall opinion of a consumer's satisfaction experience. It is noteworthy that we can meaningfully measure attitudes towards a product that a consumer has never used, but we cannot measure satisfaction for a product or brand that has never been used. There is no experience base for such a measure of satisfaction.

2. A loyalty measure (affective, behavioral):
   Would you recommend "Yoni" to your family and friends?

3. A series of attribute satisfaction measures (affective and cognitive):
   How satisfied are you with the "taste" of Yoni fresh yogurt?
   How important is "taste" in your decision to select Yoni fresh yogurt?

   Satisfaction and attitude are closely related concepts. The psychological concepts of attitude and satisfaction may both be defined as the evaluation of an object and the individual's relationship to it. The distinction is that here, satisfaction is a "post experience" state representing the emotional affect produced by the product's quality or value.

4. Intentions to repurchase (behavioral measures):
   Do you intend to repurchase Yoni fresh yogurt?

   Satisfaction can influence post-purchase/post-experience actions other than usage (these other actions might include word of mouth communications and repeat purchase behavior). Additional post-experience actions might reflect a heightened level of product involvement that in turn results in increased product or information search activity, reduced trial of alternative products, and even changes in preferences for shopping locations and choice behavior.

As shown in Figure 9.3, customer satisfaction is diagrammatically shown to be influenced by perceived quality of product and service attributes, associated product features and benefits, and is moderated by the customer's expectations regarding the product or service. The researcher may want to define and develop measures for each of these constructs that influence customer satisfaction.

Affective Measures of Customer Satisfaction

Attitudes toward a product can be developed as a result of product information or any experience with the product, whether perceived or real. Affect (liking/disliking) is best measured in the context of product attributes or benefits. Again, it may be meaningful to measure attitudes towards a product or service that a consumer has never used, but it is not meaningful to measure satisfaction when a product or service has not been used.

Figure 9.3 Building a Customer Satisfaction Survey

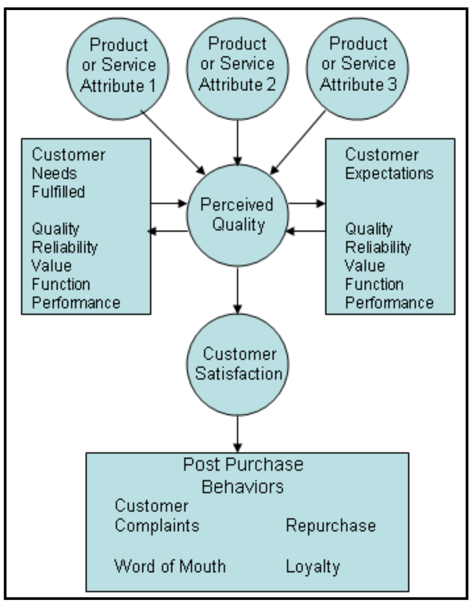
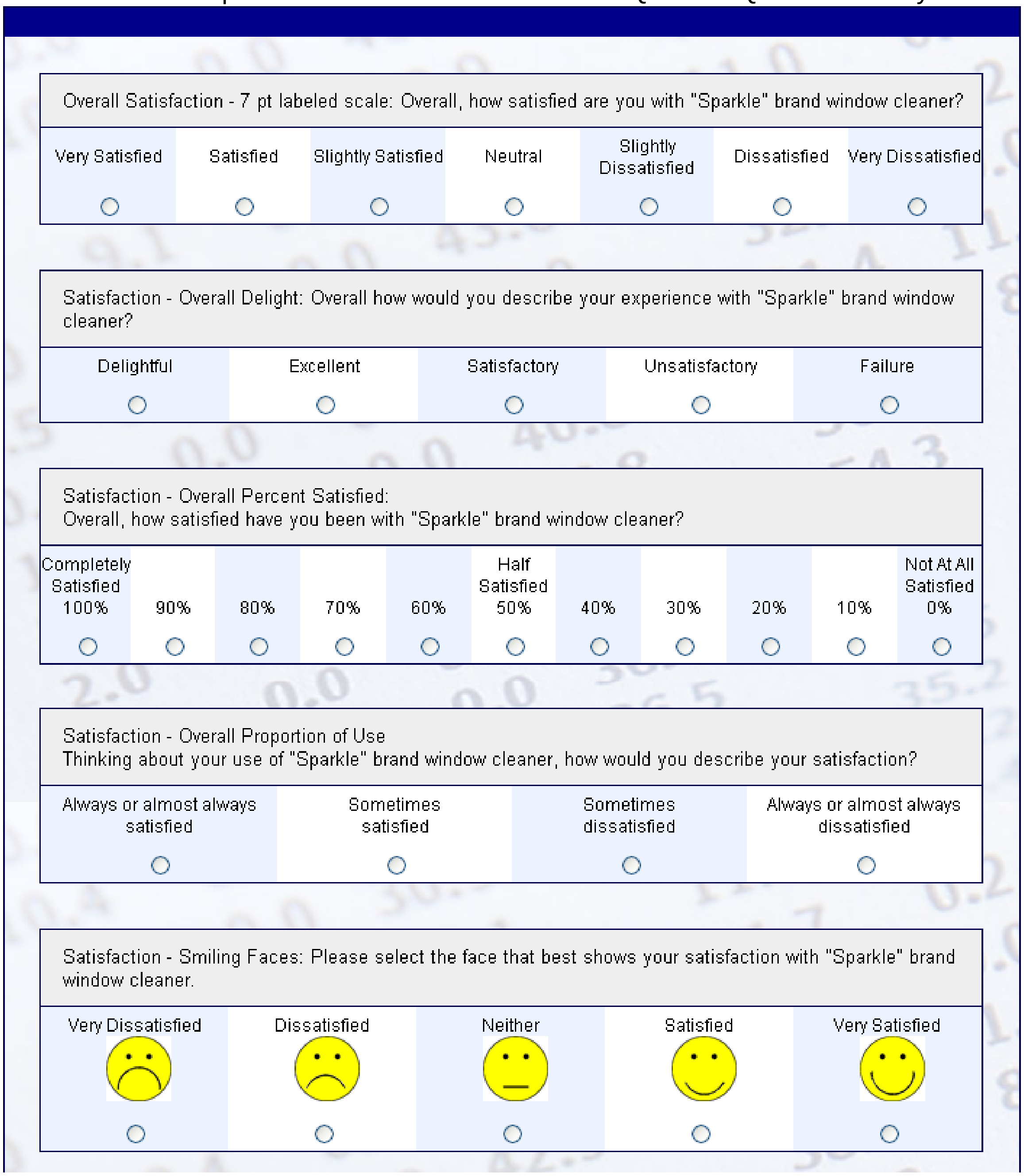


Exhibit 9.5 Sample Satisfaction Measures from the Qualtrics Question Library



Overall Satisfaction - 7 pt labeled scale: Overall, how satisfied are you with "Sparkle" brand window cleaner?

Very Satisfied	Satisfied	Slightly Satisfied	Neutral	Slightly Dissatisfied	Dissatisfied	Very Dissatisfied
○	○	○	○	○	○	○

Satisfaction - Overall Delight: Overall how would you describe your experience with "Sparkle" brand window cleaner?

Delightful	Excellent	Satisfactory	Unsatisfactory	Failure
○	○	○	○	○

Satisfaction - Overall Percent Satisfied:
Overall, how satisfied have you been with "Sparkle" brand window cleaner?

Completely Satisfied 100%	90%	80%	70%	60%	Half Satisfied 50%	40%	30%	20%	10%	Not At All Satisfied 0%
○	○	○	○	○	○	○	○	○	○	○

Satisfaction - Overall Proportion of Use
Thinking about your use of "Sparkle" brand window cleaner, how would you describe your satisfaction?

Always or almost always satisfied	Sometimes satisfied	Sometimes dissatisfied	Always or almost always dissatisfied
○	○	○	○

Satisfaction - Smiling Faces: Please select the face that best shows your satisfaction with "Sparkle" brand window cleaner.

Very Dissatisfied	Dissatisfied	Neither	Satisfied	Very Satisfied
○	○	○	○	○

Satisfaction - Pizza Scale:
Imagine that the following circles represent your satisfaction with "Sparkle" brand window cleaner.
Circle 1 has NO PLUS Marks and represents Very Dissatisfied with "Sparkle" brand window cleaner. Circle 9 has all plus marks in it, to represents Very Satisfied with "Sparkle" brand window cleaner. Other circles are in between.

Please select the circle that best represents your evaluation of "Sparkle" brand window cleaner.

Very Dissatisfied								Very Satisfied
○	○	○	○	○	○	○	○	○

Satisfaction : Performance-Importance Gap Analysis

Please indicate your satisfaction with each of the following attributes of your new Ford							**Please check the aspects that were IMPORTANT in determining your rating of this feature. Check as many as made a significant difference.**			
Noise	○	○	○	○	○	→	Road Noise ☐	Wind Noise ☐	Engine Noise ☐	Squeaks and Rattles ☐
Power	○	○	○	○	○	→	Start from a Stop ☐	Acceleration ☐	Passing ☐	Towing Capacity ☐
Interior Comfort	○	○	○	○	○	→	Seat Comfort ☐	Materials ☐	Dash Layout ☐	Color Schemes ☐

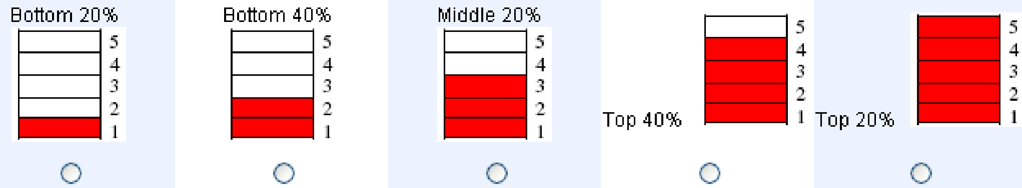
Expectations - Implicit Evaluation: Compared to your expectations, how well did "Sparkle" brand window cleaner perform?

Too High:It was poorer than I thought		Accurate:It was just as I expected		Too Low:It was better than I thought
○	○	○	○	○

Expectations - Relative Comparison: Compared to WINDEX, would you say that "Sparkle" brand window cleaner performed

Much better than WINDEX	Somewhat better than WINDEX	About the same as WINDEX	Somewhat worse than WINDEX	Much worse than WINDEX
○	○	○	○	○

Expectations - Ladder Scale Comparison with other brands:
Compared to all other brands, how would your evaluate "Sparkle" brand window cleaner ?



Satisfaction - Behavioral Evaluation:
In thinking about your experience with "Sparkle" brand window cleaner, please indicate your agreement with the following questions:

	Strongly Disagree	Sometimes Disagree	Neither Disagree Nor Agree	Sometimes Agree	Strongly Agree
If I had it to do over again, I would purchase "Sparkle" brand window cleaner	○	○	○	○	○
My choice to buy "Sparkle" brand window cleaner was a good one.	○	○	○	○	○
I feel bad about my decision concerning "Sparkle" brand window cleaner	○	○	○	○	○
I think that I did the right thing when I decided to buy "Sparkle" brand window cleaner	○	○	○	○	○
I am not happy that I purchased "Sparkle" brand window cleaner	○	○	○	○	○

Customer Satisfaction - Expectancy Value Measure:
Please tell us about your experience with "Sparkle" brand window cleaner by evaluating the performance on the following attributes, and by indicating how important that attribute is to you.

	Would you say the performance was...					How IMPORTANT are each of the following features in your evaluation?				
	Poor	Fair	Good	Very Good	Excellent	Very Important	Somewhat Important	Neither	Somewhat Unimportant	Very Unimportant
Cleaning Ability	○	○	○	○	○	○	○	○	○	○
Pleasant Scent	○	○	○	○	○	○	○	○	○	○
Easy to Use Dispenser	○	○	○	○	○	○	○	○	○	○

Please tell us your agreement with the following statements about how well "Sparkle Software" fulfills your Graphics Design Software needs

	Very Dissatisfied	Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied	Very Satisfied
Quality of Products	○	○	○	○	○	○	○
Value Received	○	○	○	○	○	○	○
Timeliness of Delivery	○	○	○	○	○	○	○
Efficiency of Sales Representatives	○	○	○	○	○	○	○
Ease of Access to the Store	○	○	○	○	○	○	○
Shopping Environment	○	○	○	○	○	○	○
Teamwork with Your Account Rep	○	○	○	○	○	○	○
Customer Service	○	○	○	○	○	○	○
Commitment to helping you	○	○	○	○	○	○	○
Innovation	○	○	○	○	○	○	○

Satisfaction - Likelihood of Recommending
How likely are you to recommend "Sparkle" brand window cleaner to other family members or close friends?

Very Unlikely	Somewhat Unlikely	Somewhat Likely	Very Likely
○	○	○	○

Satisfaction - Likelihood of Repurchase
How likely are you to repurchase "Sparkle" the next time you need window cleaner?

Very Unlikely	Somewhat Unlikely	Somewhat Likely	Very Likely
○	○	○	○

Powered by Qualtrics.com Survey Software for Success

Cognitive Measures of Customer Satisfaction

A cognitive element is defined as an appraisal or conclusion that the product was useful (or not useful), fit the situation (or did not fit), exceeded the requirements of the problem/situation (or did not exceed), or was an important part of the product experience (or was unimportant). Cognitive responses are often specific to the situation for which the product was purchased and specific to the consumer's intended use of the product, regardless if that use is correct or incorrect.

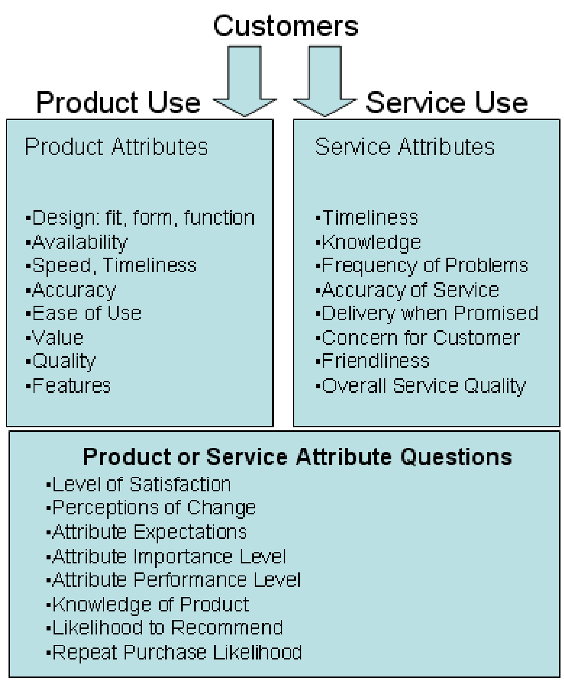
Behavioral Measures of Customer Satisfaction

It is sometimes believed that dissatisfaction is synonymous with regret or disappointment while satisfaction is linked to ideas such as, "it was a good choice" or "I am glad that I bought it." When phrased in behavioral response terms for a future or hypothetical behavior, consumers indicate that "purchasing this product would be a good choice" or "I would be glad to purchase this product." Often, behavioral measures reflect the consumer's past experience with individuals associated with the product (i.e. customer service representatives) and the intention to repeat that experience.

Expectations Measures of Customer Satisfaction

As might be expected, many different approaches to measuring satisfaction exist in the consumer behavior literature. Parasuraman, Ziethaml and Berry (1998) refined their earlier work to identify five generic dimensions of a service satisfaction scale called SERVQUAL: tangibles, reliability, responsiveness, assurance and empathy. SERVQUAL has become the standard for service quality measurement, but it should be recognized that satisfaction dimensions seem to vary depending on the application (high tech vs. health care vs. leisure services, etc.). None the less, this important work is often used for the measurement of customer satisfaction in a service environment.

Figure 9.4 Product and Service Questions



A slightly different diagnostic approach to satisfaction measurement is one that examines the gap between the customer's expectations of performance and the consumer's actual experience. This "satisfaction gap" involves measuring both perception of performance and expectation of performance along specific product or service attributes dimensions. This approach can be applied to an individual product or service experience, a brand, or even a complete product category, as proposed by Ulwick (2005).

Again we see that customer satisfaction is largely a reflection of both expectations and experiences that the customer has with a product or service. Expectations may reflect past product experiences, but will also reflect the purchase evaluation process that occurs when shopping for a product or service. For example, when we make major purchases, we research the product or service, collecting information from the advertising, salespersons, and word-of-mouth from friends and associates. This information influences our expectations and ability to evaluate quality, value, and the ability of the product or service to meet our needs.

Customer Expectations that Influence Satisfaction

Customers hold both explicit and implicit performance expectations for attributes, features and benefits of products and services. The nature of these expectations will dictate the form and even the wording of a satisfaction questionnaire.

Explicit expectations are mental targets for product performance, such as well identified performance standards. For example, if expectations for a color printer were for 17 pages per minute and high quality color printing, but the product actually delivered 3 pages per minute and good quality color printing, then the cognitive evaluation comparing product performance and expectations would be 17 PPM – 3 PPM + High – Good, with each item weighted by their associated importance.

Implicit expectations represent the norms of performance that reflect accepted standards established by business in general, other companies, industries, and even cultures. An implicit reference might include wording such as "Compared with other companies…" , or "Compared to the leading brand…"

Static performance expectations address how performance and quality for a specific application are defined. Performance measures for each application are unique, though general expectations relate to quality of outcome and may include those researched by Parasuraman or others such as: accessibility, customization, dependability, timeliness, and accuracy, tangible cues which augment the application, options, cutting edge technology, flexibility, and user friendly interfaces. Static performance expectations are the visible part of the iceberg; they are the performance we see and -- often erroneously – are assumed to be the only dimensions of performance that exist.

Dynamic performance expectations are about how the product or service evolves over time and includes the changes in support and product or service enhancement needed to meet future business or use environments. Dynamic performance expectations may help to produce "static" performance expectations as new uses, integrations, or system requirements develop and become more stable.

Technological expectations focus on the evolving state of the product category. For example, mobile phones are continually evolving. Mobile service providers, in an effort to deal with the desire to switch to new technology phones, market rate plans with high cancellation penalties for switching providers, but with liberal upgrade plans for the phones they offer. The availability of low profile phones with email, camera, MP3, email, blue tooth technology, and increased storage will change technology expectations as well as the static and dynamic performance expectations of the product. These highly involving products are not just feature based, but have expectations that enhance perceptions of status, ego, self-image, and can even invoke emotions of isolation and fear when the product is not available.

Interpersonal expectations reflect the relationship between the customer and the product or service provider. Person to person relationships are increasingly important, especially where products require support for proper use and functioning. Expectations for interpersonal support include technical knowledge and ability to solve the problem, ability to communicate, time to problem resolution, courtesy, patience, enthusiasm, helpfulness, assurance that they understood my problem and my situation, communication skills, and customer perceptions regarding professionalism of conduct, often including image, appearance.

Situational Expectations

In building a customer satisfaction survey, it is also helpful to consider reasons why pre-purchase expectations or post-purchase satisfaction may or may not be fulfilled or even measurable. In some situations the following conditions may occur:

1) Expectations may not reflect unanticipated service attributes;
2) Expectations may be quite vague, creating wide latitude of acceptable performance and expected satisfaction;
3) Expectation and product performance evaluations may not be cognitive, but instead sensory, as in expectations of taste, style or image. Such expectations are not only difficult to evaluate and understand, but may change over time.
4) The product use may attract so little attention as to produce no conscious affect or cognition (evaluation), and result in measures that are meaningless satisfaction or dissatisfaction measures;
5) There may have been unanticipated benefits or consequences of purchasing or using the product (such as a use, usage situation, or feature not anticipated with purchase);
6) The original expectations may have been unrealistically high or low;
7) The product purchaser, influencer and user may have each been a different individual, each having different expectations.

When fulfilled, expectations result in customer satisfaction (or when expectations are not fulfilled, result in dissatisfaction and complaining behavior). Each of the above discussed types of expectation should be considered in the context of the unique research project to determine if special consideration is warranted during questionnaire development. The research study may also benefit from consideration of expectations related to perceived quality and value.

Satisfaction Measurement: Perceived Quality Measures

Perceived quality is often measured through three measures: overall quality, perceived reliability, and the extent to which a product or service is able to fulfill the customer's needs. Customer experiences that result in attributions of quality are the single greatest predictor of customer satisfaction.

Satisfaction Measurement: Perceived Value Measures

Perceived value may conceptually be defined as the overall price divided by quality or the overall quality divided by price. Perceived value is measured in many ways including overall evaluation of value, expectations of price that would be paid, and more rigorous methodologies including the Van Westendorp pricing analysis and conjoint analysis. Note that these are Qualtrics advanced option question types, and that the Qualtrics online survey university contains white papers and tutorials about these topics.

The consumer behavior literature has long shown that price is a primary indicator of quality when other attributes and benefits are relatively unknown. However when repeat purchases are made in some product categories, price may be reduced in importance. This may reflect a decision simplification strategy that makes shopping or purchase decisions easier, or may reflect increased customer loyalty.

Satisfaction Measurement: Customer Loyalty Measures

Customer loyalty reflects the likelihood of repurchasing products or services. Customer satisfaction is a major predictor of repurchase, but is strongly influenced by explicit performance evaluations of product performance, quality, and value.

Loyalty is often measured as a combination of measures including overall satisfaction, likelihood of repurchase, and likelihood of recommending the brand to a friend. For a common measure of loyalty might be the sum of the following three questions measured on an agreement scale (Strongly agree – Strongly disagree):

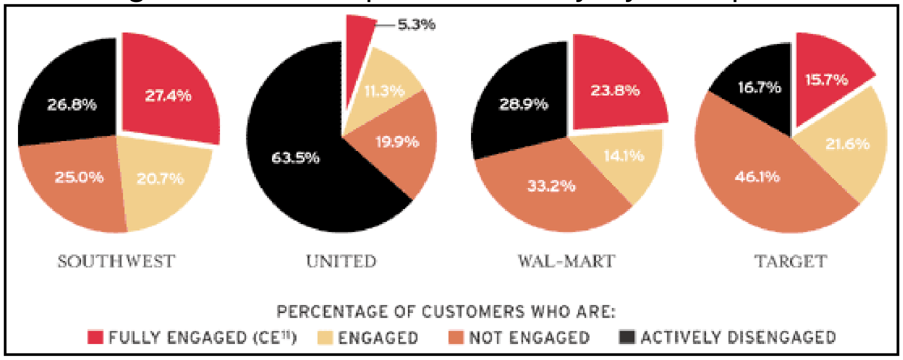
1. Overall, how satisfied are you with [brand]?
2. How likely are you to continue to choose/repurchase [brand]?
3. How likely are you to recommend [brand] to a friend or family member?

Measures of loyalty are generally not perfect predictors of the brand switching that often occurs in an actual consumer purchase situation. Switching may occur because of out of stock situations, inconvenience of going to a store that has the preferred brand, or the lack of commitment to the brand. This latter condition has been addressed by Gallup in the identification of four emotional states that are indicative of the degree of commitment and brand loyalty (Table 9.3).

Table 9.3 Brand Loyalty Measures

Construct	*Measure*
Confidence in the brand	[Brand] is a name I can always trust. [Brand] always delivers on what they promise.
Integrity of the brand	[Brand] always treats me fairly. If a problem arises, I can always count on [brand] to reach a fair and satisfactory resolution
Pride in the brand	I feel The to be a [brand] customer. [Brand] always treats me with respect.
Passion for the brand	[Brand] is the perfect company for people like me. I can't imagine a world without [brand].

Figure 9.5 Gallup Customer Loyalty Example



The application of these loyalty scales enables the differentiation of brands, companies and even industries according to how well they are "engaging" their customers (Figure 9.5).hat

This approach to measuring customer loyalty opens the door for the creation of performance benchmarks and for management decision making. An improved customer experience can be tracked and improved as it instills confidence in the brand through an image of trust, confidence, and fairness. Customer loyalty is created by front-line employees who interact with the customer. It must always be remembered that all companies face the challenge of meeting increasing customer expectations. Dealing with expectations in a timely and effective manner increases customer loyalty and retention.

Measuring Customer Expectations and Satisfaction

Expectations are beliefs that are measured as the (likelihood or probability that a product or service (with certain attributes, features or characteristics) will produce certain outcomes (benefits-values). As we have discussed, these expectations are based on the consumer's reservoir of affective, cognitive and behavioral experiences. Expectations are seen as related to satisfaction and can be measured in the following ways:

1) As an Importance-Value relationship that fulfills the expectations for the product/service. This can include overall expectations or expectations for the individual product attributes most important in the purchase decision or experience.

2) Affect-Satisfaction Expectations: The (liking/disliking) of the product/service or of the attributes that comprise the product/service.

3) Fulfillment of Expectations: the expected level of performance vs. the desired expectations. This is "Predictive Fulfillment" and is a respondent specific index of the performance level necessary to satisfy.

4) Expected Value from Use: Satisfaction is often determined by the frequency of use. If a product/service is not used as often as expected, the result may not be as satisfying as anticipated. For example a Harley Davidson motorcycle that is purchased with high expectations for a fun and adventurous lifestyle, but sits in the garage, an unused year subscription to the local fitness center/gym or a little used season pass to the local ski resort or amusement park may produce more dissatisfaction with the decision to purchase than with the actual product/service.

SUMMARY

Customer satisfaction is the most important element of business success and is a key measure of fulfillment of business strategies, including those involving segmentation and concept development.

In this chapter our discussion has provided an introduction to satisfaction measurement that illustrates the theoretical and methodological underpinnings of satisfaction research for business. The different satisfaction survey measures discussed were presented as components of a satisfaction survey, along with sample satisfaction survey questions. The Qualtrics survey library contains sample satisfaction questionnaires, and the Qualtrics question library contains sample individual questions. These templates will assist you in further understanding what measures should be included in your satisfaction research and why those measures are necessary and of value.

REFERENCES

Cochran, C. (2001, November). Customer satisfaction: The elusive quality. *Quality Digest*, 45–50.

Freeman, 2001 Freeman, L. (2001, September 24). Small, smaller, smallest: New analytical tools can slice markets too thin. *Marketing News, 35,* 1ff.

James, D. (2002, May 13). Better together. *Marketing News, 36,* 15–16.Kahle (1983

Mitchell, A. (1983). *The nine American lifestyles.* New York: Warner Books

Neal, W. D. (2000, June 5). When measuring loyalty satisfactorily, don't measure CS. *Marketing News, 34,* 19.

Parasuraman, A, V. Zeithaml, and L. Berry."SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality," *Journal of Retailing*, Spring 1988, pp. 12-40.

Ulwick, A. (2005) What Customers Want: Using Outcome-Driven Innovation to Create Breakthrough Products and Services, New York: McGraw Hill

Wagner, J., & Hanna, S. (1983, December). The effectiveness of family life cycle variables in consumer expenditures research. *Journal of Consumer Research*, *10*, 281–291.

Wells, W. D., & Gubar, G. (1966, November). Life cycle concept in marketing research. *Journal of Marketing Research*, *3,* 355–363.

This page is intentionally blank

Chapter 10
GENERAL CONCEPTS OF MEASUREMENT AND SCALING

Questions focus on the problem we are trying to solve, while answers are more closely associated with the measurement scale we are using to achieve our analysis.

Survey research, as a source of marketing information addresses many topics of practical interest, including concept testing for new products, corporate image measurement, ad copy evaluation, purchase intentions, customer satisfaction, and so forth. Regardless of the research topic, useful data is obtained only when the researcher exercises care in making procedural decisions such as:

1. Defining what is to be measured
2. Deciding how to make the measurements
3. Deciding how to conduct the measuring operations
4. Deciding how to analyze the resulting data

Definitions and decisions play a significant role in scientific inquiry, especially in marketing research and the behavioral sciences.

In the first section of this chapter, we focus on conceptual and operational definitions and their use in research. Increasingly, behavioral scientists are paying greater attention to defining the concepts measured in their specific disciplines, and refining operational definitions that specify how to measure and quantify the variables defining the concepts.

In the next section we discuss measurement scales and their relationship to the interpretation of statistical techniques. This section serves as useful background for the discussion of statistical techniques covered in later chapters. We then discuss the pragmatics of writing good questions.

The overall quality of a research project depends not only on the appropriateness and adequacy of its research design and sampling techniques, but also on the measurement procedures used. The third section of this chapter looks at measurement error and how we may control the reliability and validity in these measurements.

Definitions in Marketing Measurement

Marketers measure marketing program success as increased brand awareness, ad awareness, ratings of brand likeability and uniqueness, new product concept ratings and purchase intent, and customer satisfaction (Morgan, 2003). Researchers will often model these constructs.

Models are representations of reality and therefore raise the fundamental question of how well each model confidently represents reality on all significant issues. The quality of a model is judged against the criteria of *validity* and *utility*. Validity refers to a model's accuracy in describing and predicting reality, whereas utility refers to the value it adds to the making of decisions. A sales forecasting model that does not forecast sales with reasonable accuracy is probably worse than no sales forecasting model at all.

Model quality also depends on completeness and validity, two drivers of model accuracy. Managers should not expect a model to make decisions for them, but instead models should be viewed as one additional piece of information to help make decisions.

Clearly, managers will probably benefit from models that are simple enough to understand and deal with. But models used to help make multi-million-dollar decisions should be more complete than those used to make hundred-dollar decisions. The required sophistication of a model to be used depends on the model's purpose. We measure the value of a model based on its efficiency in helping us arrive at a decision. Models should be used only if they can help us arrive at results faster, with less expense, or with more validity.

Building Blocks for Measurement and Models

We cannot measure an attitude, a market share, or even sales, without first specifying how it is defined, formed, and related to other marketing variables. To better understand this, we must briefly study the building blocks of measurement theory: concepts, constructs, variables, operational definitions, and propositions.

Concepts and Constructs

A *concept* is a theoretical abstraction formed by a generalization about particulars. "Mass", "strength", and "love" are all concepts, as are "advertising effectiveness", "consumer attitude", and "price elasticity". Constructs are also concepts, but they are observable, measurable, and are defined in terms of other constructs. For example, the construct "attitude" may be defined as "a learned tendency to respond in a consistent manner with respect to a given object of orientation."

Variables

Researchers loosely call the constructs that they study *variables*. Variables are constructs in measured and quantified form. A variable can take on different values (i.e., it can vary).

Operational Definitions

We can talk about "consumer attitudes" as if we know what it means, but the term makes little sense until we define it in a specific, measurable way. An *operational definition* assigns meaning to a variable by specifying what is to be measured and how it is to be measured. It is a set of instructions defining how we are going to treat a variable. For example, the variable "height" could be operationally defined in a number of different ways, including measures in inches with a precision ruler with the person (1) wearing shoes, or (2) not wearing shoes, (3) by an altimeter or barometer, or (4) for a horse, by the number of "hands"

As another example, measuring "purchase intentions" for Brand X window cleaner might be operationally defined as the answer to the following question:

Please indicate your intention to purchase Brand X window cleaner the next time you purchase a window-cleaning product:

I Definitely Will Not	I Probably Will Not	I Probably Will	I Definitely Will Not
○	○	○	○

The researcher could have just as appropriately defined a measurement of "purchase intention" in other ways. For example, the concepts of "attitude" and "importance", which have been summed as the multiplicative product of a series of attributes like cleaning ability, fresh smell, etc. This would appear as:

P	≈	*BI*	=	Σ A_i	×	B_i
Purchase Behavior Toward		*Purchase Intention For Brand X*		*Attitudes about Brand X Window Cleaner Attributes*		*Importance of Attributes of Brand X Window Cleaner*

Propositions

A *proposition* defines the relationships between variables, and specifies both the variables influencing the relationship and the form of the relationship. It is not enough to simply state that the concept "sales" is a function of the concept "advertising", such that $S = f(Adv)$. Intervening variables must be specified, along with the relevant ranges for the effect, including where we would observe saturation effects, threshold effects, and the symbolic form of the relationship.

Integration into a Systematic Model

A proposition is quite similar to a *model*. A model is produced by linking propositions together to provide a meaningful explanation for a system or a process. When concepts, constructs, variables and propositions are integrated into a model for a research plan, we should conceptually ask the following questions:

- Are concepts and propositions specified?
- Are the concepts relevant to solving the problem at hand?
- Are the principal parts of the concept clearly defined?
- Is there consensus as to which concepts are relevant in explaining the problem?
- Are the concepts properly defined and labeled?
- Is the concept specific enough to be operationally reliable and valid?
- Do clear assumptions made in the model link the concepts?
- Are the limitations of the model stated?
- Can the model explain and predict?
- Can the model provide results for managerial decision making?
- Can the model be readily quantified?
- Are the outcomes of the model supported by common sense?

If the model does not meet the relevant criteria, it probably should be revised. Concept definitions may be made more precise; variables may be redefined, added, or deleted; operational definitions and measurements may be tested for validity; and/or mathematical forms revised.

INACCURACIES IN MEASUREMENT

Before delving into measurement scales and question types, it is helpful to remember that measurements in marketing research are rarely "exact." Inaccuracies in measurement arise from a variety of sources or factors. A portion of this variation among individual scores may represent true differences in what is being measured, while other variation may be error in measurement. For any given research project, not all will necessarily be operative, but the many possible sources causing variations in respondent scores can be categorized as follows:

- True differences in the characteristic or property
- Other relatively stable characteristics of individuals which affect scores (intelligence, extent of education, information processed)
- Transient personal factors (health, fatigue, motivation, emotional strain)
- Situational factors (rapport established, distractions that arise)

- Variations in administration of measuring instrument, such as interviewers
- Sampling of items included in the instrument
- Lack of clarity (ambiguity, complexity, interpretation of words and context)
- Mechanical factors (lack of space to record response, appearance of instrument)
- Factors in the analysis (scoring, tabulation, statistical compilation)
- Variations not otherwise accounted for (chance), such as guessing an answer

In the ideal situation, variation within a set of measurements would represent only true differences in the characteristic being measured. For instance, a company wanting to measure attitudes toward a possible new brand name and trademark would like to feel confident that measurement differences concerning the proposed names represent only the individuals' differences in this attitude. Obviously the ideal situation for conducting research seldom, if ever exists. Measurements are often affected by characteristics of individual respondents such as intelligence, education level, and personality attributes. Therefore, the results of a study will reflect not only differences among individuals in the characteristic of interest but also differences in other characteristics of the individuals. Unfortunately, this type of situation cannot be easily controlled unless the investigator knows all relevant characteristics of the population members such that control can be introduced through the sampling process.

There are many influences in a measurement other than the true characteristic of concern—that is, there are many sources of potential error in measurement. Measurement error has a constant (systematic) dimension and a random (variable) dimension. If the error is truly random, (it is just as likely to be greater than the true values as less) then the expected value of the sum of all errors for any single variable will be zero, and therefore less worrisome than nonrandom measurement error (Davis, 1997). Systematic error is present because of a flaw in the measurement instrument or the research or sampling design. Unless the flaw is corrected, there is nothing the researcher can do to get valid results after the data are collected. These two subtypes of measurement error affect the validity and reliability of measurement, topics that are discussed in the later part of this chapter. But now that we are aware of the conceptual building blocks and errors in measurement that should be considered in developing measurement scales, we will consider the types of measurement and associated questions that are commonly used in marketing research today.

MEASUREMENT CONCEPTS

Measurement can be defined as a way of assigning symbols to represent the properties of persons, objects, events, or states. These symbols should have the same relevant relationship to each other as do the things they represent. Another way of looking at this is that measurement is "the assignment of numbers to objects to represent amounts or degrees of a property possessed by all of the objects" (Torgerson, 1958, p. 19). If a characteristic, property, or behavior is to be represented by numbers, a one-to-one correspondence between the number system used and the various quantities (degrees) of that being measured must exist. There are three important characteristics or features of the real number series:

1. Order. Numbers are ordered.
2. Distance. Differences exist between the ordered numbers.
3. Origin. The series has a unique origin indicated by the number zero.

A scale of measurement allows the investigator to make comparisons of amounts and changes in the variable being measured. It is important to remember that it is the attributes or characteristics of objects we measure, not the objects themselves.

Primary Types of Scales

To many people, the term *scale* suggests such devices as a bathroom scale, pan balances, yard sticks, gasoline gauges, measuring cups, and similar instruments for finding length, weight, volume, and the like. We ordinarily tend to think about measurement in the sense of well-defined scales possessing a natural zero and constant unit of measurement. In the behavioral sciences (including marketing research), however, we must frequently settle for less-precise data. Scales can be classified into four major categories, designated as **Nominal, Ordinal, Interval, and Ratio** scales.

Each scale possesses its own set of underlying assumptions about order, distance and origin, and how well the numbers correspond with real-world entities. As our rigor in conceptualizing concepts increases, we can upgrade our measurement scale. One example is the measurement of color. We may simply categorize colors (nominal scale), or measure the frequency of light waves (ratio scale).

> *The specification of scale is extremely important in all research, because the type of measurement scale dictates the specific analytical (statistical) techniques that are most appropriate to use in analyzing the obtained data.*

Nominal Scales

Nominal scales are the least restrictive and, thus, the simplest of scales. They support only the most basic analyses. The nominal scale serves only as labels or tags to identify objects, properties or events. The nominal scale does not possess order, distance, or origin. For example, we can assign numbers to baseball players. We have a one-to-one correspondence between number and player and are careful to make sure that no players receive the same number (or that a single player is assigned two or more numbers). The classification of supermarkets into categories that "carry our brand" versus those that "do not carry our brand" is further illustration of the nominal scale.

It should be clear that nominal scales permit only rudimentary mathematical operations. We can count the stores that carry each brand in a product class and find the modal (highest number of mentions) brand carried. The usual statistical operations involving the calculations of means, standard deviations, etc. are not appropriate or meaningful for nominal scales.

Ordinal Scales

Ordinal scales are ranking scales and possess only the characteristic of order. These scales require the ability to distinguish between objects according to a single attribute and direction. For example, a respondent may be asked to rank a group of floor polish brands according to "cleaning ability". An ordinal scale results when we assign the number 1 to the highest-ranking polish, 2 to the second-highest ranking polish, and so on. Note, however, that the mere ranking of brands does not quantify the differences separating brands with regard to cleaning ability. We do not know if the difference in cleaning ability between the brands ranked 1 and 2 is larger, less than, or equal to the difference between the brands ranked 2 3. In dealing with ordinal scales, statistical description can employ positional measures such as the median, quartile, and percentile, or other summary statistics that deal with order among.

An ordinal scale possesses all the information of a nominal scale in the sense that equivalent entities receive the same rank. Also, like the nominal scale, arithmetic averaging is not meaningful for ranked data.

Interval Scales

Interval scales possess a constant unit of measurement and permit one to make meaningful statements about differences separating two objects. This type of scale possesses the properties of order and distance, but the zero point of the scale is arbitrary. Among the most common examples of interval scaling are the Fahrenheit and Centigrade scales used to measure temperature, and various types of indexes like the Consumer Price Index. While an arbitrary zero is assigned to each temperature scale, equal temperature differences are found by scaling equal volumes of expansion in the liquid used in the thermometer. Interval scales permit inferences to be made about the differences between the entities to be measured (say, warmth), but we cannot meaningfully state that any value on a specific interval scale is a multiple of another.

An example should make this point clearer. It is not empirically correct to say that 50°F is twice as hot as 25°F. Converting from Fahrenheit to Centigrade, we find that the corresponding temperatures on the centigrade scale are 10°C and –3.9°C, which are not in the ratio 2:1. We can say, however, that differences between values on different temperature scales are multiples of each other. That is, the difference of 50°F–0°F is twice the difference of 25°–0°F. The corresponding differences on the Centigrade scale are 10°C – (–17.7°C) = 27.7°C and –3.9°C – (–17.7°C) = 13.8°C are in the same 2:1 ratio.

Interval scales are unique up to a transformation of the form $y = a + bx$; $b > 0$. This means that interval scales can be transformed from one to another by adding or multiplying a constant. For example, we can convert from a Fahrenheit to Celsius using the formula:

$$T_C = 5/9\ (T_F - 32)$$

Most ordinary statistical measures (such as arithmetic mean, standard deviation, and correlation coefficient) require only interval scales for their computation.

Ratio Scales

Ratio scales represent the elite of scales and contain all the information of lower-order scales and more besides. These are scales like length and weight that possess a unique zero point, in addition to equal intervals. All types of statistical operations can be performed on ratio scales.

An example of ratio-scale properties is that 3 yards is three times 1 yard. If transformed to feet, then 9 feet and 3 feet are in the same 3:1 ratio. It is easy to move from one scale to another merely by applying an appropriate positive multiplicative constant; this is the practice followed when changing from grams to pounds or from feet to inches.

Relationships Among Scales

To provide some idea of the relationships among nominal, ordinal, interval, and ratio scales, the marketing researcher who uses descriptive statistics (arithmetic mean, standard deviation) and tests of significance (*t*-test, F-test) should require that the data are (at least) interval-scaled .

From a purely mathematical point of view, you can obviously do arithmetic with any set of numbers—and any scale. What is at issue here is the interpretation and meaningfulness of the results. As we select more powerful measurement scales, our abilities to predict, explain, and otherwise understand respondent ratings also increase.

Table 7.1 Scales of Measurement

Scale	*Mathematical Group Structure*	*Permissible Statistics*	*Typical Elements*
Nominal	Permutation group y = *f*(*x*), where *f*(*x*) means any one-to-one correspondence	Mode Contingency Coefficient	Numbering of football players Assignment of type or model numbers to classes
Ordinal	Isotonic group y = *f*(*x*), where *f*(*x*) means any strictly increasing function	Median Percentile Order correlation Sign test; run test	Hardness of minerals Quality of leather, lumber, wool, etc. Pleasantness of odors
Interval	General linear group $y = a+bx$ $b > 0$	Mean Average deviation Standard deviation Product-moment correlation *t*-test, F-test	Temperature (Fahrenheit and centigrade) Energy Calendar dates
Ratio	Similarity group $y = cx$ $c > 0$	Geometric mean Harmonic mean Coefficient of variation	Length, width, density, resistance Pitch scale, loudness scale

(Stevens 1946, p. 678)

Basic Question and Answer Formats

Underlying every question is a basic reason for asking it. This reason reflects the construct to be measured, the problem to be solved or hypothesis to be tested. Constructing a question that reflects this reason will result in a higher probability that the desired response will be obtained. Table 7.2 shows nine different types of questions (based on the nature of content), the broad reason underlying asking each type of question, and some examples of each type.

Table 7.2 Basic Question Types

Type of Question	*Goal of Question*	*Positioning of Question*
Factual or behavioral	To get information.	Questions beginning with what, where, when, why, who and how.
Explanatory	To get additional information or to broaden discussion.	How would that help? How would you go about doing that? What other things should be considered?
Attitudinal	To get perceptions, motivations, feelings, etc., about an object or topic.	What do you believe to be the best? How strongly do you feel about XYZ?
Justifying	To get proof to challenge old ideas and to get new ones.	How do you know? What makes you say that?
Leading	To introduce a thought of your own.	Would this be a possible solution? What do you think of this plan?
Hypothetical	To use assumptions or suppositions.	What would happen if we did it this way? If it came in blue would you buy it today?
Alternative	To get a decision or agreement.	Which of these plans do you think is best? Is one or two o'clock best for you?
Coordinative	To develop common agreement. To take action.	Do we all agree that this is our next step?
Comparative	To compare alternatives or to get a judgment anchored by another item.	Is baseball more or less exciting to watch on TV than soccer?

Based on this structure, and the information in Table 7.3, which deals with standard answer formats, we are able to distinguish four basic question/answer types:

1. Free-answer (open-ended text)
2. Choice answers: dichotomous, single choice and multiple choice (select *k* of *n*)
3. Rank order answers
4. Constant sum answers

Table 7.3 Standard Answer Formats Based on Task

Measurement Scale*	Format Type	Description
N,O,I	Select 1/n—pick-1:	The respondent is given a list of *n* options and is required to choose one option only.
N,O	Select k/n—pick-k:	The respondent gets a set of *n* options to select from but this time chooses up to *k* options (*k* = *n*).
N,O	Select k1/k2/n—pick-and-pick:	The respondent is asked to select *k*1 options in Category 1 and *k*2 options in Category 2. Each option can be selected in only one of the two categories.
N, I	Sort and rank:	The respondent picks k items and allocates them into L buckets, then items allocated to each bucket are assigned ranks.
N,O	Rank k/n—rank:	In this question the respondent gets *n* options and is asked to rank the top *k* (*k* = *n*).
N,O	Select k1/n and Rank k2/k1—pick and rank:	This question type is similar to pick-*k*, but in addition to selecting *k*1 options from a list of *n* options, the respondent is then asked to rank some fraction, *k*2/k1 of those selected.
O,I	Integer Rating:	The respondent is asked to rate on a linear scale of 1 to *n* the description on the screen or accompanying prop card (for example, 1 for completely disagree to 5 for completely agree). Only integer responses are accepted.
O,I	Continuous Rating:	This is similar to integer rating, except that the response can be any number (not necessarily an integer number) within the range (for example, 5.2 on a scale of 0 to 10).
R	Constant Sum:	The respondent is provided with a set of attributes (5, 10, etc.) and is asked to distribute a total of *p* points across those attributes.
N,O	Yes/No:	This question entails a yes/no answer and is of course, a Select 1/2---pick-1 question type.
I	Integer—integer-#:	The respondent is asked for a fact that can be expressed in integer number form. A valid range can be provided for error checking. Example: Age.
I,R	Real—real-#:	Similar to integer-# except that the answer expected is in the form of a real (not necessarily an integer) number. Example: Income. A valid range can be provided for error checking.
C	Character:	The respondent types in a string of characters as a response. Example: Name.
I	Multiple Integer Ratings:	This question type is identical to integer-scale except that multiple questions (classified as "options") can appear on a single screen. Each question is answered and recorded separately.
I,R	Multiple Real Number Ratings:	This question type is identical to real-scale except that multiple questions (classified as "options") can appear on a single screen. Each question is answered and recorded separately.

*Legend: N = Nominal, O = Ordinal, I = Interval, R = Ratio, C = Alpha-Numeric Text Characters

Free Answer or Open-Ended Text Questions/Answers

The free answer (or open-ended text question) has no fixed alternatives to which the answer must conform. The respondent answers in his or her own words and at the length he or she chooses, subject of course to any limitations imposed by the questionnaire itself. Interviewers are usually instructed to make a verbatim record of the answer.

While free-answer questions are usually shorter and less complex than multiple-choice and dichotomous questions, they place greater demands on the ability of the respondents to express themselves. As such, this form of question provides the opportunity for greater ambiguity in interpreting answers. To illustrate, consider the following verbatim transcript of one female respondent's reply to the question:

> ***What suggestions could you make for improving tomato juice?***
> "I really don't know. I never thought much about it. I suppose that it would be nice if you could buy it in bottles because the can turns black where you pour the juice out after it has been opened a day or two. Bottles break, though."

Did she have "no suggestion", "suggest packaging in a glass container", or "suggest that some way be found to prevent the can from turning black around the opening"?

One way to overcome some of these problems, at least in personal and telephone surveys is to have interviewers probe respondents for clarity (rather than additional information). One practitioner has gone so far as to suggest that questionnaires should clearly instruct interviewers to probe only once for additional information, and to continue to probe for clarity until the interviewer understands a respondent's reply.

Compared with other question forms (see Exhibit 7.1), we may tentatively conclude that the free-answer question provides the lowest probability of the questions being ambiguous, but the highest probability of the answers being ambiguous,.

Exhibit 7.1 Open-Ended Questions and Answers

The advantages of the open-ended format are considerable, but so are its disadvantages (Sudman and Bradburn, 1982). In the hands of a good interviewer, the open format allows and encourages respondents to give their opinions fully and with as much nuance as they are capable of. It also allows respondents to make distinctions that are not usually possible with the fixed alternative formats, and to express themselves in language that is comfortable for them and congenial to their views. In many instances it produces vignettes of considerable richness and quotable material that will enliven research reports.

The richness of the material can be a disadvantage if there is need to summarize the data into simple response categories. Coding of free-response material is known as *content analysis* and is not only time consuming and costly, but also introduces some amount of coding error.

Open-ended questions also take somewhat more time and psychological work to answer than closed questions. They also require greater interviewer skill to recognize ambiguities of response and to probe and draw respondents out, particularly those who are reticent and not highly verbal, to make sure that they give answers that can be coded. Open-ended response formats may work better with telephone interviews, where a close supervision of interview quality can be maintained, although there is a tendency for shorter answers to be given on the telephone. No matter how well controlled the interviewers may be, however, factors such as carelessness and verbal facility will generate greater individual variance among respondents than would be the case with fixed alternative response formats.

Dichotomous and Multiple-Choice Answers

The select *k* of *n* format is the workhorse of survey building, and provides the general form for both dichotomous and multiple-choice answer types. Three general forms of questions are frequently used:

Select Exactly 1 of n Answers:

When selecting *k* = 1/n, the type of answer scale is dependent on *n*, the number of answers. A dichotomous question has two fixed answer alternatives of the type “Yes/No”, “In favor/Not in favor”, “Use/Do not use”, and so on. The question quoted earlier, *“Do you like the taste of tomato juice?”* is an example of a dichotomous question. Multiple-choice questions are simply an extension of the dichotomous question that have more answer points and often take the form of an ordered or interval measurement scale.

Traditional multiple–choice answers also are of the select *1* of *n* answer form, but have more than two available answers. For example, an agreement scale could have three, five, or seven available answers.

Three answers:	Agree/Neutral/Disagree
Five answers:	Strongly Agree/Agree/Neither/Disagree/Strongly Disagree
Seven answers:	Very Strongly Agree/Strongly Agree/Agree/Neither Agree nor Disagree/ Disagree/ Strongly Disagree/Very Strongly Disagree

As with all select *1* of *n* answers, the specific text associated with the answer options is variable and could measure many different constructs such as affect (liking), satisfaction, loyalty, purchase likelihood, and so forth.

Select Exactly k of n Answers

When questions are developed that accept or require multiple responses within a set of answers, the form “exactly *k* of *n*” or “as many as *k* of *n*” can be used. This general form asks the respondent to indicate that several answers meet the requirements of the question. In this case, the data collected would be of type categorical or even loosely ordered if presence or absence of a characteristic is being measured (data is coded as 0 if not selected, and 1 if selected). This type of question might be

Please identify the three (3) service activities that are most likely to be outsourced in the next 12 months.

☐ Retirement benefits	☐ Training, education
☐ Recruitment	☐ Management/Executive selection
☐ Medical benefits	☐ Travel services
☐ Security services	☐ Organization development
☐ Health services/medical	☐ Work/Life programs

Select as Many as k of n Answers

A variable number of answers may also be appropriate, particularly where long lists of attributes or features are given. In these cases, the respondent is asked to select as many as *k* of the *n* possible answers, where *k* can be any number from 2 to *n*. For example, in the previous question, the respondent could select as many as three (one, two, or three) of 10 possible answers. The question might be reworded to read something like . . .

Please identify which service activities are most likely to be outsourced in the next 12 months (check all that apply).

Whitlark and Smith (2004) show an application of pick *k of n* data that asks respondents to pick a small number of attributes that they feel best describe a brand from a list of 10, 20, or even 30 attributes. Collecting the pick data is much faster than asking a respondent to rate brands with respect to a long list of attributes. In an online survey environment, respondents can quickly scan down columns or across screens and quickly complete the pick data task for a familiar brand, thereby saving time and reducing respondent fatigue and dropout rates.

Having people describe a brand by picking attributes from a list is a quick and simple way to assess brand performance and positioning. Whitlark and Smith (2004) show that when respondents are asked to pick from one third to one half of the viewed items, the pick *k of n* data can be superior to scaled data in terms of reliability and power to discriminate between attributes.

Rank-Order Questions/Answers

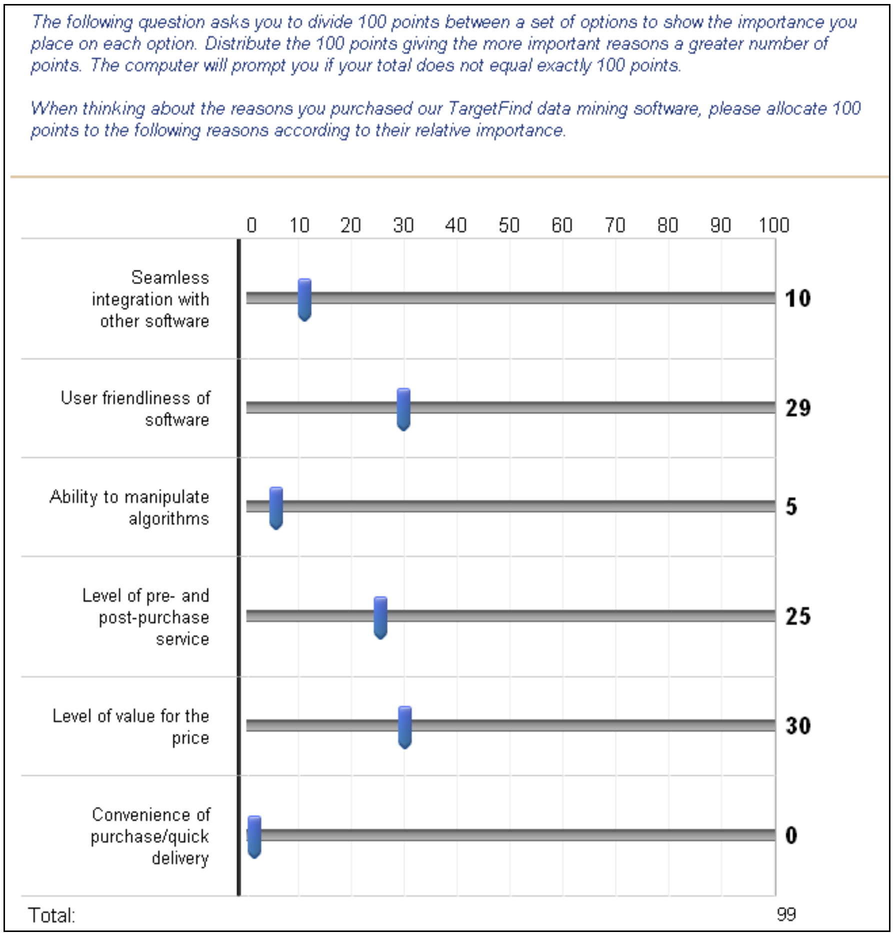
The next level of measurement rank-orders the answers and thereby increases the power of the measurement scale over categorical measurement by including the characteristic of order in the data. Whereas the categorical data associated with many dichotomous or multiple-choice items does not permit us to say that one item is greater than another, rank-order data allows for the analysis of differences. Rank-order questions use an answer format that requires the respondent to assign a rank position to all items, or a subset of items in the answer list. The first, second, and so forth up to the *n*th item would be ordered. Procedures for assigning position numbers can be very versatile, resulting in different types of questions that can be asked. Typical questions might include identifying preference rankings, or attribute associations from first to last, most recent to least recent or relative position (most, next most, and so forth, until either a set number of items is ordered or all items may be ordered).

When this type of question is administered online or using a CATI (Computer Aided Telephone Interviewing) system, additional options for administration may include randomization and acceptance/validation of ties in the ranking. Randomization of the answer list order helps to control for presentation order bias. It is well established that in elections, being the first in a ballot candidate list increases chances of receiving the voter's election.

Tied rankings are another issue to be considered for rank-order questions. When ties are permitted, several items may be evaluated as having the same rank. In general, this is not a good idea because it weakens the data. However, if ties truly exist, then the ranking should reflect this. Rank-order questions are generally a difficult type of question for respondents to answer, especially if the number of items to be ranked goes beyond five or seven.

Constant Sum Questions/Answers

A constant sum question is a powerful question type that permits collection of ratio data, meaning that the data is able to express the relative value or importance of the options (option A is twice as important as option B). This type of question is used when you are relatively sure of the answer set (i.e., reasons for purchase, or you want to evaluate a limited number of reasons that you believe are important). The following example of a constant sum question from Qualtrics, uses sliding scales to select a sum of 100 points:



ADVANCED MEASUREMENT AND SCALING CONCEPTS

Continuing our discussion of scales, we now focus on some of the more common scaling techniques and models. We focus on broad concepts of attitude scaling—the study of scaling for the measurement of managerial and consumer or buyer perception, preference, and motivation. All attitude (and other psychological) measurement procedures are concerned with having people—consumers, purchasing agents, marketing managers, or whomever—respond about certain stimuli according to specified sets of instructions. The stimuli may be alternative products or services, advertising copy themes, package designs, brand names, sales presentations, and so on. The response may involve judging which copy theme is more pleasing than another, which package design is more appealing than another, what mental images do new brand names evoke, which adjectives best describe each salesperson, and so on.

Scaling procedures can be classified in terms of the measurement properties of the final scale (nominal, ordinal, interval, or ratio), the task that the subject is asked to perform, or in still other ways, such as whether the scale measures the subject, the stimuli, or both (Torgerson, 1958).

We begin with a discussion of various methods for collecting ordinal-scaled data (paired comparisons, rankings, ratings, etc.) in terms of their mechanics and assumptions regarding their scale properties. Then specific procedures for developing these actual scales are discussed. Techniques such as Thurstone Case V scaling, semantic differential, the Likert summated scale, and the Thurstone differential scale are illustrated. The chapter concludes with some issues and limitations of scaling.

Advanced Ordinal Measurement Methods

The variety of ordinal measurement methods includes a number of techniques:

- Paired comparisons
- Ranking procedures
- Ordered-category sorting
- Rating techniques

We discuss each of these data collection procedures in turn.

Paired Comparisons

As the name suggests, paired comparisons require the respondent to choose one of a pair of stimuli that "has more of", "dominates", "precedes", "wins over", or "exceeds" the other with respect to some designated property of interest. If, for example, six laundry detergent brands are to be compared for "sudsiness", a full set of paired comparisons would involve (n x n – 1)/2 = (6 × 5) / 2, or 15, paired comparisons (if order of presentation is not considered). Respondents are asked which one of each pair has the most sudsiness.

A sample question format for paired comparisons is shown in Table 7.4. The order of presentation of the pairs and which item of a pair is shown first are typically determined and/or presented randomly. Consider the following hypothetical brand names (and numerical categories): Arrow (1), Zip (2), Dept (3), Advance (4), Crown (5), and Mountain (6).

Table 7.4 Example of the Paired Comparisons Question

For each of the pairs of laundry detergent brands shown below, indicate which one has the most sudsiness:

I prefer the left Brand		I prefer the right brand
	Arrow ----- Zip	
○		○
	Arrow ----- Advance	
○		○
	Dept ----- Arrow	
○		○
	Crown ----- Arrow	
○		○
	Crown ----- Mountain	
○		○

Original Data

		Brand					
		Arrow	**Zip**	**Advance**	**Dept**	**Crown**	**Mountain**
Brand	**Arrow**	X	0	1	1	1	1
	Zip	1	X	1	1	1	1
	Advance	0	0	X	0	0	0
	Dept	0	0	1	X	0	0
	Crown	0	0	1	1	X	1
	Mountain	0	0	1	1	0	X

Permutated Rows and Columns

		Brand						
		Zip	**Arrow**	**Crown**	**Mountain**	**Dept**	**Advance**	**Total**
Brand	**Zip**	X	1	1	1	1	1	5
	Arrow	0	X	1	1	1	1	4
	Crown	0	0	X	1	1	1	3
	Mountain	0	0	0	X	1	1	2
	Dept	0	0	0	0	X	1	1
	Advance	0	0	0	0	0	X	0

A cell value of 1 implies that a row brand exceeds the column brand, "0" otherwise.

Ranking Procedures

Ranking procedures require the respondent to order stimuli with respect to some designated property of interest. For example, instead of using the paired-comparison technique, respondent might have been asked to directly rank the detergents with respect to sudsiness. Similarly, ranking can be used to determine key attributes for services.

In a survey conducted by Subaru of America, new Subaru car purchasers were asked questions regarding the purchase and delivery processes. One question required ranking:

From the following list, please choose the three most important factors (other than price or deal) which attracted you to shop at this Subaru dealership. Please rank these three factors in order of importance to you by writing the number 1 in the box which was most important, followed by numbers 2 and 3 in the appropriate boxes. Rank three boxes only.

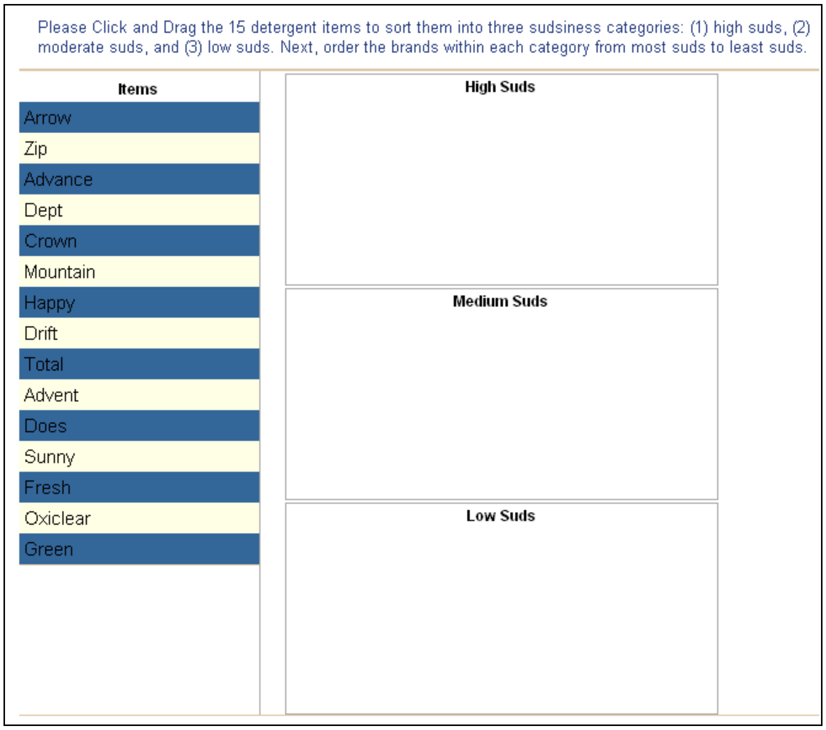
Location	1
Previous experience	2
Experiences of others	3
Dealer's reputation	4
Had specific model you wanted	5
Financing	6
Advertising reputation	7
Service	8

A variety of ordering methods may be used to order *k* items from a full set of *n* items. These procedures, denoted by Coombs (1964) as "order *k/n*" (*k* out of *n*), expand the repertory of ordering methods quite markedly. The various ordering methods may pre-specify the value of *k* ("order the top three out of six brands with respect to sudsiness") as illustrated by the Subaru study, or allow *k* to be chosen by the respondent ("select those of the six brands that seem to exhibit the most sudsiness, and rank them").

When the groups can be ordered by some category, a procedure known as category sorting can be used. For example, if it is desired that a respondent rank all items in a longer list of items, the pick-group-rank procedure may be used. The respondent first sorts the items into a number of ordered categories or piles (each of which has a relatively small, equal number of items). Then, the request is made to rank the items within each pile. This task can be completed in personal interviews or in online surveys using something like the Qualtrics pick-group-rank question that facilitates the ordered ranking within the ordered-category sorting tasks.

Ordered-Category Sorting

Pick-Group-Rank is one of a variety of data collection procedures that have as their purpose the assignment of a set of stimuli to a set of ordered categories. For example, if 15 varieties of laundry detergents represented the stimulus set, the respondent might be asked to complete the following task:



The pick-group-rank could be used with ordered categories to sort all of a large list of items, where there is:

1. free versus forced assignment of names to grouping categories
2. free verses forced assignment of stimuli to grouping categories
3. the assumption of equal intervals between category boundaries versus the weaker assumption of category boundaries that are merely ordered with regard to the attribute of interest

In ordinal measurement methods one assumes only an ordering of category boundaries. The assumption of equal intervals separating boundaries is part of the interval/ratio measurement set of methods. Ordered-category sorting appears especially useful when the researcher is dealing with a relatively large number of stimuli (over 15 or so) and it is believed that a subject's discrimination abilities do not justify a strict (no ties allowed) ranking of the stimulus objects.

Rating Techniques

Rating scales are ambiguous as to whether or not they meet the criterion of equal intervals separating the category boundaries. In some cases, the scaled responses are considered by the researcher to be only ordinal, while in other cases, the researcher treats them as interval-or ratio-scaled. The flexibility of rating procedures makes them appropriate for either the ordinal or interval/ratio measurement data collection methods (depending on the nature of the scale values).

The rating task typically involves having a respondent place that which is being rated (a person, object, or concept) along a continuum or in one of an ordered set of categories. Ratings allow the respondent to designate a degree or an amount of a characteristic or attribute as a point on a scale. The task of rating is one of the most popular and easily applied data collection methods, and is used in a variety of scaling approaches, such as the semantic differential and the Likert summated scale.

Rating scales can be either monadic or comparative. In monadic scaling, each object is measured (rated) by itself, independently of any other objects being rated. In contrast, comparative scaling objects are evaluated in comparison with other objects. For example, a recent in-flight survey conducted by United Airlines asked the following questions:

Monadic Example:
Please rate the service you received from the United reservations agent

	Poor Among the worst	Fair Not as Good as Most	Good About the Same as Most	Very Good Better than Most	Excellent Among the Best
Courtesy/Friendliness	○	○	○	○	○
Knowledge/helpfulness	○	○	○	○	○
Efficiency on completing transaction	○	○	○	○	○

The rating is monadic. United then asked respondents another question:

Comparative Example:
Please rate today's fight attendants compared to flight attendants on other airlines on each of the following items.

	Poor Among the worst	Fair Not as Good as Most	Good About the Same as Most	Very Good Better than Most	Excellent Among the Best
Courtesy/Friendliness	○	○	○	○	○
Assistance in cabin before departure	○	○	○	○	○
Responsiveness to your needs	○	○	○	○	○

Ratings are used very widely because they are easier and faster to administer and yield data that are amenable to being analyzed as if they are interval-scaled. But there is a risk of lack of differentiation among the scores when the particular attributes are worded positively or are positive constructs, such as values, and the respondents end-pile their ratings toward the positive end of the scale. Such lack of differentiation may potentially reduce the variance of the items being rated and reduce the ability to detect relationships with other variables.

McCarty and Shrum (2000) offer an alternative to simple rating. Respondents first picked their most and least important values (or attributes or factors), and then rated them. The remaining values were then rated. Their results indicate that, compared with a simple rating of values, the most-least procedure reduces the level of end-piling and increases the differentiation of values ratings, both in terms of dispersion and the number of different rating points used.

Respondents may have trouble choosing a rating scale value on the high end, but not so at the low end. One person's rating of 9 or 10 may be equal in meaning to another's 7 or 8. Semon (1999) suggests that one way to find the real difference in perception or attitude is to ask each respondent three questions at the start of an interview:

1. On this scale, how do you rate the brand you now use or that you know best?
2. How do you rate the best brand you know about?
3. What rating represents the minimum acceptable level?

Questions such as these are often asked in product and brand studies, to interpret ratings and provide anchor points for a respondent's ratings. A respondent's actual ratings can be translated into responses relative to one or more of these anchors to produce real-meaning relative ratings that can be reliably aggregated and analyzed without depending upon assumptions that may be questionable.

Rating methods can take several forms, numerical, graphic and verbal. Often two or more of these formats appear together, as illustrated in Figures 7.1 and 7.2. Many other types of rating methods are in use (Haley & Case, 1979).

Figure 7.1 Examples of Rating Scales Used in Marketing Research

a.

I Definitely Disagree	I Generally Disagree	I Moderately Disagree	I Moderately Agree	I Generally Agree	I Definitely Agree
○	○	○	○	○	○

b.

Below Average	About Average	Slightly Better Than Average	Considerably Better Than Average	One of the Best Available	Definitely the Best
○	○	○	○	○	○

c.

Definitely Would Not Buy 1	2	3	May or May Not Buy 4	5	6	Definitely Would Buy 7
○	○	○	○	○	○	○

d.

	Not at All Important	Somewhat Unimportant	Neither Important nor Unimportant	Somewhat Important	Very Important
Cleanliness of the store	○	○	○	○	○
Location	○	○	○	○	○

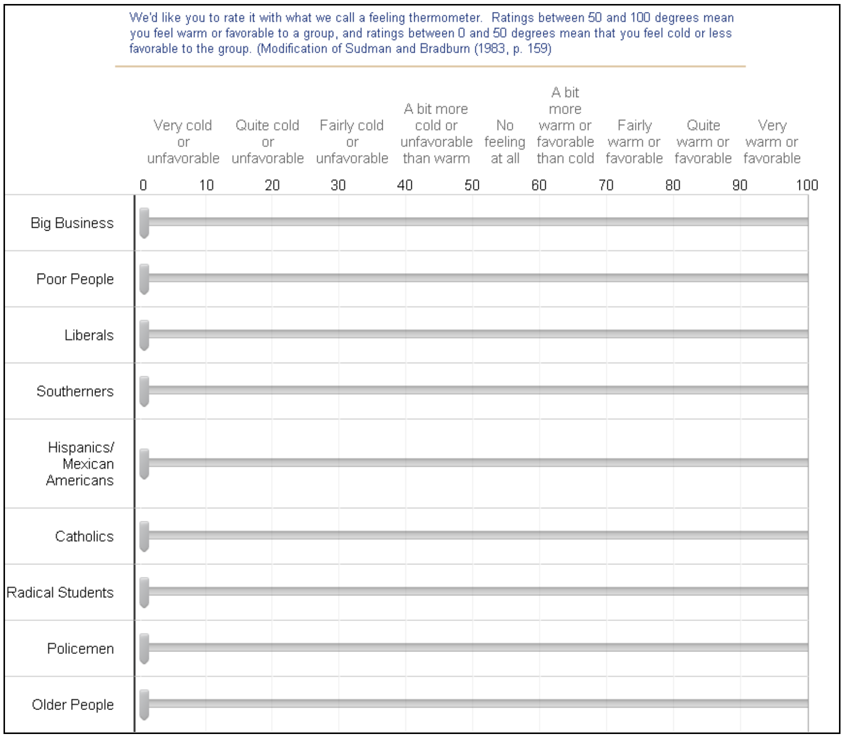
e.

○	1 Very Harsh
○	2 Somewhat Harsh
○	3 Slightly Harsh
○	4 Neither Gentile Nor Harsh
○	5 Slightly Gentle
○	6 Somewhat Gentle
○	7 Very Gentle

In many instances where rating scales are used, the researcher assumes not only that the items are capable of being ranked, but also that the descriptive levels of progress are in equal-interval steps psychologically. That is, the numerical correspondences shown in Panels (c) and (e) of Figure 7.1 may be treated—sometimes erroneously—as interval- or ratio-scaled data. Even in cases represented by Panels (a), (b), and (d), it is not unusual to find that the researcher assigns successive integer values to the various category descriptions and subsequently works with the data as though the responses *were* interval-scaled.

Treating rating scales as interval or ratio measurements is a practice that is well documented and widespread. Research shows that there is little error in treating the data as being of a higher level of measurement than it is. Research evidence supports this practice, in that often when ordinal data are treated as interval and parametric analysis are used, the conclusions reached are the same as when the data are treated as ordinal and tested using non-parametric analyses.

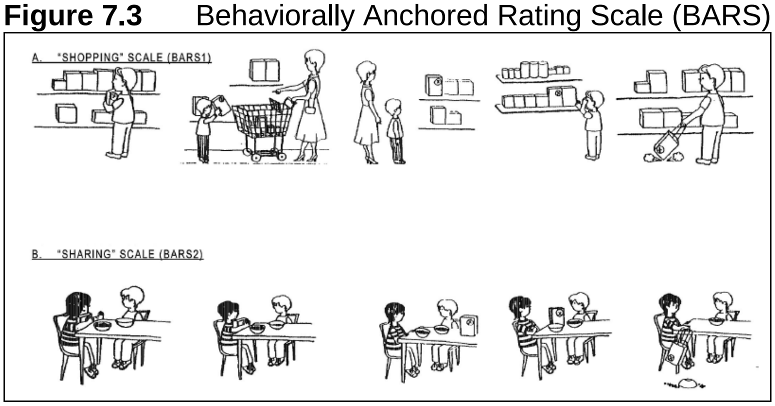
Figure 7.2 A Rating Thermometer



Constructing a Behaviorally Anchored Rating Scale

One type of itemized rating scale that has merit in cases where leniency error (lack of discrimination) may be troublesome is the behaviorally-anchored rating scale, or BARS (see Figure 7.3). This scale uses behavioral incidents to define each position on the rating scale rather than verbal, graphic, or numeric labels. Thus, providing specific behavioral anchors can reduce leniency errors and increase discrimination. Developing scales such as these requires a great amount of testing and refinement to find the right anchors for the situation under examination.

Figure 7.3 Behaviorally Anchored Rating Scale (BARS)



The basic process of developing a behaviorally anchored rating scale consists of four steps:

1. Construct definition—the construct being measured must be explicitly defined and the key dimensions identified

2. Item generation—statements must be generated describing actual behaviors that would illustrate specific levels of the construct for each dimension identified
3. Item testing—to unambiguously fit behavioral statements to dimensions
4. Scale construction—lay out the scale with behavioral statements as anchors

In following this process, sets of judges are used. It should be clear that developing BARS is a time-consuming and costly task. Thus, they should be reserved for those applied settings where they can minimize the errors they are designed to curtail, especially leniency error. As an example, families with elderly members were surveyed to determine their need for in home health-care services. A BARS was used for one critical measure of how well elderly members of the household were able to perform everyday living activities:

Now about your ability to perform everyday living activities. Which of the following best describes your everyday living capacities:

- ○ You can perform all physical activities of daily living without assistance. (Excellent capacity)
- ○ You can perform all physical activities without assistance, but may need some help with the heavy work (such as laundry and housekeeping). (Good capacity)
- ○ You regularly require help with certain physical activities and/or heavy work, but can get through any single day without help. (Moderate capacity)
- ○ You need help each day, but not necessarily throughout the day or night. (Severely impaired capacity)
- ○ You need help throughout the day and/or night to carry out the activities of daily living. (Completely impaired capacity)

Table 7.5 identifies nine questions that must be answered when a scale is constructed.

Table 7.5 Issues in Constructing a Rating Scale

1. Should negative numbers be used?
2. How many categories should be included?
3. Related to the number of categories is: Should there be an odd number or an even number? That is, should a neutral alternative be provided?
4. Should the scale be balanced or unbalanced?
5. Is it desirable to not force a substantive response by giving an opportunity to indicate “don’t know,” “no opinion,” or something similar?
6. What does one do about halo effects—that is, the tendency of raters to ascribe favorable property levels to all attributes of a stimulus object if they happen to like a particular object in general?
7. How does one examine raters’ biases—for example, the tendency to use extreme values or, perhaps, only the middle range of the response scale, or to overestimate the desirable features of the things they like (i.e., the generosity error)?
8. How should descriptive adjectives for rating categories be selected?
9. How anchoring phrases for the scale’s origin should be chosen?

Some research on these questions has been conducted showing that errors are not made when a neutral option is provided. Our suggestion is that it always be included, unless the researcher has a compelling reason to not do so (e.g., the problem situation/sample mix is such that each sample member can be expected to have a non-neutral attitude). Expected voting in a survey of voters is an example.

Question 4 deals with the interesting issue of *Balance*, referring to having an equal number of negative response alternatives as positive ones. When using importance scales for attributes, the alternatives provided may be "very important", "important", "neither important nor unimportant", "unimportant", and "very unimportant", or additional categories may be included. Thomas Semon (2001) has questioned the use of balance (or symmetry, as he calls it) in importance scales. He argues that importance is not a bipolar concept. Importance ranges from some positive amount to none, not a negative amount. Although this appears to have conceptual appeal, researchers continue to use successfully use importance scales from some mid-point—specified or implied. There would seem to be three keys to successful importance scale use:

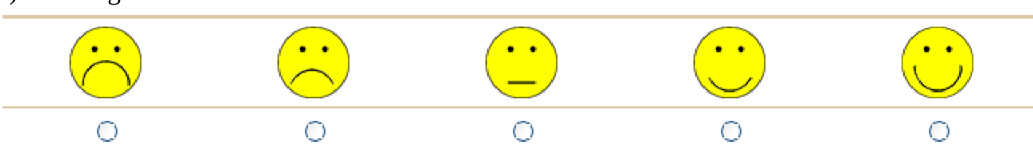
1. Isolating any findings of unimportance
2. Recognizing that importance is ordinally scaled
3. Accurately interpreting the relative nature of importance findings

Answers to questions such as these will vary by the researcher's approach, and by the problem being studied. The effects of research design on reliability and validity of rating scales are discussed in two excellent review papers (Churchill and Peter, 1984; Peter and Churchill, 1986).

Exhibit 7.4 Measuring Preferences of Young Children Calls for Creativity

The children's market is a multi-billion dollar market in direct purchasing power and an even greater market in purchasing influence. Among the areas of most concern are better scaling techniques for measuring children's product preferences. Widely used approaches for assessing children's preferences are itemized rating scales using a series of stars (a scale from 1 to 5 stars) or a series of facial expressions (a scale anchored at one end with a happy face and at the other end with a sad face), as illustrated below:

(a) Smiling Faces Scale



(b) Star Scale

★★★★★ ★★★★ ★★★ ★★ ★

Children are asked to indicate how much they like a product, or how much they like a particular feature of a product, by pointing to one of the visual anchors on the scale.

Although these scales have done well in varied research applications, there is often a problem with leniency that emerges, particularly when used with young children under the age of eight. This error emerges when young children consistently use the extreme positions (usually on the positive side) with relatively little use of intermediate scale positions. If this is done for all products tested, the overall sensitivity of existing (traditional) rating scales is lowered, resulting in inconclusive findings about children's preferences.

In summary, rating methods—depending on the assumptions of the researcher—can be considered to lead to ordinal-, interval-, or even ratio-scaled responses. The latter two scales are taken up next. We shall see that rating methods figure prominently in the development of quantitative-judgment scales.

Interval/Ratio Procedures

Direct-judgment estimates, fractionation, constant sum and rating methods assume more than ordinal properties about respondents' judgments) are all variants of interval/ratio procedures or metric measurement methods**.**

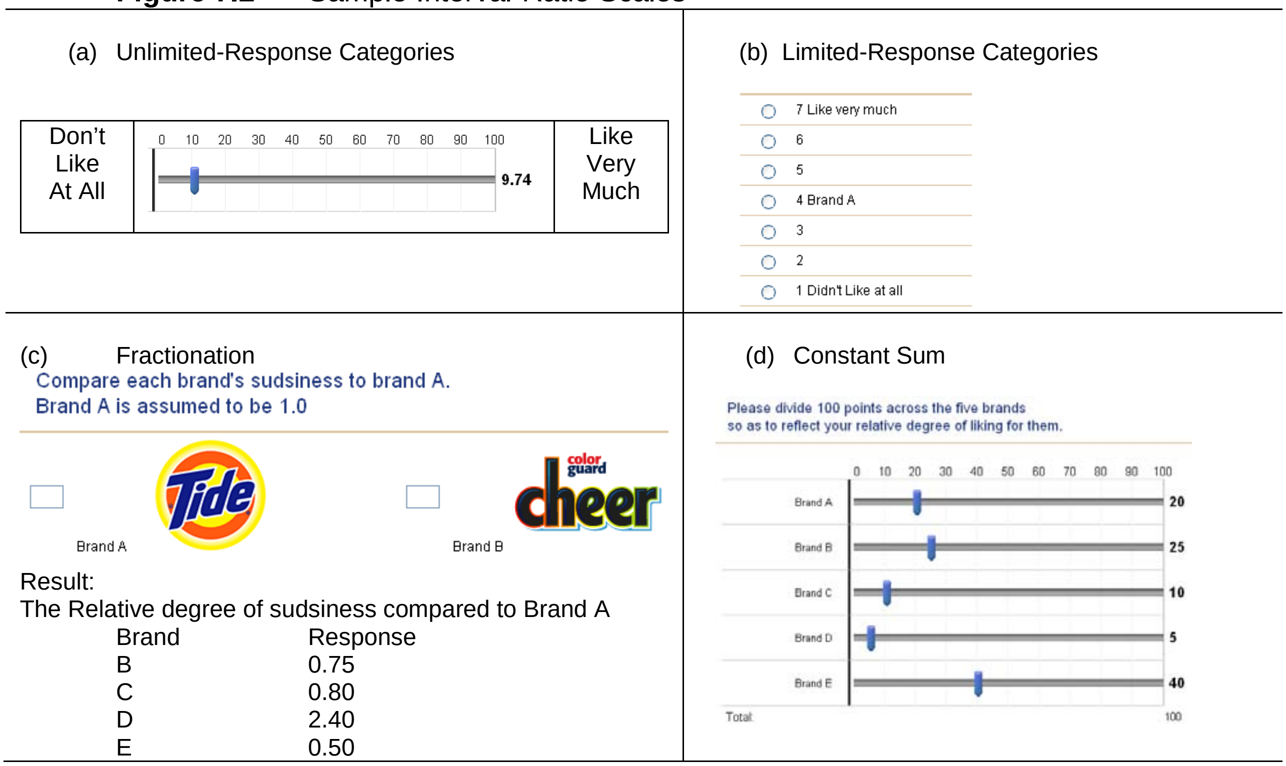
Direct-Judgment Methods

In direct-judgment methods, the respondent is asked to give a numerical rating to each stimulus with respect to some designated attribute. In the case of continuous rating scales, the respondent is free to choose his or her own number along some line that represents his or her judgment about the magnitude of the stimulus relative to some reference points (Figure 7.4).

These continuous scales work effectively in a semantic differential context (to be discussed later in this chapter) (Albaum, Best, & Hawkins, 1981), and appear to be insensitive to fluctuations in the length of the line used (Hubbard, Little, & Allen, 1989).

The limited-response category sub case, illustrated by Panel (b) in Figure 7.4, is nothing more than a straight rating procedure, with the important addition that the ratings are now treated as either interval- or ratio-scaled data (depending on the application).

Figure 7.2 Sample Interval-Ratio Scales



Fractionation

Fractionation is a procedure in which the respondent is given two stimuli at a time (e.g., a standard laundry detergent and a test brand) and asked to give some numerical estimate of the ratio between them, with respect to some attribute, such as sudsiness. The respondent may answer that the test brand, in his or her judgment, is three-fourths as sudsy as the standard. After this is done, a new test brand is compared with the same standard, and so on, until all test items are judged. Panel (c) in Figure 7.2 illustrates this procedure.

In other cases, the test item can be more or less continuously varied by the respondent. For example, in an actual test of the attribute is sweetness of lemonade, the respondent may be asked to add more sweetener until the test item is “twice as sweet” as the standard.

Constant Sum

Constant-sum methods have become quite popular in marketing research, primarily because of their simplicity and ease of instructions. In constant-sum methods the respondent is given some number of points—typically 10 or 100—and asked to distribute them over some set of stimuli or attribute alternatives in a way that reflects their relative importance or magnitude (Figure 7.2d). Constant sum forces the respondent to make comparative evaluations across the stimuli, and effectively standardizes each scale across persons, since all scores must add to the same constant. Generally, it is assumed that a subjective ratio scale is obtained by this method.

In summary, unlike ordinal measurement methods, the major assumption underlying ratio/interval measurement methods) is that a unit of measurement can be constructed directly from respondents’ estimates about scale values associated with a set of stimuli. The respondent’s report is taken at face value and any variation in repeated estimates (over test occasions within respondent or over respondents) is treated as error; repeated estimates are usually averaged over persons and/or occasions. The problems associated with Interval-Ratio Scaling methods include the following:

1. Respondents’ subjective scale units may differ across each other, across testing occasions, or both.
2. Respondents’ subjective origins (zero points) may differ across each other, across occasions, or both.
3. Unit and origin may shift over stimulus items within a single occasion.
4. Subjective distance between stimuli may not equal one’s perception of the distance on the scale.

These problems should not be treated lightly, but considered in the design of the question and scale points.

Most ratings measurement methods have the virtue of being easy to apply. Moreover, little additional work beyond averaging is required to obtain the unit of measurement directly. Indeed, if a unique origin can be established (e.g., a zero level of the property), then the researcher obtains both an absolute origin and a measurement unit. As such, a subjective ratio scale is obtained.

Techniques for Scaling Stimuli

Mission Impossible?

You are making good progress in your first internship and have just been asked by the head of research to present a summary of how your brand is perceived relative to the four major competitor brands… for tomorrow's meeting.

You find the results of a survey conducted just before you came on board. The survey used a paired comparison task to rate preference (10 pairs: 1 vs. 2, 1 vs. 3, 1 vs. 4, etc.), and another simple rank order question for the 5 brands. But your manager specifically asked that preference be displayed on a single continuous (interval or ratio scale).

You're not yet in full blown panic… you average the preference rankings and put them into a symmetric square matrix (Table 7-7)…. but how do you convert a data matrix into a single dimension interval preference scale?

The answer to this problem is that ranking methods may undergo a further transformation (via an intervening scaling model) to produce set of scale values that are interval-scaled. One such transformation, Thurstone's Case V method, is capable of transforming ordinal data obtained from ranking methods. It should be noted that technically speaking, the raw data obtained from ratings methods also requires an intervening model in order to prepare an interval scaled summary measure. However, in this case the model may be no more elaborate than averaging the raw data across respondents and/or response occasions.

Osgood's semantic differential is an illustration of a procedure for dealing with raw data obtained from interval-ratio scale ratings methods. We consider each of these techniques in turn.

Case V Scaling

Thurstone's Case V Scaling model, based on his Law of Comparative Judgment, permits the construction of a unidimensional interval scale using responses from ordinal measurement methods, such as paired comparisons (Thurstone, 1959). This model can also be used to scale ranked data or ordered-category sorts. Several sub cases of Thurstone's model have been developed. We shall first describe the general case and then concentrate on Case V, a special version particularly amenable to application in marketing situations.

Essentially, Thurstone's procedure involves deriving an interval scale from comparative judgments of the type "A is fancier than B", "A is more prestigious than B", "A is preferred to B", and so on. Scale values may be estimated from data in which one individual makes many repeated judgments on each pair of a set of stimuli or from data obtained from a group of individuals with few or no replications per person.

The example should make the Case V procedure easier to follow. Assume that the survey you found had asked 100 homemakers to compare five brands of "fortified juice" with respect to "overall preference of flavor". The homemakers sipped a sample of each brand paired with a sample of every other brand (a total of ten pairs) from paper cups that were marked merely with identifying numbers. Table 7.7 shows the empirically observed proportion for each comparison.

From this table we see that 69 percent of the respondents preferred Juice C to Juice A and the remainder, 31 percent preferred Juice A to Juice C (if we arbitrarily let column dominate row). It is customary to set self-comparisons (the main-diagonal entries of Table 10.4) to 0.5; this has no effect on the resulting scale values (Edwards, 1957). From the data of this table we next prepare Table 7.8, which summarizes the *Z*-values appropriate for each proportion. These *Z*-values were obtained from Table A.1 in Appendix A at the end of this book. If the proportion is less than 0.5, the *Z*-value carries a negative sign; if the proportion is greater than 0.5, the *Z*-value carries a positive sign. The *Z*-values are standard unit variates associated with a given proportion of total area under the normal curve. The Thurstonian model assumes normally distributed scale differences in mean = 0 and standard deviation = 1.0.

For example, from Table 7.7 we note that the proportion of respondents preferring Juice B over Juice A is 0.82. We wish to know the *Z*-value appropriate thereto. This value (labeled *Z* in the standard unit normal table of Table Appendix A.1) is 0.92. That is, 82 percent of the total area under the normal curve is between $Z = -\infty$ and $Z = 0.92$. All remaining entries in Table 10.5 are obtained in a similar manner, a minus sign being prefixed to the *Z*-value when the proportion is *less* than 0.5.

Column totals are next found for the entries in Table 7.8. Scale values are obtained from the column sums by taking a simple average of each column's *Z*-values. For example, from Table 10.5, we note that the sum of the *Z*s for the first column (Juice A) is –0.36. The average *Z* for column A is simply:

$$Z = -\frac{0.36}{5} = -0.072$$

This scale value expresses Juice A as a deviation from the mean of all five scale values. The mean of the five values, as computed from the full row of *Z*s, will always be zero under this procedure. Similarly, we find the average *Z*-value for each of the remaining four columns of Table 7.8.

Table 7.7 Observed Proportions Preferring Brand X to Brand Y

Proportion Preferring Row Brand to Column Brand	*Proportion Preferring Column Brand to Row Brand*				
	A	B	C	D	E
A	0.50	0.82	0.69	0.25	0.35
B	0.18	0.50	0.27	0.07	0.15
C	0.31	0.73	0.50	0.16	0.25
D	0.75	0.93	0.84	0.50	0.59
E	0.65	0.85	0.75	0.41	0.50

Table 7.8 Z-Values Related to Preference Proportions in Table 7.7

	Brand				
Brand	A	B	C	D	E
A	0	0.92	0.50	-0.67	-0.39
B	-0.92	0	-0.61	-1.48	-1.04
C	-0.50	0.61	0	-0.99	-0.67
D	0.67	1.48	0.99	0	0.23
E	0.39	1.04	0.67	-0.23	0
Total	-0.36	4.05	1.55	-3.37	-1.87
Mean (Z)	-0.072	0.81	0.31	-0.674	-0.374
R	0.602	1.484	0.984	0	0.3

Next, since the zero point of an interval scale is arbitrary, we can transform the minimum scale so that it becomes zero. Since Juice D has the lowest value ($R_D = Z_D = -0.674$), we force it to become the reference point (or origin) of zero by adding .674. We then simply add 0.674 to each of the other *Z*-values to obtain the Case V scale values of the other four brands. These are denoted by R^* and appear in the last row of Table 7.8.

The scale values of Juices A through E indicate the preference ordering B > C > A > E > D. Moreover, assuming that an interval scale exists, we can say, for example, that the difference in "goodness of flavor" between Juices B and A is 2.3 times the difference in "goodness of flavor" between Juices C and A, since

$$B - A = 2.3\,(C - A)$$
$$1.484 - 0.602 = 2.3(0.984 - 0.602)$$
$$0.882 = 2.3(0.382) \quad \text{(within rounding error).}$$

The test of this model is how well scale values can be used to work backward—that is, to predict the original proportions. The Case V model appears to fit the data in the example quite well. For any specific brand, the highest mean absolute proportion discrepancy is 0.025 (Juice A). Moreover, the overall mean absolute discrepancy is only .02 (rounded). Even the simplest version (Case V) of the Thurstonian model leads to fairly accurate predictions. The R^* scale values of the Case V model preserve the original rank ordering of the original proportions data.

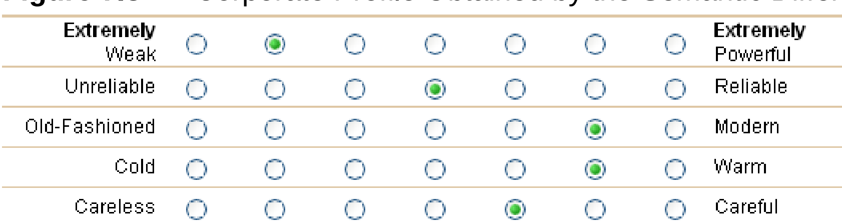
The Semantic Differential

The semantic differential (Osgood, Suci, & Tannenbaum, 1957) is a ratings procedure that results in (assumed interval) scales that are often further analyzed by such techniques as factor analysis (see Chapter 12). Unlike the Case V model, the semantic differential provides no way to test the adequacy of the scaling model itself. It is simply assumed that the raw data are interval-scaled; the intent of the semantic differential is to obtain these raw data for later processing by various multivariate models.

The semantic differential procedure permits the researcher to measure both the direction and the intensity of respondents' attitudes (i.e., measure psychological meaning) toward such concepts as corporate image, advertising image, brand or service image, and country image.

As shown in Figure 7.3, the respondent may be given a set of pairs of antonyms, the extremes of each pair being separated by seven intervals that are assumed to be equal. For each pair of bi-polar adjectives (e.g., powerful/weak), the respondent is asked to judge the concept along the seven-point scale with *implicit* descriptive phrases.

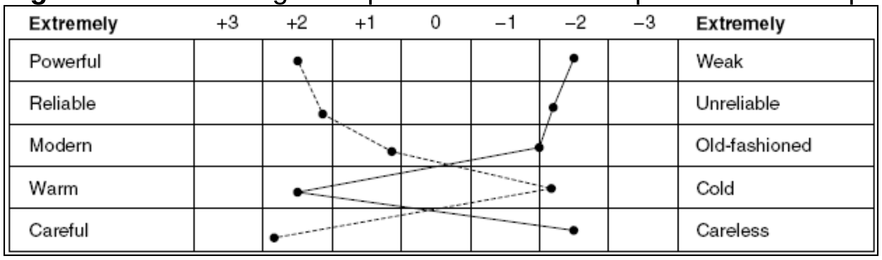
Figure 7.3 Corporate Profile Obtained by the Semantic Differential



In practice, however, profiles would be built up for a large sample of respondents, with many more bipolar adjectives being used than given here.

By assigning a set of integer values, such as +3, +2, +1, 0, –1, –2, –3, to the seven gradations of each bipolar scale in Figure 7.4, the responses can be quantified under the assumption of equal-appearing intervals. These scale values, in turn, can be averaged across respondents to develop semantic differential profiles. For example, Figure 10.5 shows a chart comparing evaluations of Companies X and Y. The average score for the respondents show that the Company X is perceived as very weak, unreliable, old-fashioned, and careless, but rather warm. Company Y is perceived as powerful, reliable, and careful, but rather cold as well; it is almost neutral with respect to the modern/old-fashioned scale.

Figure 7.4 Average-Respondent Profile Comparisons of Companies X and Y

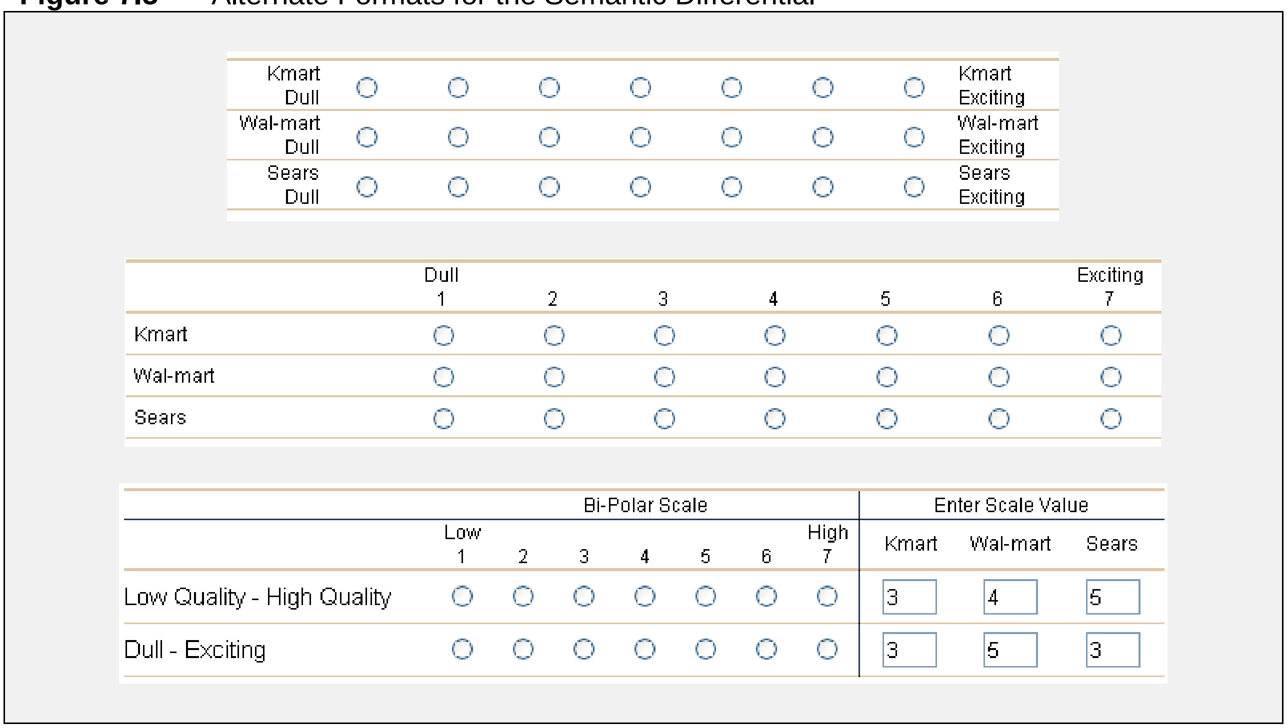


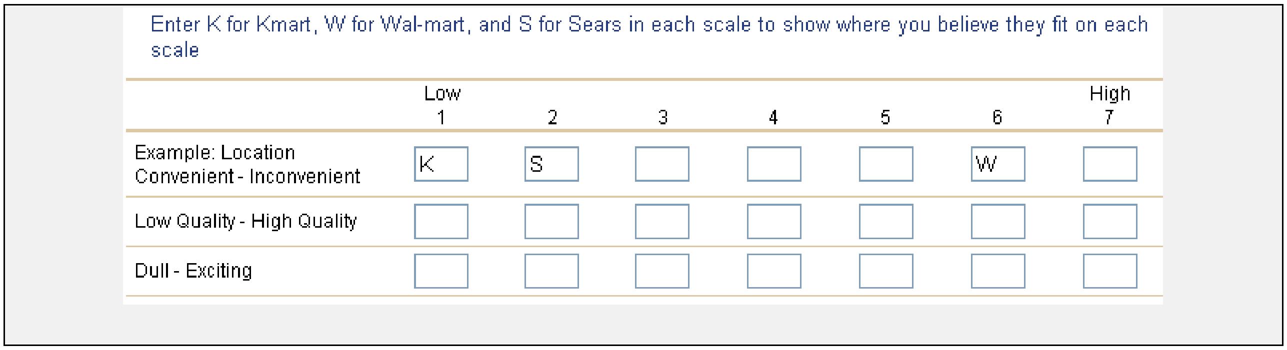
Company X = ———— Company Y = ----------

In marketing research applications, the semantic differential often uses bipolar descriptive phrases rather than simple adjectives, or a combination of both types. These scales are developed for particular context areas, so the scales have more meaning to respondents, thus leading usually to a high degree of reliability.

The same issues of scale construction presented in Table 7.5 apply to the semantic differential. In addition, the researcher must select an overall format for presentation of the scales. As Figure 7.5 illustrates, there are many formatting variations to semantic differential scaling, such as scales that include the numbers 1-7 at the scale points, or numerical comparative scale where respondents make their judgments for KMart, Wal-Mart, and Sears on one attribute before moving to the next one (Golden, Brockett, Albaum, & Zatarain, 1992).

Figure 7.5 Alternate Formats for the Semantic Differential





Enter K for Kmart, W for Wal-mart, and S for Sears in each scale to show where you believe they fit on each scale

	Low 1	2	3	4	5	6	High 7
Example: Location Convenient - Inconvenient	K	S				W	
Low Quality - High Quality							
Dull - Exciting							

The number and type of stimuli to evaluate and the method of administration (personal interview, mail, telephone and email) should determine at least which format the researcher should use. Comparative studies of data quality (including reliability) seem to indicate that the choice of a format may be appropriately made on the basis of ease of subject understanding, ease of coding and interpretation for the researcher, ease of production and display, and cost. If a large number of stimuli are to be evaluated, this would tend to favor use of the graphic positioning or numerical comparative scales.

A Concluding Remark

The semantic differential technique is appropriate for use in a variety of applications:

- Comparing corporate images, both among suppliers of particular products and against an ideal image of what respondents think a company should be
- Comparing brands and services of competing suppliers
- Determining the attitudinal characteristics of purchasers of particular product classes or brands within a product class, including perceptions of the country of origin for imported products
- Analyzing the effectiveness of advertising and other promotional stimuli toward changing attitudes

The comparatively widespread use of the semantic differential by marketing researchers suggests that this method provides a convenient and reasonably reliable way for scaling stimuli (scaling images of brands, corporations, services, etc.), and developing profiles of consumer/buyer attitudes on a wide variety of topics.

Techniques for Scaling Respondents

In contrast to the approaches for scaling stimuli just discussed, researchers also have available techniques whose primary purpose is to scale respondents along some attitude continuum of interest. Two better-known procedures for doing this are the summated scale, and the Q-sort technique. Each of these is described in turn.

The Summated Scale

The summated scale was originally proposed by Rensis Likert (pronounced "lick-ert"), a psychologist (Likert, 1967; Kerlinger, 1973). To illustrate, assume that the researcher wishes to scale some characteristic, such as the public's attitude toward travel and vacations.

To illustrate the Likert scale, a set of seven statements regarding travel and vacations used in a study by a travel company are shown in Figure 7.6. Each of the seven test items has been classified as “favorable” (items 1, 3, and 7) or “unfavorable” (items 2, 4, 5, and 6). Each subject would be asked to indicate their agreement with the statement. The responses are scored +2 for “strongly agree”, +1 for “agree”, 0 for “neither”, –1 for “disagree”, and –2 for “strongly disagree”. Since, we reverse scaled items 2, 4, 5, and 6 on “unfavorable” statements, we would reverse the order of the scale values so as to maintain a consistent direction (+2 would stand for “strongly disagree,” and so on).

Suppose that a subject evaluated the seven items as follows such that the respondent would receive a total score of:

$$+ 2 + 1 + 1 + 2 + 1 + 2 + 2 = 11$$

Suppose that another respondent responded to the seven items by marking (1) strongly disagree, (2) neither, (3) disagree, (4) strongly agree, (5) strongly disagree, (6) strongly agree, and (7) neither. This person’s score would be:

$$- 2 + 0 - 1 - 2 - 2 - 2 + 0 = -9$$

This listing indicates that the second respondent would be ranked “lower” than the first—that is, as having a less-favorable attitude regarding travel and vacations. However, as indicated earlier, a given total score may have different meanings.

Figure 7.6 A Direction-Intensity Scale for Measuring Attitudes Toward Travel and Vacations

Please select the number that best describes your reaction

	Strongly Disagree 1	Disagree 2	Neither Agree nor Disagree 3	Agree 4	Strongly Agree 5
In the winter I need to go south to the sun	○	○	○	○	○
When you take trips with the children you're not really on vacation	○	○	○	○	○
I look for travel bargains	○	○	○	○	○
I "hate" to spend money	○	○	○	○	○
I do not like the fresh air and out-of-doors	○	○	○	○	○
I would feel lost if I were alone in a foreign country	○	○	○	○	○
A good vacation shortens the year and makes life longer	○	○	○	○	○

In applying the Likert summated-scale technique, the steps shown in Table 7.9 are typically carried out.

Table 7.9 Steps in Constructing a Likert Summated Scale

1. The researcher assembles a large number (e.g., 75 to 100) of statements concerning the public’s sentiments toward travel and vacations.
2. Each of the test items is classified by the researcher as generally “favorable” or “unfavorable” to the attitude under study. No attempt is made to scale the items; however, a pretest is conducted that involves the full set of statements and a limited sample of respondents. Ideally, the initial classification should be checked across several judges.
3. In the pretest the respondent indicates approval (or not) with *every* item, checking one of the following direction-intensity descriptors:

Strongly Disaprove or disagree	Disapprove or disagree	Undecided or neither agree nor disagree	Approve or agree	Strongly approve or agree
○	○	○	○	○

4. Each response is given a numerical weight (e.g., +2, +1, 0, −1, −2, or +1 to +5).
5. The individual's *total-attitude score* is represented by the algebraic summation of weights associated with the items checked. In the scoring process, weights are assigned such that the direction of attitude— favorable to unfavorable—is consistent over items. For example, if a + 2 were assigned to "strongly approve/agree" for favorable items, a + 2 should be assigned to "strongly disapprove/disagree" for unfavorable items.
6. On the basis of the results of the pretest, the analyst selects only those items that appear to discriminate well between high and low *total* scorers. This may be done by first finding the highest and lowest quartiles of subjects on the basis of *total* score. Then, the mean differences on each *specific* item are compared between these high and low groups (excluding the middle 50 percent of subjects).
7. The 20 to 25 items finally selected are those that have discriminated "best" (i.e., exhibited the greatest differences in mean values) between high versus low total scorers in the pretest.
8. Steps 3 through 5 are then repeated in the main study.

When analysis is completed, many researchers assume only ordinal properties regarding the placement of respondents along the continuum. Nonetheless, two respondents could have the same total score even though their response patterns to individual items were quite different. That is, the single (summated) score ignores the details of just which items were agreed with and which ones were not. Moreover, the total score is sensitive to how the respondent reacts to the descriptive intensity scale.

Often, a researcher will reverse the polarity of some items in the set (i.e., word items negatively) as a way to overcome the possibility of acquiescence bias (being overly agreeable). Having positively and negatively worded statements hopefully forces respondents with strong positive or negative attitudes to read carefully and use both ends of a scale. A researcher should reverse the polarity of some items, but may need to adjust the scoring, as appropriate. That is, a "strongly agree" response to a positive statement and a "strongly disagree" to a negative statement should be scored the same, and so forth.

Another approach to wording the summated scale adapts the statements into a set of non-directional questions, thereby alleviating the problems associated with mixed-wording scales (Wong, Rindfleisch, & Burroughs, 2003). As an illustration, a non-directional format for one item would be:

"How much pleasure do you get from traveling? [Very little…A great deal]"

In contrast, the normal Likert format for this item is:

"Traveling gives me a lot of pleasure [strongly agree, agree, neither agree nor disagree, disagree, strongly disagree]"

Some final comments are in order. When using this format, Likert (1967) stated that a key criterion for statement preparation and selection should be that all statements be expressions of desired behavior and not statements of fact. Because two persons with decidedly different attitudes may agree on fact, it is recognized that direction is the only meaningful measure obtained when using statements of facts.

The second concern is that the traditional presentation of a Likert scale is one-stage, with both intensity and direction combined. As stated earlier, this may lead to reluctance on the part of respondents to either give extreme scores or use the extreme position on an individual scale item (*central tendency error*). To compensate for this situation the longer two-stage format, whereby direction and intensity are separate evaluations, can be used.

The Q-Sort Technique

The Q-sort technique has aspects in common with the summated scale. Very simply, the task required of a respondent is to sort a number of statements (on individual cards or a pick and group Qualtrics question) into a predetermined number of categories (usually 11) with a specified number having to be placed in each category.

In illustrating the Q-sort technique, assume that four respondents evaluate the test items dealing with travel and vacations. For purposes of illustration, only three groups will be used. The respondents are asked to sort items into:

Group 1 LEAST AGREED WITH	***Group 2*** NEUTRAL ABOUT	***Group 3*** MOST AGREED WITH
-1	0	+1

Suppose that the responses toward seven items by the four respondents, A, B, C, and D, result in the following scale values:

	Respondent			
Item	A	B	C	D
1	+1	+1	–1	–1
2	0	0	0	0
3	+1	0	0	–1
4	-1	–1	+1	+1
5	0	0	0	0
6	–1	–1	+1	+1
7	0	+1	–1	–1

As can be noted, the respondent pairs A & B and C & D seem to be the “most alike” of the six distinct pairs that could be considered. We could, of course, actually correlate each respondent’s scores with every other respondent and, similar to semantic differential applications, and then conduct factor or cluster analyses (see Chapter 14) to group the respondents or items. Typically, these additional steps *are* undertaken in Q-sort studies.

Multi-Item Scales

Each of the types of scales discussed in this chapter can be used either alone or part of a multi-item scale used to measure some construct. A multi-item scale consists of a number of closely related individual rating scales whose responses are combined into a single index, composite score, or value (Peterson, 2000). Often the scores are summated to arrive at a total score. Multi-item scales are used when measuring complex psychological constructs that are not easily defined by just one rating scale or captured by just one question.

The major steps in constructing a multi-item scale. The first, and perhaps most critical, step is to clearly and precisely define the construct of interest. A scale cannot be developed until it is clear just what the scale is intended to measure. This is followed by design and evaluation of the scale. A pool of items is developed and then subject to analysis to arrive at the initial scale. Along the way, a pilot study is conducted to further refine the scale and move toward the final version. Validation studies are conducted to arrive at the final scale. Of concern is construct validation, in which an assessment is made that the scale measures what it is supposed to measure. At the same time that validity data are collected, normative data can also be collected. Norms describe the distributional characteristics of a given population on the scale. Individual scores on the scale then can be interpreted in relation to the distribution of scores in the population (Spector, 1992, p. 9).

A good multi-item scale is both reliable and valid. Reliability is assessed by the scale's stability (test-retest reliability) and internal consistency reliability (coefficient alpha). According to Spector (1992), there are several other characteristics of a good multi-item scale:

- The items should be clear, well-written, and contain a single idea.
- The scale must be appropriate to the population of people who use it, such as having an appropriate reading level.
- The items should be kept short and the language simple.
- Consider possible biasing factors and sensitive items.

Table 7.11 gives an example of a multi-item scale developed to measure consumer ethnocentrism within a nation, the CETSCALE (Shimp & Sharma, 1987). This scale is formatted as a 7 point agreement question in the Likert scale format. A Cetscale score for an individual respondent is obtained as a sum of item ratings, and ranges from 17 to 119 with higher numbers indicating greater consumer ethnocentrism. A compilation of multi-item scales frequently used in consumer behavior and marketing research is provided by Bearden and Netemeyer (1999).

Table 7.11 Example of a Multi-Item Scale: Consumer Ethnocentrism (CETSCALE)

1. American people should always buy American-made products instead of imports.
2. Only those products that are unavailable in the United States should be imported.
3. Buy American-made products. Keep America working.
4. American products first, last and foremost.
5. Purchasing foreign-made products is un-American.
6. It is not right to purchase foreign products.
7. A real American should always buy American-made products.
8. We should purchase products in America instead of letting other countries get rich off us.
9. It is always best to purchase American products.
10. There should be very little trading or purchasing of goods from other countries unless out of necessity.
11. Americans should not buy foreign products, because this hurts American business and causes unemployment.
12. Curbs should be put on all imports.
13. It may cost me in the long run, but I prefer to support American products.
14. Foreigners should not be allowed to put their products on our markets.
15. Foreign products should be taxed heavily to reduce their entry into the United States.
16. We should buy from foreign countries only those products that we cannot obtain within our own country.
17. American consumers who purchase products made in other countries are responsible for putting their fellow Americans out of work.

NOTE: Items composing the 10-item reduced version are items 2, 4 through 8, 11, 13, 16, and 17.

Predictions from attitude scales, preference ratings, and the like still need to be transformed into measures (sales, market share) of more direct value to the marketer. We still do not know, in many cases, how to effectively translate verbalized product ratings, attitudes about corporations, and so on into the behavioral and financial measures required to evaluate the effectiveness of alternative marketing actions.

The Art of Writing Good Questions

Strength of Question Wording

The wording of questions is a critical consideration when obtaining information from respondents. Consider that the following question differing only in the use of the words "should", "could" and "might" was shown to three matched samples of respondents (Payne, 1951, pp. 8–9).

> Do you think anything should be done to make it easier for people to pay doctor or hospital bills? (82 percent replied "Yes".)

For the sample shown the sentence with the word "could", 77 percent replied "Yes", and with "might", 63 percent replied "Yes". These three words are sometimes used as synonyms, and yet at the extreme, responses are 19 percentage points apart.

As another example, Rasinski (1989) posed a question where labels for the topic issue was changed:

> Are we spending too much, too little, or about the right amount on welfare?

In this case, 23.1 percent of respondents replied "too little", but when the label was changed to assistance to the poor, 62.8 percent replied "too little". Questions portraying a more descriptive and positive position may show a large difference in the evaluation score.

Reducing Question Ambiguity

Reducing ambiguity and bias is critical in both the respondent's understanding and proper consideration of the question and in the researcher's understanding of the answer's meaning. In this section we discuss issues of question structure and form that can greatly influence and improve the quality of your questionnaire.

The Qualtrics "Survey University" and survey and question libraries provide many suggestions and helpful examples for writing unambiguous questions. Writing questions is an art, which like all arts requires a great amount of work, practice, and help from others. In Exhibit 7.4 we provide an overview of the common pitfalls we often see in "bad questionnaires" that lead to various forms of ambiguity.

Exhibit 7.4 Key Considerations for Reducing Question Ambiguity

1. **Strength of Question Wording** The wording of questions is a critical consideration when obtaining information from respondents. One study "should", "could" and "might" was shown to three matched samples of respondents (Payne, 1951, pp. 8–9). Do you think anything should be done to make it easier for people to pay doctor or hospital bills? (82 percent replied "Yes".) For the sample shown the sentence with the word "could", 77 percent replied "Yes", and with "might", 63 percent replied "Yes". These three words are sometimes used as synonyms, and yet at the extreme, responses are 19 percentage points apart. Questions portraying a more descriptive and positive position may show a large difference in the evaluation score.

2. **Avoid loaded or leading words or questions.** Slight wording changes can produce great differences in results. Could, Should, Might all sound almost the same, but may produce a 20% difference in agreement to a question (The supreme court could.. should.. might.. have forced the breakup of Microsoft Corporation). Strong words that represent control or action, such as prohibit produces similar results (Do you believe that congress should prohibit insurance companies from raising rates?) Sometimes wording is just biased: You wouldn't want to go to Rudolpho's Restaurant for the company's annual party would you?

3. **Framing effects.** Information framing effects reflect the difference in response to objectively equivalent information depending upon the manner in which the information is labeled or framed. Levin, Schneider, and Gaeth (1998) and Levin et al. (2001) identify three distinct types of framing effects:

   - Attribute framing effects occur when evaluations of an object or product are more favorable when a key attribute is framed in positive rather than negative terms.
   - Goal framing effects occur when a persuasive message has different appeal depending on whether it stresses the positive consequences of performing an act to achieve a particular goal or the negative consequences of not performing the act.
   - Risky choice framing effects occur when willingness to take a risk depends upon whether potential outcomes are positively framed (in terms of success rate) or negatively framed (in terms of failure rate).

   Which type of potential framing effects should be of concern to the research designer depends upon the nature of the information being sought in a questionnaire. At the simplest level, if intended purchase behavior of ground beef was being sought, the question could be framed as "80 percent lean" or "20 percent fat." This is an example of attribute framing. It should be obvious that this is potentially a pervasive effect in question design, and is something that needs to be addressed whenever it arises. More detailed discussion of these effects is given by Hogarth (1982).

4. **Misplaced questions.** Questions placed out of order or out of context should be avoided. In general, a funnel approach is advised. Broad and general questions at the beginning of the questionnaire as a warm-up (What kind of restaurants do you most often go to?). Then more specific questions, followed by more general easy to answer questions (like demographics) at the end of the questionnaire.

5. **Mutually non-exclusive response categories.** Multiple choice response categories should be mutually exclusive so that clear choices can be made. Non-exclusive answers frustrate the respondent and make interpretation difficult at best.

6. **Nonspecific questions.** Do you like orange juice? This is very unclear...do I like what? Taste, texture, nutritional content, Vitamin C, the current price, concentrate, fresh squeezed? Be specific in what you want to know about. Do you watch TV regularly? (what is regularly?).

7. **Confusing or unfamiliar words.** Asking about caloric content, acrylamide, phytosterols, and other industry specific jargon and acronyms are confusing. Make sure your audience understands your language level, terminology and above all, what you are asking.

8. **Non-directed questions give respondents excessive latitude.**
   What suggestions do you have for improving tomato juice? The question is about taste, but the respondent may offer suggestions about texture, the type of can or bottle, mixing juices, or something related to use as a mixer or in recipes.

9. **Forcing answers.** Respondents may not want, or may not be able to provide the information requested. Privacy is an important issue to most people. Questions about income, occupation, finances, family life, personal hygiene and beliefs (personal, political, religious) can be too intrusive and rejected by the respondent.

10. **Non-exhaustive listings.** Do you have all of the options covered? If you are unsure, conduct a pretest using the "Other (please specify) __________" option. Then revise the question making sure that you cover at least 90% of the respondent answers.

11. **Unbalanced listings.** Unbalanced scales may be appropriate for some situations and biased in others. When measuring alcohol consumption patterns, one study used a quantity scale that made the heavy drinker appear in the middle of the scale with the polar ends reflecting no consumption and an impossible amount to consume. However, we expect all hospitals to offer good care and may use a scale of excellent, very good, good, fair. We do not expect poor care.

12. **Double barreled questions.** What is the fastest and most convenient Internet service for you? The fastest is certainly not the most economical. The double barreled question should be split into two questions.

13. **Independent answers.** Make sure answers are independent. For example the question "Do you think basketball players as being independent agents or as employees of their team?" Some believe that yes, they are both.

14. **Long questions.** Multiple choice questions are the longest and most complex. Free text answers are the shortest and easiest to answer. When you Increase the length of questions and surveys, you decrease the chance of receiving a completed response.

15. **Questions on future intentions.** Yogi Berra (Famous New York Yankees Baseball Player) once said that making predictions is difficult, especially when they are about the future. Predictions are rarely accurate more than a few weeks or in some case months ahead.

VALIDITY AND RELIABILITY OF MEASUREMENT

The content of a measurement instrument includes a subject, theme, and topics that relate to the characteristics being measured. However the measuring instrument does not include all of the possible items that could have been included. When measuring complex psychological constructs such as perceptions, preferences, and motivations, hard questions must be asked to identify the items most relevant in solving the research problem:

1. Do the scales really measure what we are trying to measure?
2. Do subjects' responses remain stable over time?
3. If we have a variety of scaling procedures, are respondents consistent in their scoring over those scales that purport to be measuring the same thing?

By solving these problems, we establish the validity and reliability of scaling techniques.

Note that our focus is only on the general concepts and measures of validity and reliability that are used in cross-sectional studies. There is little documented research on issues of measure reliability and validity for time-series analysis.

Validity

Validity simply means that we are measuring what we believe we are measuring. The data must be unbiased and relevant to the characteristic being measured. The validity of a measuring instrument reflects the absence of systematic error. Systematic error may arise from the instrument itself, the user of the instrument, the subject, or the environment in which the scaling procedure is being administered. Since in practice we rarely know true scores, we usually have to judge a scaling procedure's validity by its relationship to other relevant standards.

The validity of a measuring instrument hinges on the availability of an external criterion that is thought to be correct. Unfortunately the availability of such outside criteria is often low. What makes the problem even more difficult is that the researcher often is not interested in the scales themselves, but the underlying theoretical construct that the scale purports to measure. It is one thing to define IQ as a score on a set of tests; it is quite another to infer from test results that a certain construct, such as intelligence, or a dimension of intelligence is being measured.

In testing the validity of a scale, the researcher must be aware that many forms of validity exist, including (1) Content validity, (2) Criterion validity, and (3) Construct validity.

Content Validation

Content Validity concerns how the scale or instrument represents the universe of the property or characteristic being measured. It is essentially judgmental and is ordinarily measured by the personal judgments of experts in the field. That is, several content experts may be asked to judge whether the items being used in the instrument are representative of the field being investigated. Closely related to this approach for assessing content validation is a method involving known groups. For instance, a scale purported to measure attitudes toward a brand could be tested by administering it to a group of regular buyers of the product (which presupposes a favorable attitude) and compared with those from a group of former buyers or other non-buyers (who presumably have a negative attitude). If the scale does not discriminate between the two groups, then its validity with respect to measuring brand attitude is highly questionable. Caution must be exercised in using this method in that other group differences besides their known behavior might exist, and account for the differences in measurement.

Face Validity is a preliminary or exploratory form of content validity. It is based on a cursory review of items by non-experts such as one's wife, mother, tennis partner, or similarly convenient to access persons. A simple approach is to show the measurement instrument to a convenient group of untrained people and ask whether or not the items look okay (Litwin, 2003).

Logical Validation refers simply to an intuitive, or common-sense, evaluation. This type of validation is derived from the careful definition of the continuum of a scale and the selection of items to be scaled. Thus, in an extreme case, the investigator reasons that everything that is included is done so because it is obvious that it should be that way. Because things often do not turn out to be as obvious as believed, it is wise for the marketing researcher not to rely on logical validation alone.

> **Example**: of research lacking content validity is the Coca-Cola Company's introduction many years ago of New Coke. Since the product represented a major change in taste, thousands of consumers were asked to taste New Coke. Overwhelmingly, people said they liked the new flavor. With such a favorable reaction, why did the decision to introduce the product turn out to be a mistake? Executives of the company acknowledge that the consumer survey conducted omitted a crucial question. People were asked if they like the new flavor, but they were not asked if they were willing to give up the old Coke. In short, they were not asked if they would buy the new product in place of the old one.

Criterion Validation

Criterion Validity, also known as *pragmatic validity*, has two basic dimensions known as *predictive validity* and *concurrent validity.* They question if the instrument works and are better decisions can be made with it than without it?

The New Coke example also illustrates a case of poor predictive validity. The measures of liking, and so on, were not very good predictors of purchase, which was the real measure of managerial interest.

In concurrent validity, a secondary criterion, such as another scale, is used to compare results. Concurrent validity can be assessed by correlating the set of scaling results with some other set, developed from another instrument administered at the same time. Often product researchers will ask a question like "Overall, how much do you prefer Brand A soft drink?", and then follow with another question such as, "Given the following four brands, indicate the percentage of your total soft drink purchases that you would make for each brand."

Alternatively, the correlation may be carried out with the results of the same question asked again later in the survey or on another testing occasion.

Construct Validation

In *construct validation* the researcher is interested both in the question, "Does it work?" (i.e., predict), and in developing criteria that permit answering theoretical questions of why it works and what deductions can be made concerning the theory underlying the instrument. Construct validity involves three subcases: convergent, discriminant, and nomological validity.

Convergent Validity: The correspondence in results between attempts to measure the same construct by two or more independent methods. These methods need not all be scaling techniques.

Discriminant Validity: Refers to properties of scaling procedures that do differ when they are supposed to—that is, in cases where they measure different characteristics of stimuli and/or subjects. That is, more than one instrument and more than one subject characteristic should be used in establishing convergent-discriminant validity. Discriminant validity concerns the extent to which a measure is unique (and not simply a reflection of other variables), and as such it provides the primary test for the presence of method variance.

Nomological Validity: "Understanding" a concept (or construct). In nomological validity the researcher attempts to relate measurements to a theoretical model that leads to further deductions, interpretations, and tests, gradually building toward a nomological net, in which several constructs are systematically interrelated.

Ideally, the marketing researcher would like to attain construct validity, thus achieving not only the ability to make predictive statements but understanding as well. Specifically, more emphasis should be placed on the theories, the processes used to develop the measures, and the judgments of content validity.

Reliability

Reliability is concerned with the consistency of test results over groups of individuals or over the same individual at different times. A scale may be reliable but not valid. Reliability, however, establishes an upper bound on validity. An unreliable scale cannot be a valid one. Reuman (1982, p. 1099) states that "according to classical test theory, highly reliable measures are necessary, but not sufficient, for demonstrating high construct validity or high criterion validity."

The achievement of scale reliability is, of course, dependent on how consistent the characteristic being measured is from individual to individual (homogeneity over individuals) and how stable the characteristic remains over time. Just how reliable a scaling procedure turns out to be will depend on the dispersion of the characteristic in the population, the length of the testing procedure, and its internal consistency. Churchill and Peter (1984) concluded that rating scale estimates were largely determined by measuring characteristics such as number of items in a scale, type of scale, and number of scale points. They further concluded that sampling characteristics and measurement development processes had little impact.

In general, a measurement of the reliability of a scale (or measurement instrument) may be measured by one of three methods: test-retest, alternative forms, or internal consistency. The basics of reliability in a marketing context are reviewed by Peter (1979).

Test-Retest

The *test-retest method* examines the stability of response over repeated applications of the instrument. Do we achieve consistent results, assuming that the relevant characteristics of the subjects are stable over trials? One potential problem, of course, is that the first measurement may have an effect on the second one. Such effects can be reduced when there is a sufficient time interval between measurements. If at all possible, the researcher should allow a minimum of two weeks to elapse between measurements. Reliability may be estimated by any appropriate statistical technique for examining differences between measures.

Alternative Forms

The *alternative forms method* attempts to overcome the shortcomings of the test-retest method by successively administering equivalent forms of the measure to the same sample. Equivalent forms can be thought of as instruments built such that the same types and structures of questions are included on each form, but where the specific questions differ. The forms of the measurement device may be given one after the other or after a specified time interval, depending upon the investigator's interest in stability over time. Reliability is estimated by correlating the results of the two equivalent forms.

Internal Consistency

Internal consistency refers to estimates of reliability within single testing occasions. In a sense it is a modification of the alternative form approach, but differs in that alternatives are formed by grouping variables. The basic form of this method is split-half reliability, in which items are divided into equivalent groups (say, odd- versus even-numbered questions, or even a random split) and the item responses are correlated. In practice, any split can be made.

A potential problem arises for split-half in that results may vary depending on how the items are split in half. A way of overcoming this is to use coefficient alpha, known also as Cronbach's alpha, which is a type of mean reliability coefficient for all possible ways of splitting an item in half (Cronbach, 1951). Whenever possible, alpha should be used as a measure of the internal consistency of multi-item scales. Alpha is perhaps the most widely used measure of internal consistency for multiple-item measures within marketing research. One caution however, is that there should be a sufficient number of items in the measure so that alpha becomes meaningful. Alpha has been used for as few as two items, and this essentially amounts to a simple correlation between the two. Although there is no generally acceptable heuristic covering the number of items, common sense would indicate that the minimum number of items should be four or perhaps even six. What is clear, however, is that alpha is a function of the number of items in a scale (i.e., the more items, the greater alpha will tend to be), and also a function of the intercorrelation of the items themselves (Cortina, 1993; Voss, Stem, & Fotopoulos, 2000). Consequently, when interpreting an obtained alpha, the number of items must always be kept in mind.

The usual application of coefficient alpha is to calculate it using a statistical analysis packages, report it, and assess whether the value obtained exceeds some rule-of-thumb minimum value, typically 0.70. There now exist methods to make inferential tests about the size of alpha, and to attach confidence intervals to the measure (Iacobucci & Duhachek, 2003).

In many projects, measurements or evaluations are made by more than a single evaluator. Sometimes this is done when coding answers to open-ended questions. In these situations, the researcher is interested in the reliability of these evaluations. This is known as *interrater* or *interobserver reliability*. The most common measure used is a correlation (Litwin, 2003).

A Concluding Comment

Although it is not our objective to pursue in detail the methods by which reliability or validity can be tested, we hope to provide an appreciation of the difficulties encountered in designing and analyzing psychological measure. One question that has not been answered is, "What is a satisfactory level of reliability, or what minimum level is acceptable?" There is no simple definitive answer to this question. Much depends on the investigator's or decision maker's primary purpose in measurement and on the approach used to estimate reliability. In trying to arrive at what constitutes satisfactory reliability, the investigator must at all times remember that reliability can affect certain qualities of a study including (1) Validity, (2) The ability to show relationships between variables, and (3) The making of precise distinctions among individuals and groups.

SUMMARY

This chapter focused on general concepts of measurement. We discussed the role of definitions and defined concepts, constructs, variables, operational definitions, and propositions. We then turned to measurement and examined what it is and how measurement relates to development of scales. Also discussed, but rather briefly, were alternative sources that cause variations within a set of measurements derived from a single instrument. This was followed by a description of different types of scales that are commonly used in marketing research. Advanced scaling techniques for scaling stimuli and respondents were also discussed. We concluded with a brief overview of measurement validity and reliability, and the various types of each that are of concern to an investigator.

REFERENCES

Albaum, G., Best, R., & Hawkins, D. (1981). Continuous vs. discrete semantic differential scales. *Psychological Reports, 49,* 83–86.

Bearden, W. O., & Netemeyer, R. G. (1999). *Handbook of Marketing Scales* (2nd ed.). Thousand Oaks, CA: Sage.

Churchill, G. A., Jr., & Peter, J. P. (1984, November). Research design effects on the reliability of rating scales: A meta analysis. *Journal of Marketing Research*, *21*, 360–375.

Coombs, C. H. (1964). *A Theory of Data*. New York: Wiley.

Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104.

Cronbach, L. J. (1951, September). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Davis, D. W. (1997). Nonrandom measurement error and race of interviewer effects among African-Americans. *Public Opinion Quarterly*, *61*, 183–207.

Golden, L., Brockett, P., Albaum, G., & Zatarain, J. (1992). The golden numerical comparative scale format for economical multi-object/multiattribute comparison questionnaires. *Journal of Official Statistics*, *8* (1), 77–86.

Guttman, L. (1985). Measuring the true-state of opinion. In R. Ferber & H. Wales (Eds.), *Motivation and market behavior*. Homewood, IL: Richard D. Irwin, 393–415.

Haley, R., & Case, P. (1979, Fall). Testing thirteen attitude scales for agreement and brand determination. *Journal of Marketing*, *43*, 20–32.

Hogarth, R. M. (Ed.) (1982). *Question framing and response consistency*. San Francisco: Jossey-Bass.

Hubbard, R., Little, E. L., & Allen, S. J. (1989). Are responses measured with graphic rating scales subject to perceptual distortion? *Psychological Reports*, *69*, 1203–1207.

Iacobucci, D., & Duhachek, A. (2003). *Applying confidence intervals to coefficient alpha.* Unpublished working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

Kerlinger, F. (1973). *Foundation of behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.

Levin, I. P., Gaeth, G. J, Evangelista, F., Albaum, G., & Schreiber, J. (2001). How positive and negative frames influence the decisions of persons in the United States and Australia. *Asia Pacific Journal of Marketing and Logistics*, *13*(2), 64–71.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998, November). All frames are not created equal: A typology of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188.

Likert, R. (1967). The method of constructing an attitude scale. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley, pp. 90–95.

Litwin, M. S. (2003). *How to assess and interpret survey psychometric* (2nd ed.). Thousand Oaks, CA: Sage.

McCarty, J. A., & Shrum, L. J. (2000). The measurement of personal values in survey research: A test of alternative rating procedures. *Public Opinion Quarterly, 64,* 271–298.

Morgan, M. (2003, March 31). Be careful with survey data. *Marketing News, 37,* 26.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Peter, J. P., & Churchill, G. A., Jr. (1986, February). Relationships among research design choices and psychometric properties of rating scales: A meta-analysis. *Journal of Marketing Research, 23,* 1–10.

Peterson, R. A. (2000). *Creating effective questionnaires*. Thousand Oaks, CA: Sage.

Payne, S. L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.

Rasinski, K. A. (1989). The effect of question wording on public support for government spending. *Public Opinion Quarterly, 53,* 388–394.

Reuman, D. A. (1982). Ipsative behavioral variability and the quality of thematic apperceptive measurement of the achievement motive. *Journal of Personality and Social Psychology, 43*(5), 1098–1110.

Semon, T. T. (1999, August 30). Scale ratings always betrayed by arithmetic. *Marketing News, 33,* 7.

Semon, T. T. (2001, October 8). Symmetry shouldn't be goal for scales. *Marketing News, 35,* 9.

Shimp, T. A., & Sharma, S. (1987, August). Consumer ethnocentrism: Construction and validation of the CETSCALE. *Journal of Marketing Research, 24,* 280–289.

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 677–680.

Sudman, S., & Bradburn, N. (1983). *Asking questions*. San Francisco: Jossey-Bass.

Thurstone, L. L (1959). *The measurement of values*. Chicago: University of Chicago Press.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Voss, K. E., Stem, D. E, Jr., and Fotopoulos, S. (2000). A comment on the relationship between coefficient alpha and scale characteristics. *Marketing Letters*, *11*(2), 177—191.

Whitlark, D., & Smith, S. (2004). "Pick and Choose " Marketing Research Volume 16, Issue 4, Pages 8-14, American Marketing Association, Chicago, December, 2004

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003, June). Do reverse-worded items confound measures in cross-cultural research? The case of the material values scale. *Journal of Consumer Research, 30,* 72–91.

Chapter 11
HYPOTHESIS TESTING AND UNIVARIATE ANALYSIS

Scientific research is directed at the inquiry and testing of alternative explanations of what appears to be fact. For behavioral researchers, this scientific inquiry translates into a desire to ask questions about the nature of relationships that affect behavior within markets. It is the willingness to formulate hypotheses capable of being tested to determine (1) what relationships exist, and (2) when and where these relationships hold.

The first stage in the analysis process is identified to include editing, coding, and making initial counts of responses (tabulation and cross tabulation). In the current chapter, we then extend this first stage to include the testing of relationships, the formulation of hypotheses, and the making of inferences.

In *formulating hypotheses* the researcher uses "interesting" variables, and considers their relationships to each other, to find suggestions for working hypotheses that may or may not have been originally considered. In *making inferences*, conclusions are reached about the variables that are important, their parameters, their differences, and the relationships among them. A *parameter* is a summarizing property of a collectivity—such as a population—when that collectivity is not considered to be a sample (Mohr, 1990, p.12).

Although the sequence of procedures, (a) formulating hypotheses, (b) making inferences and (c) estimating parameters is logical, in practice these steps tend to merge and do not always follow in order. For example, the initial results of the data analysis may suggest additional hypotheses that in turn require more and different sorting and analysis of the data. Similarly, not all of the steps are always required in a particular project; the study may be exploratory in nature, which means that it is designed more to formulate the hypotheses to be examined in a more extensive project, than to make inferences or estimate parameters.

AN OVERVIEW OF THE ANALYSIS PROCESS

The overall process of analyzing and making inferences from sample data can be viewed as a process of refinement that involves a number of separate and sequential steps that may be identified as part of three broad stages:

1. *Tabulation*: identifying appropriate categories for the information desired, sorting the data by categories, making the initial counts of responses, and using summarizing measures to provide economy of description and thereby facilitate understanding.
2. *Formulating additional hypotheses*: using the inductions derived from the data concerning the relevant variables, their parameters, their differences, and their relationships to suggest working hypotheses not originally considered.
3. *Making inferences*: reaching conclusions about the variables that are important, their parameters, their differences, and the relationships among them.

The Data Tabulation Process

Seven steps are involved in the process of data tabulation:

1. *Categorize.* Define appropriate categories for coding the information collected.
2. *Editing and Coding Data.* Assign codes to the respondent's answers.
3. *Create the Data File.* Enter the data into the computer and create a data file.
4. *Error Checking and Handling Missing Data.* Check the data file for errors by performing a simple tabulation analysis to identify errors in coding or data entry. Once errors are identified, data may be edited or recoded to collapse, combine or delete responses or categories.
5. *Generate New Variables.* New variables may be computed by data manipulations that multiply, sum, or otherwise transform variables.
6. *Weight Data Subclasses.* Weights are often used to adjust the proportionate representation of sample subgroups so that they match the proportions found in the population.
7. *Tabulate.* Summarize the responses to each variable included in the analysis.

As simple as these steps are from a technical standpoint, data management is most important in assuring a quality analysis and thereby merit an introductory discussion. A more in depth discussion of survey-based data management is provided by Fink (2003, Chap.1).

Defining Categories

The raw input to most data analyses consists of the basic data matrix, as shown in Table 11.1. In most data matrices, each row contains a respondent's data and the columns identify the variables or data fields collected for the respondent. The analyses of a column of data might include a tabulation of data counts in each of the categories or the computation of the mean and standard deviation. This analysis is often done simply because we want to summarize the meaning of the entire column of values. In so doing we often (willingly) forgo the full information provided by the data in order to understand some of its basic characteristics, such as central tendency, dispersion, or categories of responses. Because we summarize the data and make inferences from it, it is doubly important that the data be accurate.

Tabulation of any sizable array of data often requires that responses be grouped into categories or classes. The identification of response categories early in the study has several advantages. Ideally, it forces the analyst to consider all possible interpretations and responses to the questionnaire. It often leads to improvements in the questionnaire or observation forms. It permits more detailed instruction of interviewers and results in higher consistency in interpreting responses. Editing problems are also reduced.

The definition of categories allows for identification of the database columns and values assigned to each question or variable and to indicate the values assigned to each response alternative. Depending on the data collection method, data code sheets can be prepared and pre-coded. Data files are often formatted as comma separated variable (CSV) files, meaning that each variable appears in the same relative position for each respondent with a comma separating each of the variables. The major data analysis software programs read data files and then display them in a spreadsheet-like database (see Table 11.1). Often the data are entered directly into a Microsoft Excel spreadsheet for import into the statistical program to be used for analysis. Where data is not collected and formatted electronically, pre-coding of printed questionnaires will eliminate transcription and thereby decreasing both processing errors and costs. Most of today's computer-based software for telephone (CATI) or internet surveys (Qualtrics.com) automate this entire process. They not only define the question categories in the database but also automatically build the database and record the completed responses as they are submitted. The data may then be analyzed online, exported to Microsoft Excel™ or imported into a dedicated statistical analysis program such as PASW (formerly known as SPSS). Response categories are coded from 1 for the first category to the highest value for the last category. Category values can be recoded to assign different numbers as desired by the researcher.

As desirable as the early definition of categories is, it can sometimes only be done after the data have been collected. This is usually the case when open-end text questions, unstructured interviews, and projective techniques are used.

The selection of categories is controlled by both the purposes of the study and the nature of the responses. Useful classifications meet the following conditions:

1. *Similarity of response within the category*. Each category should contain responses that, for purposes of the study, are sufficiently similar that they can be considered homogenous.
2. *Differences of responses between categories*. Differences in category descriptions should be great enough to disclose any important distinctions in the characteristic being examined.
3. *Mutually exclusive categories*. There should be an unambiguous description of categories, defined so that any response can be placed in only one category.
4. *Categories should be exhaustive*. The classification schema should provide categories for all responses.

The use of extensive open-end questions often provides rich contextual and anecdotal information, but is a practice often associated with fledgling researchers. Open-end questions, of course, have their place in marketing research. However, the researcher should be aware of the inherent difficulties in questionnaire coding and tabulation, not to mention their tendency to be more burdensome to the respondent. All of this is by way of saying that any open-end question should be carefully checked to see if a closed-end question (i.e., check the appropriate box) can be substituted without doing violence to the intent of the question. Obviously, sometimes this substitution should not be made.

Editing and Coding

Editing is the process of reviewing the data to ensure maximum accuracy and clarity. This applies to the editing of the collection forms used for pretesting as well as those for the full-scale project. Careful editing during the pre-test process will often catch misunderstandings of instructions, errors in recording, and other problems so as to eliminate them for the later stages of the study. Early editing has the additional advantage of permitting the questioning of interviewers while the material is still relatively fresh in their minds. Obviously, this has limited application for printed questionnaires, though online or CATI surveys can be edited even when data is being collected.

Editing is normally centralized so as to ensure consistency and uniformity in treatment of the data. If the sample is not large, a single editor usually edits all the data to reduce variation in interpretation. In those cases where the size of the project makes the use of more than one editor mandatory, it is usually best to assign each editor a different portion of the data collection form to edit. In this way the same editor edits the same items on all forms, an arrangement that tends to improve both consistency and productivity.

Typically, interviewer and respondent data are monitored to ensure that data requirements are fulfilled. Each collection form should be edited to ensure that data quality requirements are fulfilled. Regarding data obtained by an interviewer (and to an extent self-report) the following should be specifically evaluated:

1. *Legibility of entries*. Obviously the data must be legible in order to be used. Where not legible, although it may be possible to infer the response from other data collected, where any real doubt exists about the meaning of data it should not be used.

2. *Completeness of entries*. On a fully structured collection form, the absence of an entry is ambiguous. It may mean either that the respondent could not or would not provide the answer, that the interviewer failed to ask the question, or that there was a failure to record collected data.

3. *Consistency of entries*. Inconsistencies raise the question of which response is correct. (If a respondent family is indicated as being a non-watcher of game shows, for example, and a later entry indicates that they watched *Wheel of Fortune* twice during the past week, an obvious question arises as to which is correct.) Discrepancies may be cleared up by questioning the interviewer or by making callbacks to the respondent. When discrepancies cannot be resolved, discarding both entries is usually the wisest course of action.

4. *Accuracy of entries*. An editor should keep an eye out for any indication of inaccuracy in the data. Of particular importance is the detection of any repetitive response patterns in the reports of individual interviews. Such patterns may well be indicative of systematic interviewer or respondent bias or dishonesty.

Coding is the process of assigning respondent answers to data categories and numbers are assigned to identify them with the categories. *Pre-coding* refers to the practice of assigning codes to categories. Sometimes these codes are printed on structured questionnaires and observation forms before the data are collected. Using these predefined codes, the interviewer is able to code the responses when interpreting the response and marking the category into which it should be placed.

Post-coding is the assignment of codes to responses after the data are collected, and is most often required when responses are reported in an unstructured format (open-ended text or numeric input). Careful interpretation and good judgment are required to ensure that the meaning of the response and the meaning of the category are consistently and uniformly matched.

When not using CATI or online data collection technologies, a formal coding manual or codebook is often created and made available to those who will be entering or analyzing the data. The codebook used for a study of supermarkets in the United States is shown in Figure 11.1 as an illustration.

Like good questionnaire construction, good coding requires training and supervision. The editor-coder should be provided with written instructions, including examples. He or she should be exposed to the interviewing of respondents and become acquainted with the process and problems of collecting the data, thus providing aid in its interpretation. The coder also should be aware of the computer routines that are expected to be applied, insofar as they may require certain kinds of data formats.

Whenever possible (and when cost allows) more than one person should do the coding, specifically the post-coding. By comparing the results of the various coders, a process known as determining *inter-coder reliability*, any inconsistencies can be brought out. In addition to the obvious objective of eliminating data coding inconsistencies, the need for recoding sometimes points to the need for additional categories for data classification and may sometimes mean that there is a need to combine some of the categories. Coding is an activity that should not be taken lightly. Improper coding leads to poor analyses and may even constrain the types of analysis that can be completed.

Qualtrics has an interesting feature that uses a "wizard" to take a survey by selecting random choices and following the various logic paths available. The resulting test data conforms to the "sample size" specified by the researcher and the pre-specified logic and coding can be checked for errors that the researcher has made.

Figure 11.1 Codebook for Comparative Supermarket Study

Question Number	Variable		
	01	Sample group 1=Group A 2=Group B	
	02	Respondent ID number xxx=actual number	
1	03	Residential district 1=SE 2=SW 3=NW 4=NE	
2	04	How often shop at Albertson's xx=actual number	
2	05	How often shop at Raley's xx=actual number	
2	06	How often shop at Wal-Mart xx=actual number	
2	07	How often shop at Smith's xx=actual number	
3	08	Primary shopper 1=self 2=spouse 3=parent(s) 4=housekeeper 5=other	
4A	09	Store most likely to shop at 1=Albertson's 2=Raley's 3=Wal-Mart 4=Smith's 5=Other	
4B	10	Time to get to store xxx=actual number	
4C	11	How to get there 1=car/taxi 2=bus 3=walk	
5	12	Amount spent at Albertson's xxxxx=actual number	
5	13	Amount spent at Raley's xxxxx=actual amount	
5	14	Amount spent at Wal-Mart xxxxx=actual number	
5	15	Amount spent at Smith's xxxxx=actual number	
	16	BLANK	
6	17	Supermarket evaluated 1=Albertson's 2=Smith's	
6	18-35	Semantic scales for Albertson's x=1 to 7, starting from the left-side of scale	
		location (18)	layout (27)
		prices (19)	shopping experience (28)
		atmosphere (20)	reputation (29)
		quality of products (21)	service (30)
		modern (22)	helpfulness of clerks (31)
		friendliness of clerks (23)	dull (32)
		customers (24)	selection of products (33)
		cluttered (25)	dirty (34)
		check-out (26)	like (35)
6	36-53	Semantic scales for Smith's X=1 to 7, starting from the left-side of the scale	
		location (36)	layout (45)
		prices (37)	shopping experience (46)
		atmosphere (38)	reputation (47)
		quality of products (39)	service (48)
		modern (40)	helpfulness of clerks (49)
		friendliness of clerks (41)	dull (50)
		customers (42)	selection of products (51)
		cluttered (43)	dirty (52)
		check-out (44)	like (53)
7	54	Gender 1=female 2=male	
8	55	Marital status 1=single 2=married 3=divorced/separated/widowed 4=other	
9	56	Age xx=actual number	
10	57	Employment status 1=full time 2=part time 3=not employed	

Tabulation: Cleaning the Data

The purpose of the initial data cleaning tabulation is to identify outliers, missing data and other indications of data, coding, transcription, or entry errors. The tabulation of responses will invariably reveal codes that are out of range or otherwise invalid. For example, one tabulation might reveal 46 males (category 1), 54 females (category 2), and one category 5 response, which is obviously an error. Some errors, such as the preceding one, represent entry of values that are out-of-range or wild codes (Lewis-Beck, 1995, p. 7). That is, the value is not one that has been assigned to a possible response to a question. A miscoding error that is more difficult to detect is one where an erroneous recording of a response category is made using a number that is assigned to another response category. That is, in the coding shown in Figure 11.1 a response of self to question number 3 (code = 1) might have been coded as spouse (code = 2). Hopefully, not too many errors of this type occur.

Table 11.1 Illustration of Data Matrix

	Variable						
Object	*1*	*2*	*3*	…	*j*	…	*m*
1	X_{11}	X_{12}	X_{13}	…	X_{1j}	…	X_{1m}
2	X_{21}	X_{22}	X_{23}	…	X_{2j}	…	X_{2m}
3	X_{31}	X_{32}	X_{33}	…	X_{3j}	…	X_{3m}
.	.	.	.		.		.
.	.	.	.		.		.
i	X_{i1}	X_{i2}	X_{i3}	…	X_{ij}	…	X_{im}
.	.	.	.		.		.
.	.	.	.		.		.
n	X_{n1}	X_{n2}	X_{n3}	…	X_{nj}	…	X_{nm}

An aspect of cleaning the data is dealing with missing data. That is, some respondents may not provide responses for all the questions. One way of handling this is to use statistical imputation. This involves estimating how respondents who did not answer particular questions would have answered if they had chosen to. Researchers are mixed in their views about this process. A much safer way to handle a nonresponse situation is to treat the nonresponse as missing data in the analysis. Statistical software programs can handle this either question-by-question or by deleting the respondent with a missing value from all analyses. Also, the researcher can choose to eliminate a respondent from the data set if there is too much missing data. Yet another way is to simply assign the group's mean value to the missing items. Or, when an item is missing from a multi-item measure, a respondent's mean value for the rest of the items can be used for the missing value.

Another issue that can arise is how to deal with outliers (Fink, 2003, pp. 22–23). Outliers are respondents whose answers appear to be inconsistent with the rest of the data set. An easy way to check for outliers is by running frequency analyses, or counts, of responses to questions. Regression analysis also can be used to detect outliers. This is discussed in Chapter 13. Outliers can be discarded from the analysis, but one must be careful to not throw out important and useful information as would be the case when the outliers belong to a unique category of respondents heretofore unidentified. If an outlier is retained then it may be best to use the median rather than the mean as the measure of central tendency when such a measure is part of the analysis.

Short of having two or more coders create the data file independently of each other and then assessing intercoder reliability, there is not much that can be done to prevent coder error except to impress upon coders the necessity of accurate data entry. Multiple coders can be very time-consuming and costly, particularly for large data files. Each error that is identified should be traced back to the questionnaire to determine the proper code. The cleaning process is complete when either the data file has been edited to correct the errors or the corrections have been made in the analysis program.

Tabulation: Basic Analysis

Tabulation may be thought of as the final step in the data collection process and the first step in the analytical process. Tabulation is simply the counting of the number of responses in each data category (often a single column of the data matrix contains the responses to all categories).

The most basic is the simple tabulation, often called the marginal tabulation and familiar to all students of elementary statistics as the frequency distribution. A simple tabulation or distribution consists of a count of the number of responses that occur in each of the data categories that comprise a variable. A cross-tabulation involves the simultaneous counting of the number of observations that occur in each of the data categories of two or more variables. An example is given in Table 11.2. We shall examine the use of cross-tabulations in detail later in the chapter. A cross-tabulation is one of the more commonly employed and useful forms of tabulation for analytical purposes.

The flexibility and ease of conducting computer analysis increases the importance of planning the tabulation analysis. There is a common tendency for the researcher to decide that, because cross-tabulations (and correlations) are so easily obtained, large numbers of tabulations should be run. Not only is this methodologically unsound, but in commercial applications it is often costly in analyst time as well. For 50 variables, for example, there are 1,225 different two-variable cross-tabulations that can be made. Only a few of these are potentially of interest in a typical study.

Table 11.2 Cross-Tabulation: Energy Drink Purchases by Income Classes of Respondents*

	Number of liters purchased				
Income Class	*Zero*	*One*	*Two*	*Three or more*	*Total*
Less than $15,000	160	25	15	0	**200**
$15,000 - $34,999	120	15	10	5	**150**
$35,000 - $54,999	60	20	15	5	**100**
$55,000 - $74,999	5	10	5	5	**25**
$75,000 and over	5	5	5	10	**25**
Total	**350**	**75**	**50**	**25**	**500**

**Hypothetical Data*

FORMULATING HYPOTHESES

As a beginning point in the discussion of hypotheses testing, we ask: what is a hypothesis? A hypothesis is *an assertion that variables (measured concepts) are related in a specific way such that this relationship explains certain facts or phenomena*. From a practical standpoint, hypotheses may be developed to solve a problem, answer a question, or imply a possible course of action. Outcomes are predicted if a specific course of action is followed. Hypotheses must be empirically testable. A hypothesis is often stated as a research question when reporting either the purpose of the investigation or the findings. The hypothesis may be stated informally as a research question, or more formally as an alternative hypothesis, or in a testable form known as a *null hypothesis.* The null hypothesis makes a statement that no difference exists (see Pyrczak, 1995, pp. 75-84).

Research questions state in layman's terms the purpose of the research, the variables of interest, and the relationships to be examined. Research questions are not empirically testable, but aid in the important task of directing and focusing the research effort. To illustrate, a sample research question is developed in the following scenario:

EXHIBIT 11.1 Development of a Research Question for Mingles

Mingles is an exclusive restaurant specializing in seafood prepared with a light Italian flair. Barbara C., the owner and manager, has attempted to create an airy contemporary atmosphere that is conducive to conversation and dining enjoyment. In the first three months, business has grown to about 70 percent of capacity during dinner hours.

Barbara wants to track customer satisfaction with the Mingles concept, the quality of the service, and the value of the food for the price paid. To implement the survey, a questionnaire was developed using a five-point expectations scale with items scaled as values from -2 to +2. The questionnaire asks, among other things:

"How would you rate the value of Mingles food for the price paid"?

The response format for the five answers appeared as:

Much Worse Than Expected	A Little Worse Than Expected	About Average Than Expected	A Little Better Than Expected	Much Better Than Expected
○	○	○	○	○

When tabulated, the average response was found to be +0.89 with a sample standard deviation of 1.43. The research question asks if Mingles is perceived as being better than average when considering the price and value of the food.

Additional questions measure customer satisfaction by addressing "How satisfied Mingles customers are with the concept, service, food, and value".

Null Hypotheses (H_0) are statements identifying relationships that are statistically testable and can be shown not to hold (nullified). The logic of the null hypothesis is that if we hypothesize no difference, and we "reject" the hypotheses if a difference is found. If, however we confirm that no difference exists, then we "tentatively accept" the null hypothesis. We may only accept the null on a "tentative" basis because another testing of the null hypothesis using a new sample may reveal that sampling error was present and that the null hypothesis should be rejected.

For example, to compare the population and the sample, the null hypothesis might be: "There is no difference between the perceived price-value of Mingles food and what is expected on average. In this example, the difference between the population average, which is assumed to be the middle scale value of 0 = "about average" and the sample's mean evaluation of Mingles can be tested using the z distribution.

A null hypothesis may also be used to specify other types of relationships that are being tested, such as the difference between two groups, or the ability of a specific variable to predict a phenomenon such as sales or repeat business. Two examples:

1: <u>Comparing two sample groups</u>: H_0: There is no difference in the value of the food for the price paid as perceived by first time patrons and repeat patrons. This is tested by a *t*-test of the difference in means between two patron groups.

2: <u>Predicting intention to return to Mingles</u>: H_0: The perceived quality of service is not related to the likelihood of returning to Mingles. This is a regression analysis problem that uses quality of service to predict likelihood of returning to Mingles.

Alternative hypotheses may be considered to be the opposite of the null hypotheses. The alternative hypothesis makes a formal statement of expected difference, and may state simply that a difference exists or that a directional difference exists, depending upon how the null hypothesis is stated. Because population differences may exist, even if not verified by the current sample data, the alternative form is considered to be empirically non-testable. The relationship between hypothesis and research questions is summarized in Table 11.3.

Table 11.3 Hypotheses and Research Questions

	Purpose	***Example***	***Decision***
Research Question	Express the purpose of the research	What is the perception of Mingles customers regarding the price-value of the food?	None used
Alternative Hypothesis	The alternative hypothesis states the specific nature of the hypothesized relationship. i.e., that there is a difference. The alternative hypothesis is the opposite of the null hypothesis. The alternative hypothesis cannot be falsified because a relationship hypothesized to exist may not have been verified, but may in truth exist in another sample. (You can never reject an alternative hypothesis unless you test the population on all possible samples.	Mingles is perceived as having superior food value for the price when compared to the average evaluation.	Not tested because we cannot reject. We may only accept that a relationship exists.
Null Hypothesis	The null hypothesis is testable in the sense that the hypothesized lack of relationship can be tested. If a relationship is found, the null hypothesis is rejected. The Null hypothesis states that there is no difference between groups (with respect to some variable) or that a given variable does not predict or otherwise explain an observed phenomena, effect or trend.	There is no difference in perceived food value for the price for Mingles and the average evaluation.	We may reject a null hypothesis (Find a relationship). We may only tentatively accept that no relationship exists.

The objectives and hypotheses of the study should be stated as clearly as possible and agreed upon at the outset. Objectives and hypotheses shape and mold the study; they determine the kinds of questions to be asked, the measurement scales for the data to be collected, and the kinds of analyses that will be necessary. However, a project will usually turn up new hypotheses, regardless of the rigor with which it was planned and developed. New hypotheses are continually suggested as the project progresses from data collection through the final interpretation of the findings.

In Chapter 2 it was pointed out that when the scientific method is strictly followed, hypothesis formulation must precede the collection of data. This means that according to the rules for proper scientific inquiry, data suggesting a new hypothesis should *not* be used to test it. New data must be collected prior to testing a new hypothesis.

In contrast to the strict procedures of the scientific method, where hypotheses formulation must precede the collection of data, actual research projects almost always formulate and test new hypotheses *during* the project. It is both acceptable and desirable to expand the analysis to examine new hypotheses to the extent that the data permit. At one extreme, it may be possible to show that the new hypotheses are not supported by the data and that no further investigation should be considered. At the other extreme, a hypothesis may be supported by both the specific variables tested and by other relationships that give similar interpretation. The converging results from these separate parts of the analysis strengthen the case that the hypothesized relationship is correct. Between these extremes of nonsupport-support are outcomes of indeterminacy: the new hypothesis is neither supported nor rejected by the data. Even this result may indicate the need for an additional collection of information.

In a position yet more extreme from scientific method, Selvin and Stuart (1966) convincingly argue that in survey research, it is rarely possible to formulate precise hypotheses independently of the data. This means that most survey research is essentially exploratory in nature. Rather than having a single pre-designated hypothesis in mind, the analyst often works with many diffuse variables that provide a slightly different approach and perspective on the situation and problem. The added cost of an extra question is so low that the same survey can be used to investigate many problems without increasing the total cost. However, researchers must resist the syndrome of "just one more question". Often, the one more question escalates into many more questions of the type "it would be nice to know", which can be unrelated to the research objectives.

In a typical survey project, the analyst may alternate between searching the data (analyzing) and formulating hypotheses. Obviously, there are exceptions to all general rules and phenomena. Selvin and Stuart (1966), therefore, designate three practices of survey analysts:

1. Snooping. The process of searching through a body of data and looking at many relations in order to find those worth testing (that is, there are no pre-designated hypotheses)

2. Fishing. The process of using the data to choose which of a number of pre-designated variables to include in an explanatory model

3. Hunting. The process of testing from the data all of a pre-designated set of hypotheses

This investigative approach is reasonable for basic research but may not be practical for decisional research. Time and resource pressures seem to require that directed problem solving be the focus of decision research. Rarely can the decision maker afford the luxury of dredging through the data to find all of the relationships that must be present. Again, it simply reduces to the question of cost versus value.

MAKING INFERENCES

Testing hypotheses is the broad objective that underlies all decisional research. Sometimes the population as a whole can be measured and profiled in its entirety. Often, however, we cannot measure everyone in the population but instead must estimate the population using a sample of respondents drawn from the population. In this case we estimate the population "parameters" using the sample "statistics". Thus, in both estimation and hypothesis testing, *inferences* are made about the population of interest on the basis of information from a sample.

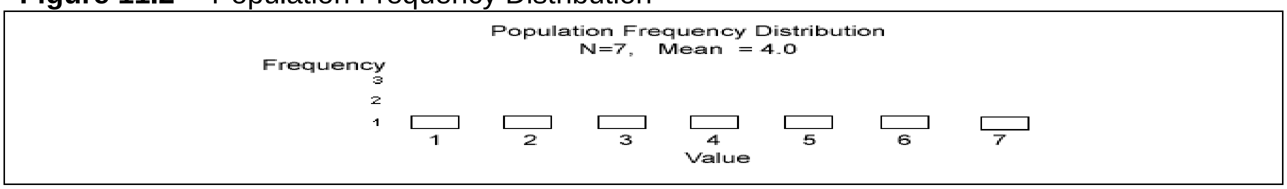
We often will make inferences about the nature of the population and ask a multitude of questions, such as: Does the sample's mean satisfaction differ from the mean of the population of all restaurant patrons? Does the magnitude of observed the differences between categories indicate that actual differences exist, or are they the result of random variations in the sample?

In other studies, it may be sufficient to simply estimate the value of certain parameters of the population, such as the amount of our product used per household, the proportion of stores carrying our brand, or the preferences of housewives concerning alternative styles or package designs of a new product. Even in these cases, however, we would want to know about the underlying associated variables that influence preference, purchase, or use (color, ease of opening, accuracy in dispensing the desired quantity, comfort in handling, etc.), and if not for purposes of the immediate problem, then for solving later problems. In yet other case studies, it might be necessary to analyze the relationships between the enabling or situational variables that facilitate or cause behavior. Knowledge of these relationships will enhance the ability to make reliable predictions, when decisions involve changes in controllable variables.

The Relationship Between a Population, a Sampling Distribution, and a Sample

In order to simplify the example, suppose there is a population consisting of only five persons. On a specific topic, these five persons have a range of opinions that are measured on a 7-point scale ranging from very strongly agree to very strongly disagree. The frequency distribution of the population is shown in the bar chart of Figure 11.2.

Figure 11.2 Population Frequency Distribution



The parameters describe this population as having a mean of μ = 4 and standard deviation = 2.

$$\sigma_x = \sqrt{\frac{\sum_{i=1} (x_i - \mu)^2}{N}} = 2$$

Now that we know the “parameters” of the population, we will consider the *sampling distribution* for our example data. Assume for a moment that like most populations, ours is so large that we are not able to measure all persons in this population, but must rely instead on a sample. In our example, the population is not large, and we will assume a sample of size n = 2.

The sampling distribution is the distribution if sample means from all possible samples of size n=2. The sampling distribution of means and standard errors are shown in Table 11.4.

Table 11.4 Computation of Sampling Distribution, Mean, and Standard Error

All possible sample distributions	*Sample mean, ($\bar{x}_i$)*	*(Error, ($\bar{x}_i$–μ)*	*Standard Error ($\bar{x}_i$–μ)^2*
1,2	1.5	–2.5	6.25
1,3	2	–2	4
1,4	2.5	–1.5	2.25
1,5	3	–1	1
1,6	3.5	–0.5	0.25
1,7	4	0	0
2,3	2.5	–1.5	2.25
2,4	3	–1	1
2,5	3.5	–0.5	0.25
2,6	4	0	0
2,7	4.5	0.5	0.25
3,4	3.5	–0.5	0.25
3,5	4	0	0
3,6	4.5	0.5	0.25
3,7	5	1	1
4,5	4.5	0.5	0.25
4,6	5	1	1
4,7	5.5	1.5	2.25
5,6	5.5	1.5	2.25
5,7	6	2	4
6,7	6.5	2.5	6.25

The mean of all possible two-member sample means is

$$\mu_{\bar{x}} = \frac{84}{21} = 4$$

and summing the standard errors of the mean for the sampling distribution gives

$$\sum (\bar{x}_i - \mu_{\bar{x}})^2 = 35.00$$

which gives a standard deviation of

$$\sigma_{\bar{x}} = \frac{\sqrt{\sum (\bar{x}_i - \mu_{\bar{x}})^2}}{N} = \frac{\sqrt{35.00}}{21} = 1.29$$

Understand that the sampling distribution becomes more normal as the sample size increases and that even in this simple case, we observe a somewhat normal shape. Also understand that the population mean $\mu = 4$ is always equal to the mean of the sampling distribution of all possible sample means ($\mu_{\bar{X}} = 4$).

The Relationship Between the Sample and the Sampling Distribution

When we draw a sample, we rarely know anything about the population, including its shape, μ, or σ. We must, therefore compute statistics from the sample ($\bar{X}$ and s) and make inferences about the population μ, and σ using the sample information.

Suppose we were to repeatedly draw samples of n = 2 (without replacement). The relevant statistics for the first of these samples having the values of (1,2) are:

$$\text{Sample mean } = \sum x_i / n = (1+2)/2 = 1.5$$

$$\text{Sample std. dev.} = s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(1-1.5)^2 + (2-1.5)^2}{(2-1)}} = .71$$

$$\text{Est. std. error} = S_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{.71}{\sqrt{2}} = .5$$

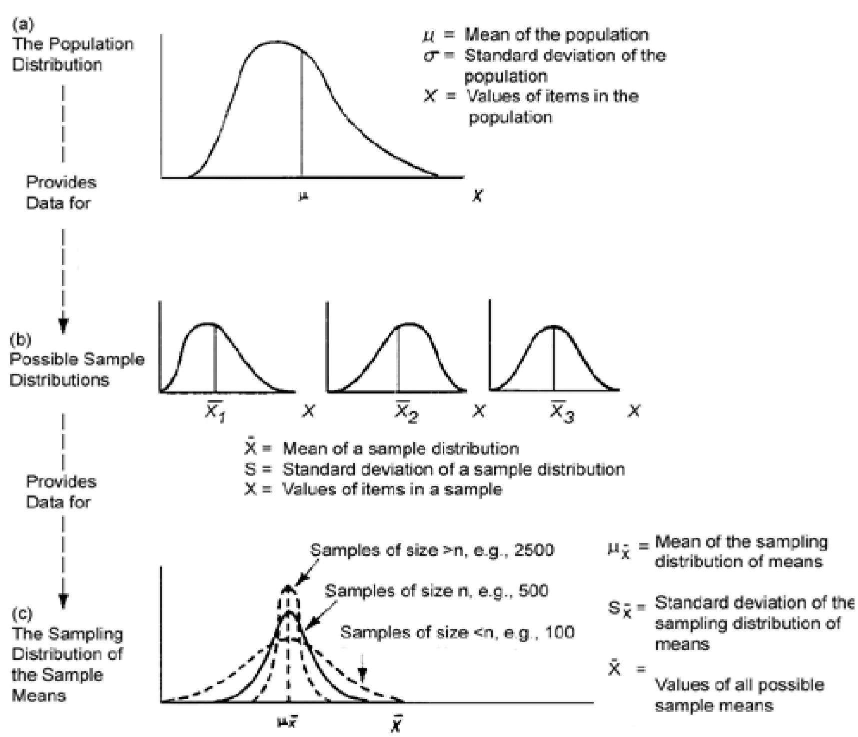
Given this $\bar{x}$ and $S_{\bar{X}}$, we can now estimate with a given probability, the intervals that give a range of possible values that could include μ, the population mean. For this single sample, they are:

68% = 1.5 ± 6.31 (.5) or -1.655 to 4.655
95% = 1.5 ± 12.71 (.5) or -4.855 to 7.855
99% = 1.5 ± 31.82 (.5) or -14.41 to 17.41

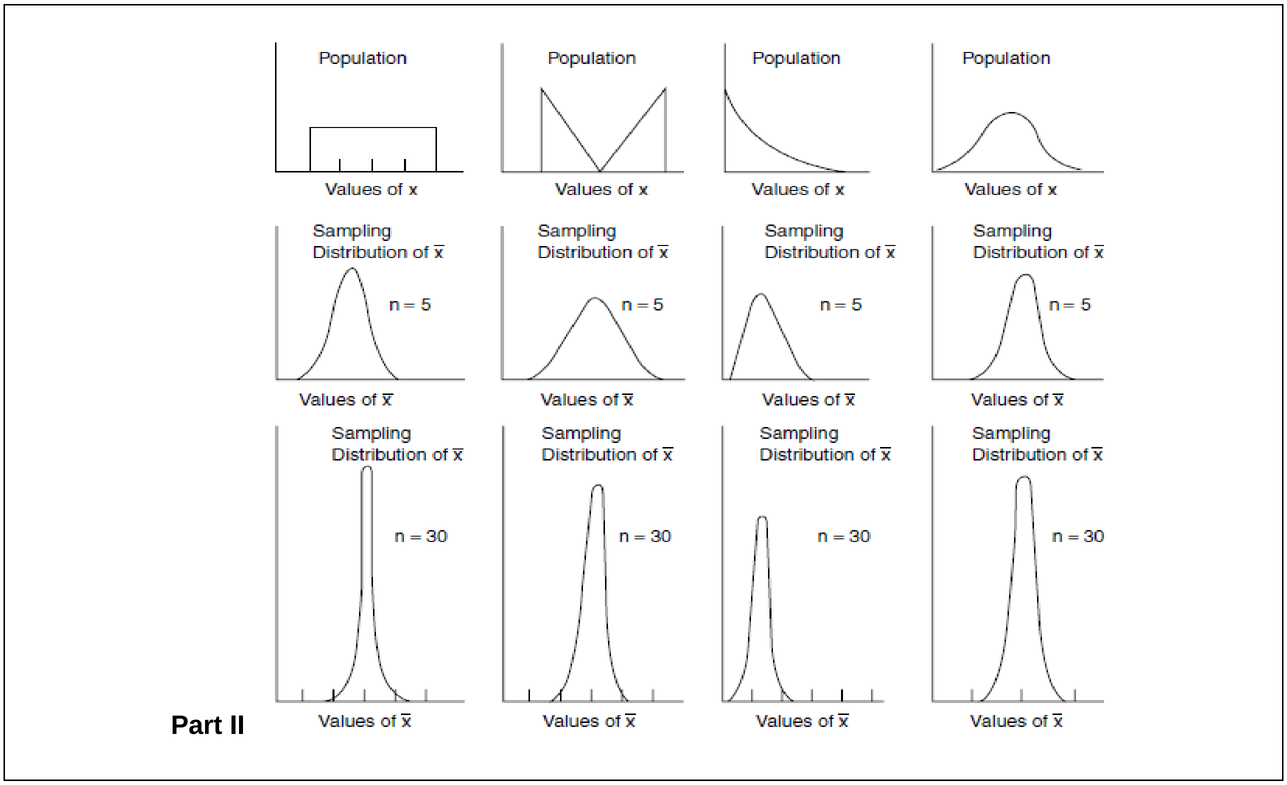
Thus, we could state that we are 99% confident that the population mean would fall within the interval –14.41 to 17.41. We note that this range is very wide and does include our population mean of 4.0. The size of the range is large because the small sample size (n=2). As the sample size increases, the numbers become larger, and gradually approximate a standard normal distribution.

The above discussion is summarized graphically in Exhibit 11.2. Part I of Exhibit 11.2 shows the relationship between the population, the sample, and the sampling distribution while Part II illustrates the impact of sample size on the shape of the sampling distribution for differently shaped population distributions.

EXHIBIT 11.2 Population, Sample, and Sampling Distribution



Part I



Acceptable Error in Hypothesis Testing

A question that continually plagues analysts is, what significance level should be used in hypothesis testing? The significance level refers to the amount of error we are willing to accept in our decisions that are based on the hypothesis test. Hypotheses testing involves specifying the value α, which is the allowable amount of Type I error.

In hypothesis testing the sample results sometimes lead us to reject H_0 when it is true. This is a Type I error. On other occasions the sample findings may lead us to accept H_0 when it is false. This is a Type II error. The nature of these errors is shown in Exhibit 11.3.

The amount of type I error, α, we are willing to accept should be set after considering (a) how much it costs to make such an error and (b) the decision rule used. A ”classic" paper dealing with criteria relevant for this question is Labovitz (1968). There are substantial differences in costs of errors in research conducted for exploratory purposes and research conducted to make a decision where large financial investments are made. The acceptable levels of significance (error) used in a basic research project may be entirely inappropriate for a managerial decision dealing with the same problem, but where millions of dollars, or a company’s market strategy is involved, managerial decisions are rarely simple or based on a single piece of research.

A tradition of conservatism exists in basic research and has resulted in the practice of keeping the Type I error at a low level (.05 or .01). The Type I error has been traditionally considered to be more important than the Type II error and, correspondingly, that it is more important to have a low α than a low β. The basic researcher typically assigns higher costs to a Type I than to a Type II error.

Exhibit 11.3 Types of Error in Making a Wrong Decision

There are two types of error that result from a mismatch between the conclusion of a research study and reality. In the null-hypothesis format, there are two possible research conclusions, to retain H_0 and to reject H_0. There are also two possibilities for the true situation: H_0 is true or H_0 is false.
These outcomes provide the definitions of Type I and Type II errors and the confidence level and power of the test:

1. A Type I error occurs when we incorrectly conclude that a difference exists. This is expressed as α, the probability that we will incorrectly reject H0, the null hypothesis, or sometimes called the hypothesis of no difference.

2. A Type II error occurs when we accept a null hypothesis when it is in reality false (we find no difference when a difference really does exist).

3. Confidence level: we correctly retain the null hypothesis (we could also say tentatively accept or it could not be rejected). This is equal to the area under the normal curve less the area occupied by α, the significance level.

4. The power of the test is the ability to reject the null hypothesis when it should be rejected (when false). That is, the power to not make an error. Because power increases as α becomes larger, researchers may choose an α of .10 or even .20 to increase power. Alternatively, sample size may be increased to increase power. Increasing sample size is the preferred option for most market researchers.

The four possible combinations are shown in the following diagram:

		True Situation	
		H_o is True	H_o is False
Conclusion of Research	Retain H_o:	No Error Prob. = $1-\alpha$ Confidence Level	Type II Error Prob. = β
	Reject H_o:	Type I Error Prob. = α Significance Level	No Error Prob. = $1-\beta$ Power of Test

In decisional research, the costs of an error are a direct result of the consequences of the errors. The cost of missing a market entry by not producing a product and foregoing gain (Type II error) may be even greater than the loss from producing a product when we should not (Type I error). The cost depends on the situation and the decision rule being used. Of course not all decision situations have errors leading to such consequences. In some situations, making a Type I error may lead to an opportunity cost (for example, a foregone gain) and a Type II error may create a direct loss.

In decisional research, the acceptable level of significance (i.e., the specification of a significance level or α) should be made by the client or the manager for whom the study has been conducted. This allows the assessment of risk by the person who will use the results as a basis for decisions (Semon, 2000). This means that the researcher should merely *report the level at which significance occurs*, letting the manager decide what this means. For basic research situations, a typical finding is often reported as significant or not as tested against a specified level, often .05. Again, it seems sensible that the researcher reports for each finding, the level at which significance occurs, letting the reader of a study decide the meaning of the alpha level.

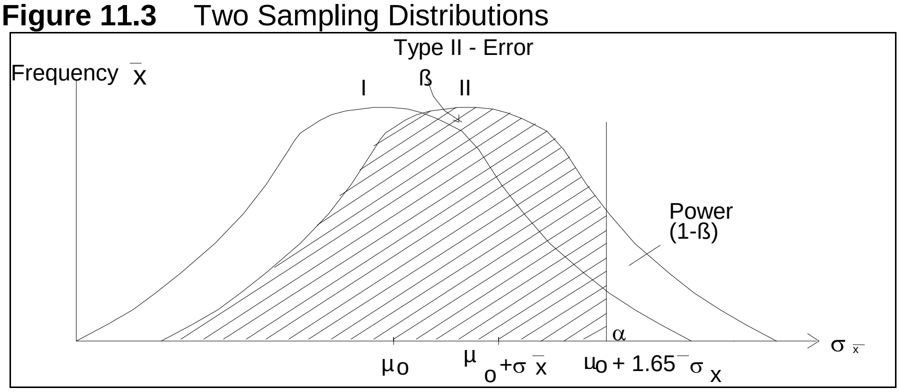
Power of a Test

The power of a hypotheses test is defined as 1-β, or 1 minus the probability of a Type II error. This means that the power of a test is the ability to reject the null hypothesis when it is false (or to find a difference when one is present).

The power of a statistical test is determined by several factors. The main factor is the acceptable amount of discrepancy between the tested hypothesis and the true situation. This can be controlled by increasing α. Power is also increased by increasing the sample size (which decreases the confidence interval).

In Figure 11.3, we observe two sampling distributions, the first having a mean μ and the second having a mean $\mu_0 + 1\alpha$. In testing the equality of the two distributions, we first identify α, which is here set to .05. (The one tail probably is .95 for 1.65σ). In Figure 11.2, we observe that for the second sampling distribution, power is the area to the right of $\mu_0 + 1.65\ \sigma$, i.e., the area in which a difference between means from the two sampling distributions exists <u>and</u> was found to exist.

Figure 11.3 Two Sampling Distributions



If σ were decreased to, say, $\mu_0 + 1.5\ \sigma$, the shaded area would move to the left, and the area defining power for curve II would increase.

The second method of increasing power is to increase sample size. Because a sample size increase directly reduces the standard error ($\sigma = \sigma / \sqrt{2}$) any increase in n decreases the absolute width of the confidence interval and narrows sampling distribution I.

In decisions that imply a greater cost for a type II error (missing a difference that does in fact exist), the analyst may want to consider expanding the probability of a type I error (σ) to .10, .15, or even .20, especially when sample sizes are small. Exhibit 11.2 discusses some ramifications of ignoring statistical power.

SELECTING TESTS OF STATISTICAL SIGNIFICANCE

Up to this point, we have considered data analysis at a descriptive level. It is now time to introduce ways to test whether the association observed is *statistically significant*. In many cases this involves testing hypotheses concerning tests of group means. These types of tests are performed on interval or ratio data using what is known as "parametric tests" and include such techniques as the F, *t*, and z tests. Often however, we have only nominal or loosely ordinal data and we are not able to meet the rigid assumptions of a parametric test. Cross tabulation analysis with the χ^2 test is often used for hypothesis testing in these situations. The χ^2 statistic is from the family of non-parametric methods.

Nonparametric methods are often called *distribution-free* methods because the inferences are based on a test statistic whose sampling distribution does not depend upon the specific distribution of the population from which the sample is drawn (Gibbons, 1993, p., 2). Thus, the methods of hypothesis testing and estimation are valid under much less restrictive assumptions than classical parametric techniques—such as independent random samples drawn from normal distributions with equal variances, interval level measurement. These techniques are appropriate for many Marketing applications where measurement is often at an ordinal or nominal level.

There are many parametric and nonparametric tests. The one that is appropriate for analyzing a set of data depends on: (1) the level of measurement of the data, (2) the number of variables that are involved, and for multiple variables, how they are assumed to be related.

As measurement scales become more restrictive and move from nominal to ordinal, and interval levels of measurement, the amount of information and power to extract that information from the scale increases. Corresponding to this spectrum of available information is an array of non-parametric and parametric statistical techniques that focus on *describing* (i.e., measures of central tendency and dispersion) and *making inferences* about the variables contained in the analysis--i.e., tests of statistical significance (as shown in Table 11.5). A useful reference for selecting an appropriate statistical technique is found in the Qualtrics Survey University.

Table 11.5 Selected Parametric and Nonparametric Univariate Analyses

	Level of Measurement		
What type of analysis do you want?	*Nominal*	*Ordinal*	*Interval (Parametric)*
Measure of central tendency	Mode	Median	Mean
Measure of Dispersion	None	Percentile	Standard Deviation
One-Sample test of statistical significance	Binomial test Π^2 one-sample test	Kolmogorov-Smirnov One-sample test One-sample runs test	t-test Z-test

EXHIBIT 11.4 IGNORING STATISTICAL POWER

Much has been written about product failure, and about the failure of advertising campaigns. But, except in rare instances, very little has been said about research failure, or research that leads to incorrect conclusions. Yet research failure can occur even when a study is based on an expertly designed questionnaire, good field work, and sophisticated analysis. The flaw may be inadequate statistical power.

Mistaking chance variation for a real difference is one risk, called Type II error and the 95% criterion largely eliminates it. By doing so, we automatically incur a high risk of Type II error-mistaking a real difference for chance variation. The ability of a sample to guard against Type II error is called statistical power.

Power was hot stuff in the 1980s. According to Semon (1990), however, but we continued to ignore statistical power, for three reasons:

1. **The concept** is more complicated than statistical significance or confidence limits.
2. **Consideration** of statistical power would often indicate that we need larger (more costly) samples than we are now using.
3. **Numerical** objectives must be specified before a research budget is fixed.

The last reason may be the worst of all. Management, even if it has a target in mind, is often unwilling to do so.

The issue really is one of the degree of sensitivity required, the answer or the question, "What is the smallest change or difference we need to measure, with a specified degree of confidence?"

Suppose we have a pricing or a package test using two monadic samples for determining which of two options to use. Suppose also that (unknown to us) Option X if superior to Y by 10 percentage points, say 30% strong buying interest vs. 20% interest.

If we use two samples of 200 each, how likely will our study identify X as being significantly superior, using a 95% significance criterion? The likelihood of that correct result is only 68%, just a shade better than 2:1, in effect a 32% risk of research failure. If we want to reduce that risk to 10% (that is, 90% statistical power), we need samples of 400 each.

Setting the desired levels of protection against the two types of error should not be a rote process. It should take into consideration the relative business risks and, therefore, requires management participation since researchers usually do not have access to enough information to make these decisions

Another way of looking at statistical power is to suggest that the researcher look at statistical insignificance rather than statistical significance. By "testing" for statistical insignificance one asks, "Is this result so likely to occur by chance that we should ignore it?" Is statistical insignificance high enough to dismiss the finding?

But, as pointed out by one observer, there is no scientific, objective way to determine what is "high enough" (Semon, 1999). For example, if the significance level associated with the difference in preference for two packages is 75%, the insignificance level is 25%. That's the same as saying the likelihood of a difference as large as observed occurring by chance is at odds of 1:3. Is that "too high?" Clearly, 1:3 odds should deserve some consideration. But, the level that such odds should have is a matter of personal preference and reflects one's risk taking philosophy.

Perhaps we should review the specs. Why do we need 95% protection against Type I error? If we lower our significance criterion to 90%, the sample size requirement drops to 300.

PARAMETRIC AND NON-PARAMETRIC ANALYSIS

Our discussion of statistical inference has emphasized that analytical procedures require assumptions about the distribution of the population of interest. For example, for normally distributed variables, we assume normality and homogeneity of variances. When a variable to be analyzed conforms to the assumptions of a given distribution, the distribution of the variable can be expressed in terms of its parameters (μ and σ). This process of making inferences from the sample to the population's parameters is called parametric analysis.

Sometimes, however, problems occur: what if we cannot assume normality, or we must question the measurement scale used. Parametric methods rely almost exclusively on either interval or ratio scaled data. In cases where data can be obtained only using ordinal or categorical scales the interpretation of the results may be questionable especially if the ordinal categories are not of equal interval. When data do not meet the rigorous assumptions of parametric method, we must rely on non-parametric methods which free us of the assumptions about the distribution.

Whereas parametric methods make inferences about parameters of the population (μ and s), non-parametric methods may be used to compare entire distributions that are based on nominal data. Other non-parametric methods that use an ordinal measurement scale test for the ordering of observations in the data set.

Problems that may be solved with parametric methods may often be solved by a non-parametric method designed to address a similar question. Often times, the researcher will find that the same conclusion regarding significance is made when data are analyzed by a parametric method and by its "corresponding" non-parametric method. We will now discuss a univariate parametric and non-parametric analyses. In the next chapter bi-variate parametric and non-parametric analyses are presented. Additional non-parametric analyses are presented in the appendix to Chapter 12.

Univariate Analyses of Parametric Data

Marketing researchers are often concerned with estimating parameters of a population. In addition, many studies go beyond estimation and compare population parameters by testing hypotheses about differences between them. Very often, the means, proportions and variances are the summary measures of concern. Our concern at this point is with differences between the means and proportions of the sample and those of the population as a whole. These comparisons involve a single variable. In the following sections, we will demonstrate three important concepts: (1) how to construct and interpret a confidence interval; (2) how to perform a hypothesis test, and (3) how to determine the power of a hypothesis test. These issues are discussed in more depth by Mohr (1990).

The Confidence Interval

The concept of a *confidence interval* is central to all parametric hypothesis testing. The confidence interval is *a range of values with a given probability (.95, .99, etc.) of covering the true population parameter.*

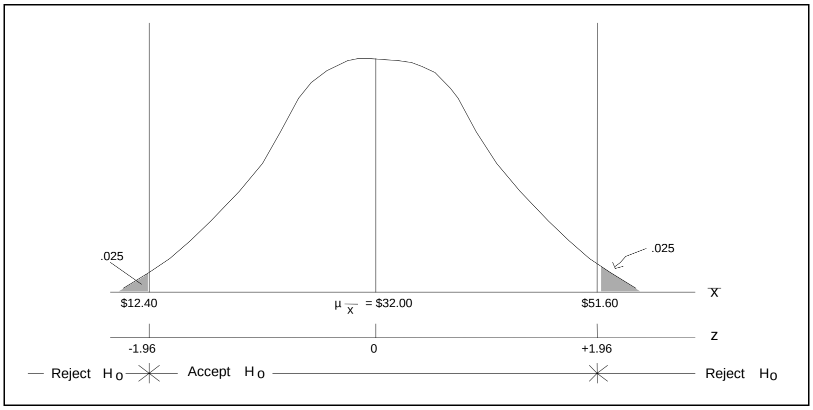
For example, assume we have a normally distributed population with population mean μ and a known population variance σ^2. Suppose we sample one item from the population, X. This single item is an estimate of μ, the population mean. Further, because the single item has been drawn randomly from a normally distributed population, the possible distribution of x values is the same as the population. This normal distribution permits us to estimate the probability associated with various intervals of values of X. For example, $p(-1.96 \leq z \geq +1.96) = .95$. Or, about 95% of the area under the normal probability curve is within this range. The confidence interval shows both the z values and the values that are included in the confidence interval. We compute this range for the sample problem that follows.

Suppose that it is a well-known fact that the average supermarket expenditure on laundry and paper products is normally distributed, with a mean of μ=$32.00 per month, and the known population standard deviation is σ =10.00. The 95% confidence interval about the population mean is computed as:

$$\text{Lower limit} \le \text{population mean} \ge \text{upper limit}$$
$$\mu - 1.96\sigma \le \mu \ge \mu + 1.96\,\sigma$$
$$32 - (1.96)(10) \le 32 \ge 32 + (1.96)(10)$$
$$32 - 19.6 \le 32 \ge 32 + 19.6$$
$$12.4 \le 32 \ge 51.6$$

Thus, we expect that 95% of all household expenditures will fall within this range, as shown in Figure 11.5.

Figure 11.5 Sampling Distribution of the Mean ($u_{\bar{x}}$)



If we expand this analysis to construct a confidence interval around the mean of a sample rather than the mean of a population we must rely on a sampling distribution to define our normal distribution. The sampling distribution is defined by the means of all possible samples of size n. Recall that the population mean μ is also the mean of the normally distributed sampling distribution, whereas $\mu \pm z\sigma$ describes the confidence interval for a population, the value ($\pm\ \alpha$ $s/\sqrt{n}$)) describes the confidence interval for the sampling distribution. This is the probability that this specified area around the sample mean covers the population mean. It is interesting to note that because n, the sample size, is included in the computation of the "standard error," we may estimate the population mean with any desired degree of precision, simply by having a large enough sample size.

Univariate Hypothesis Testing of Means

Where the Population Variance is Known

Researchers often desire to test a sample mean to determine if it is the same as the population mean. The z statistic describes probabilities of the normal distribution and is the appropriate tool to test the difference between μ, the mean of the sampling distribution, and, the sample mean when the population variance is known. The z statistic may, however, be used only when the following conditions are met:

(1) Individual items in the sample must be drawn in a random manner.
(2) The population must be normally distributed. If this is not the case, the sample must be large (>30), so that the sampling distribution is normally distributed.
(3) The data must be at least interval scaled.
(4) The variance of the population must be known.

1. When these conditions are met, or can at least be reasonably assumed to exist, the traditional hypothesis testing approach is as follows:

   (1) The null hypothesis (H_0) is specified that there is no difference between μ and $\bar{x}$. Any observed difference is due solely to sample variation.
   (2) The alpha risk (Type I error) is established (usually .05).
   (3) The z value is calculated by the appropriate z formula:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

   (4) The probability of the observed difference having occurred by chance is determined from a table of the normal distribution (Appendix B, Table B-1).
   (5) If the probability of the observed differences having occurred by chance is greater than the alpha used, then H_0 cannot be rejected and it is concluded that the sample mean is drawn from a sampling distribution of the population having mean μ.

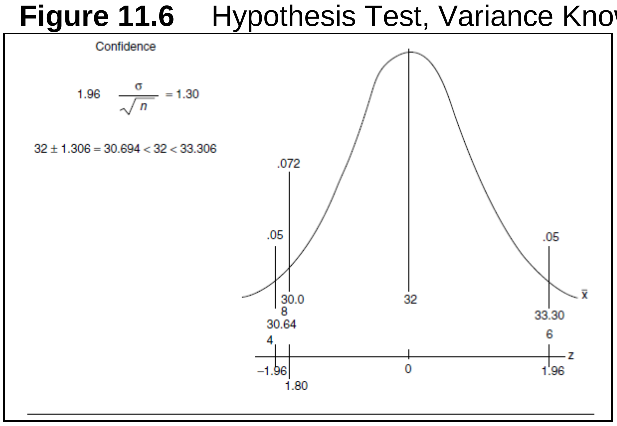
Returning to our supermarket example, suppose now that a random sample of 225 respondents was collected. The sample has a mean of =$30.80 and the population standard deviation is known to be $10.00. We want to know if the population mean μ=$32.00 equals the sample mean =30.80, given sample variation.

Steps:

1. All assumptions are met
2. $H_0 : \mu = \bar{x}$
3. $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{30.80 - 32.00}{10/\sqrt{225}} = -1.20/0.667 = -1.80$
4. $P(z = 1.80) = .928$
   $1 - P(z = 1.80) = .072$
5. Decision: Do not reject

The results are shown in Figure 11.6.

Figure 11.6 Hypothesis Test, Variance Known



Population Variance is Unknown

Researchers rarely know the true variance of the population, and must therefore rely on an estimate of σ^2, namely, the sample variance s^2. With this variance estimate, we compute the t statistic.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Let's assume that for the supermarket example, we do not know the population values, but we again want to test if the average monthly purchase of laundry and paper goods equals \$32.00, when the sample of 225 shoppers shows a mean, $\bar{x}$, of \$30.80 and a standard deviations, of \$9.00. The t statistic is computed as:

$$t = \frac{30.80 - 32.00}{9/\sqrt{225}} = \frac{-1.20}{9/15} = -2.00$$

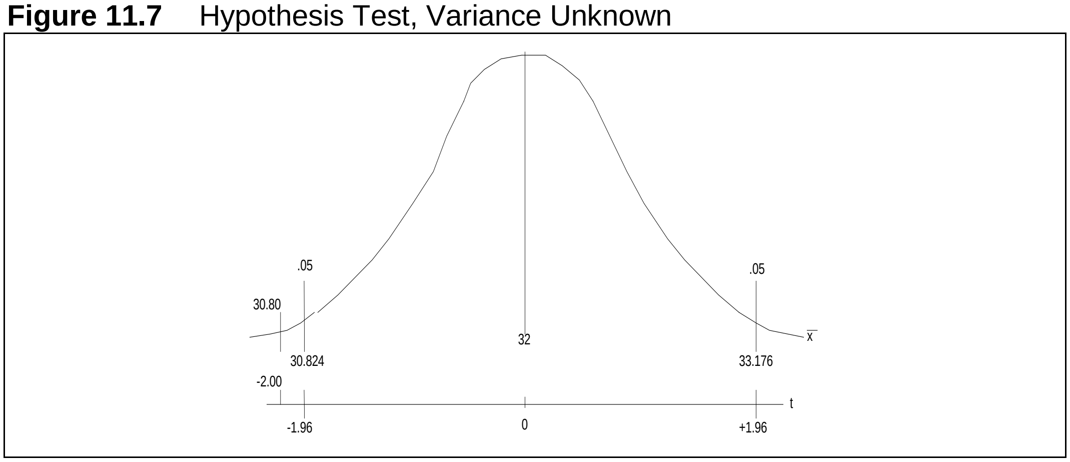
With a probability: P (t=2.00, df = *n*-1=224) = .951 (2 tailed test)
1-P (t=2.00) = .049

Confidence Interval:

$1.96\ s/\sqrt{n}) = 1.176$
$32 \pm 1.176 = 30.824 + 32 \dotplus 33.176$

We therefore reject the null hypothesis. Figure 11.7 shows these results.

Figure 11.7 Hypothesis Test, Variance Unknown



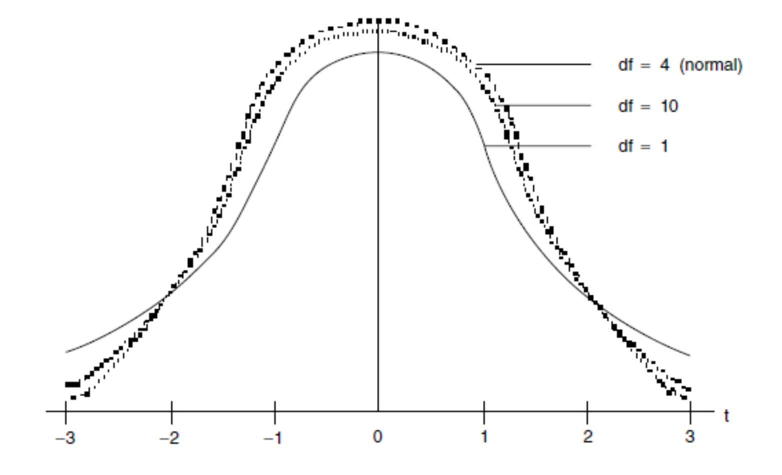
A question may arise why we reject the null hypothesis for the t-test but do not reject it for z-test. The answer is relatively simple. In the z-test, the numerator is a known constant $\sigma/\sqrt{n}$). It is the same regardless of the sample that is drawn, as long as the sample is of size n. The denominator of the t-test varies with the sample variance. Because the variance of the sample was less than the variance of the population (9 vs. 10), the size of the confidence interval was reduced, allowing us to reject the null hypothesis.

Unlike the z-distribution, the t is a family of distributions, each having a different shape, depending on the degrees of freedom (see Figure 11.8). The probability for a given value of t varies across distributions.

The appropriate t-distribution to use in an analysis is determined by the available "degrees of freedom." Unfortunately, the concept of degrees of freedom is inadequately defined. Many statisticians use the term to describe the number of values that are free to vary. A second definition is that “degrees of freedom” is a measure of how much precision an estimate of variation has. A general rule is that the degrees of freedom decrease when more parameters have to be estimated. This concept seems to something that “we all know what it is, but cannot precisely define it.” It seems reasonable to accept the meaning in terms of “values that vary” by specifying that degrees of freedom refers to a mathematical property of a distribution related to the number of values in a sample that can be *freely specified* once one knows something about the sample.

In univariate analyses, the available degrees of freedom are n-1. We lose one degree of freedom for each population parameter that we estimate (μ). To explain further, suppose we have a sample with size $n = 5$ in which the mean value of some measure is calculated to be zero. If we select at random any 4 numbers from this sample, say -6, -4, 4, and 9, we know that the last number is completely determined and must be -3. Each time we estimate a parameter we lose one degree of freedom, while n - 1 are free to estimate that parameter.

Figure 11.8 The t-distributions for *df* = 1, *df* = 4 and *df* = 10



As n becomes large ($\geq$ 30) the curve is approximately normal, with true normality reached at the limit ($n = \infty$). The t-statistic is widely used in both univariate and bi-variate market research analyses, due to the relaxed assumptions over the z-statistic, as follows:

(1) Individual items in the sample are drawn at random.
(2) The population must be normally distributed. If not, the sample must be large (>30).
(3) The data must be at least interval scaled.
(4) The population variance is not known exactly, but is estimated by the variance of the sample.

Univariate Analysis of Categorical Data: The Chi-Square Goodness of Fit Test

Chi-square analysis (χ^2) can be used when the data identifies the number of times or frequency that each category of a tabulation or cross tabulation appears. Chi-square is a useful technique for testing the following relationships:

1. Determining the significance of sample deviations from an assumed theoretical distribution; that is, does a certain model fit the data. This is typically called a goodness-of-fit test.

2. Determining the significance of the observed associations found in the cross tabulation of two or more variables. This is typically called a test of independence. (Discussed in Chapter 12).

The procedure involved in chi-square analysis is quite simple. We compare the observed (frequency) data with another set of "data" based on a set of theoretical frequencies. These theoretical frequencies may result from application of some specific model of the phenomenon being investigated--relationship 1 above. Or we might specify that the frequency of occurrence of two or more characteristics is mutually independent--relationship 2 above.

In either case we compute a measure (chi-square) of the variation between actual and theoretical frequencies, under the null hypothesis that there is no difference between the model and the observed frequencies. We may say that the model fits the facts. If the measure of variation is "high," we reject the null hypothesis at some specified alpha risk. If the measure is "low," we accept the null hypothesis that the model's output is in agreement with the actual frequencies.

Single Classification

Suppose that we are interested in how frequently a sample of respondents selects each of two test packages. Test package A contains an attached coupon involving so many cents off on a subsequent purchase of the brand by the respondent. Test package B is the same package but, rather than an attached coupon, it contains an attached premium (a ballpoint pen) that the respondent may keep. The packages are presented simultaneously to a sample of 100 respondents and each respondent is asked to choose one of the two packages.

The frequency of choice is presented in the first column (labeled "Observed") of Table 11.6. We see that 63 respondents out of 100 select package A. Suppose that the researcher believes the "true" probability of selecting A versus B to be 50-50 and that the observed 63-37 split reflects sampling variation. Given that the researcher's model would predict an estimated frequency of 50-50, are the observed frequencies compatible with this theoretical prediction? In chi-square analysis we set up the null hypothesis that the observed (sample) frequencies are consistent with those expected under application of the model.

We use the following notation. Assume that there are k categories and a random sample of n observations; each observation must fall into one and only one category. The observed frequencies are

$$f_i = (i = 1, 2, \ldots, k); \sum_{i=1}^{k} f_i = n$$

The theoretical frequencies are:

$$F_i = (i = 1, 2, \ldots, k); \sum_{i=1}^{k} F_i = n$$

In the problem above,

$$f_1=63,\ f_2=37,\ F_1=50,\ F_2=50,\ n=100$$

We compute the chi-square statistic:

$$X^2 = \sum_{i=1}^{k} \frac{(f_i - F_i)^2}{F_i}$$

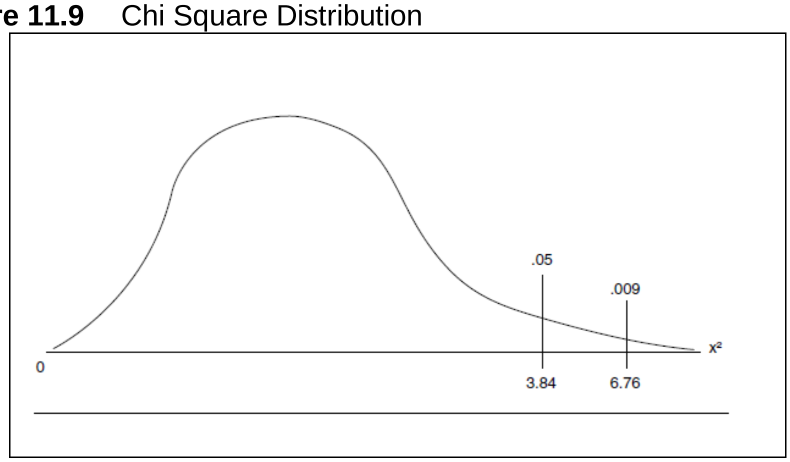
Table 11.6 Observed vs. Theoretical Frequencies (Test-Package Illustration)

Package	*f_i Observed*	*F_i Predicted*	*$f_i - F_i$*	*$(f_i - F_i)^2 / F_i$*
A	63	50	13	169/50 = 3.38
B	37	50	-13	169/50 = 3.38
Total	100	100		6.67

In the above one-way classification problem, the statistic is approximately distributed as chi square with k-1 degrees of freedom. That is, we have only two categories and, hence, 1 degree of freedom. Table A.3 in the Appendix at the end of this book shows the appropriate distribution.

In Table A.3 the tabular chi-square value for $\alpha = 0.05$ and k-1=1 is 3.84. If the null hypothesis is true, the probability of getting a chi-square value greater then 3.84 is 0.05. Since our computed chi-square value is 6.76 (see Table 11.6), we reject the null hypothesis that the output of the theoretical model corresponds with the observed frequencies (see Figure 11.9). In using the chi-square table we note that only k, the number of categories, is pertinent, rather than the sample size n. Sample size is important to the quality of the approximation and the power of the test. A good rule of thumb, however, is that chi-square analysis should be used only when the theoretical frequencies in each cell exceed five; otherwise, the distribution in Table A.3 will not be a good approximation. Pragmatically, the "risk" is that with a theoretical frequency less than five, a single cell's chi-square value may be unusually high and, thus, unduly influence the overall value.

Figure 11.9 Chi Square Distribution



Univariate Analysis: Test of a Proportion

The standard normal distribution may be used to test not only means, as explained above, but also differences in proportions. The univariate test of proportions, like the univariate test of means, compares the population proportion to the proportion observed in the sample. For a sample proportion, p,

$$Z = \frac{p - \pi}{S_p}$$

Where s_p, the estimated standard error of the proportion,

$$S_p = \sqrt{pq/n} = \sqrt{\frac{p(1-p)}{n}}$$

z = standard normal value
p = the sample proportion of successes
q = (1-p) = the sample proportion of failures
n = sample size

In a simple example, suppose the marketing manager of a snack food company is evaluating a new snack. 225 respondents are surveyed at a local shopping mall. The survey indicates that 87% are favorable toward the snack. The manager needs a 90% favorability rate. Is it safe to say that this is simply sampling variation?

$$Z = 1.338 = \frac{.87 - .90}{.022} \quad \text{where} \quad S_p = \sqrt{\frac{(.87)(.13)}{225}} = .022$$

given $\alpha = .05$ with z = 1.96, we cannot reject the null hypothesis.

$$CI = .90 - 1.96(.022) < 90 < 90 + 1.96(.022)$$

$$CI = .856 < .90 < .944$$

In this example the manager could (in a statistical sense) claim that 87% approval is no different that 90% approval with sampling variation. In terms of corporate policy, however, set cut points may be more rigidly held.

SUMMARY

Chapter 11 has introduced the basic concepts of formulating hypothesis testing and making statistical inference in the context of univariate analysis. In actual research, the analyst may alternate between analyzing the data and formulating hypotheses.

A hypothesis is a statement that variables (measured constructs) are related in a specific way. The null hypothesis, H_0, is a statement that no relationship exists between the variables tested or that there is no difference.

Statistics are based on making inferences from the sample of respondents to the population of all respondents by means of a sampling distribution. The sampling distribution is a distribution of the parameter values (means or variances) that are estimated when all possible samples are collected.

When testing hypotheses, the analyst may correctly identify a relationship as present or absent, or may commit one of two types of errors. A Type I Error occurs when a true H_0 is rejected (there is no difference, but we find there is). A Type II Error occurs when we accept a false H_0 (there is a difference, but we find that none exists). The power of a test was explained as the ability to reject H_0 when it should be rejected.

Selecting the appropriate statistical technique for investigating a given relationship depends of the level of measurement (nominal, ordinal, interval) and the number of variables to be analyzed. The choice of parametric vs. non-parametric analyses depends on the analyst's willingness to accept the distributional assumptions of normality and homogeneity of variances.

Finally, univariate hypothesis testing was demonstrated using the standard normal distribution statistic (z) to compare a mean and proportion to the population values. The t-test was demonstrated as a parametric test for populations of unknown variance and samples of small size. The chi-square goodness of fit test was demonstrated as a non-parametric test of nominal data that make no distributional assumptions. The observed frequencies were compared to an expected distribution.

In this chapter, we focused on *univariate analyses*. Chapter 12 expands this discussion to include non-parametric and parametric analyses involving two variables. In Chapter 12, we also focus on bivariate analyses involving Measures of Association. Finally, in Chapters 13 and 14, we focus on multivariate analyses, involving dependence, and interdependence of associative data.

In each chapter we will observe that the appropriate statistical technique is selected, in part based on the number of variable and their relationship, when included in the analysis. However, without higher levels of measurement and their associated characteristics of central tendency dispersion and rates of change, the ability to investigate these more complex relationships is not possible.

REFERENCES

Andrews, F.M., Klem, L., Davidson, T.N., O'Malley, P.M., and Rodgers, W.L. (1981), *A Guide for Selecting Statistical Techniques for Analyzing Social Science Data*. Ann Arbor, MI: Institute for Social Research, University of Michigan, Second Edition

Fink, A. (2003). *How to Manage, Analyze, and Interpret Survey Data*. Thousand Oaks, CA: Sage Publications, Second Edition

Gibbons, J.D. (1993). Nonparametric Statistics: An Introduction. Newbury Park, CA: Sage Publications

Labovitz, S. (1968). "Criteria for Selecting A Significance Level: A Note on the Sacredness of .05," The American Sociologist, 3 (August), 220-222.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction*. Thousand Oaks, CA: Sage.

Mohr, L.B. (1990). Understanding Significance Testing. Newbury Park, CA: Sage Publications

Pyrczak, F. (1995). Making Sense of Statistics: A Conceptual Overview. Los Angeles: Pyrczak Publishing

Selvin, H. and Stuart, A. (1966). "Data-Dredging Procedures in Survey Analysis," The American Statistician, June, 20-23

Semon, T.T. (2000). "When Determining Significance, Let Clients Decide the Risks," Marketing News October 9, 15

Semon, T.T. (1999). "Consider A Statistical Insignificance Test," Marketing News, 33, 3 (February 1), 9

Semon, T.T. (1990). :Keep Ignoring Statistical Power at Your Own Risk," Marketing news, 24, 181 (September 30), 21

Chapter 12
BIVARIATE DATA ANALYSIS

Analysis can be viewed as the categorization, aggregation, and the manipulation of data to obtain answers to research question underlying the research project. The final aspect of analysis is interpretation. The process of interpretation involves taking the results of analysis, making inferences relevant to the research relationships studied, and drawing managerially useful conclusions about these relationships.

BASIC CONCEPTS OF ANALYZING ASSOCIATIVE DATA

Our chapter begins with a brief discussion of cross-tabulation to mark the beginning of a major topic of this chapter—the analysis of associative data. A large part of the rest of the chapter will focus on methods for analyzing how the variation of one variable is associated with variation in other variables.

The computation of row or column percentages in the presentation of cross-tabulations is taken up first. We then show how various insights can be obtained as one goes beyond two variables in a cross-tabulation to three (or more) variables. In particular, examples are presented of how the introduction of a third variable can often refine or explain the observed association between the first two variables.

Bivariate Cross-Tabulation

Bivariate cross-tabulation represents the simplest form of associative data analysis. At the minimum we can start out with only two variables, such as occupation and education, each of which has a discrete set of exclusive and exhaustive categories. Data of this type is called categorical, since each variable is assumed to be nominal-scaled. Bivariate cross-tabulation is widely used in marketing applications to analyze variables at all levels of measurement. In fact, it is the single most widely used bivariate technique in applied settings. Reasons for the continued popularity of bivariate cross-tabulation include the following (Feick, 1984, p. 376):

1. It provides a means of data display and analysis that is clearly interpretable even to the less statistically inclined researcher or manager.
2. A series of bivariate tabulations can provide clear insights into complex marketing phenomena that might be lost in a simultaneous analysis with many variables.
3. The clarity of interpretation affords a more readily constructed link between market research and market action.
4. Bivariate cross-tabulations may lessen the problems of sparse cell values that often plague the interpretation of discrete multivariate analyses (bivariate cross-tabulations require the expected number of respondents in any table cell be 5).

The entities being cross-classified are usually people, objects, or events. The cross-tabulation, at its simplest, consists of a simple count of the number of entities that fall into each of the possible categories of the cross-classification. Excellent discussions of ways to analyze cross-tabulations can be found in Hellevik (1984) and Zeisel (1957).

However, we usually want to do more than show the raw frequency data. At the very least, row or column percentages (or both) are usually computed.

Percentages

The simple mechanics of calculating percentages are known to all of us. In cross-tabulation, percentages serve as a relative measure indicating the relative size of two or more categories.

The ease and simplicity of calculation, the general understanding of its purpose, and the near universal applicability have made the percent statistic, or its counterpart the proportion, the most widely used statistical tool in marketing research. Yet its simplicity of calculation is sometimes deceptive, and the understanding of its purpose is frequently insufficient to ensure sound application and interpretation. The result is that the percent statistic is often the source of misrepresentations, either inadvertent or intentional.

The sources of problems in using percentages are largely in identifying the direction in which percentages should be computed, and in knowing how to interpret percentage of change.

Both these problems can be illustrated by a small numerical example that uses a before and after with control experimental design. Let us assume that KEN's Original, a small regional manufacturer of salad dressings, is interested in testing the effectiveness of spot TV ads in increasing consumer awareness of a new brand—called Life. Two geographic areas are chosen for the test: (1) test area A and (2) control area B. The test area receives five 15-second television spots per week over an eight-week period, whereas the control area receives no spot TV ads at all. (Other forms of advertising were equal in each area.)

Assume that four independent samples (of telephone interviews) were conducted as before and after tests in each of the areas. Respondents were asked to state all the brands of salad dressing they could think of, on an aided basis. If Life was mentioned, it was assumed that this constituted consumer awareness of the brand. However, as it turned out, sample sizes differed across all four sets of interviews. This common occurrence in surveys (i.e., variation in sample sizes) increases the value of computing percentages.

Table 12.1 shows a set of frequency tables that were compiled before and after a TV ad for Life Salad Dressing was aired. Interpretation of Table 12.1 would be hampered if the data were expressed as raw frequencies and different percentage bases were reported. Accordingly, Table 12.1 shows the data, with percentages based on column and row totals. Can you see why the row percentages are more useful for analytical purposes?

Direction in Which to Compute Percentages

In examining the relationship between two variables, it is often clear from the context that one variable is more or less the independent or control variable and the other is the dependent or criterion variable. In cases where this distinction is clear, the rule is to compare percentages within levels of the dependent variable.

In Table 12.1, the control variable is the experimental area (test versus control) and the dependent variable is awareness. When comparing awareness in the test and control areas, row percentages are preferred. We note that before the spot TV campaign the percentage of respondents who are aware of Life is almost the same between test and control areas: 42 percent and 40 percent, respectively.

However, after the campaign the test-area awareness level moves up to 66 percent, whereas the control-area awareness (42 percent) stays almost the same. The small increase of 2 percentage points reflects either sampling variability or the effect of other factors that might be serving to increase awareness of Life in the control area.

On the other hand, computing percentages across the independent variable (column percent) makes little sense. We note that 61 percent of the aware group (before the spot TV campaign) originates from the test area; however, this is mainly a reflection of the differences in total sample sizes between test and control areas.

After the campaign we note that the percentage of aware respondents in the control area is only 33 percent, versus 39 percent before the campaign. This may be erroneously interpreted as indicating that spot TV in the test area depressed awareness in the control area. But we know this to be false from our earlier examination of raw percentages.

It is not always the case that one variable is clearly the independent or control variable and the other is the dependent or criterion variable. This should pose no particular problem as long as we agree, for analysis purposes, which variable is to be considered the control variable. Indeed, cases often arise in which each of the variables in turn serves as the independent and dependent variable.

Table 12.1 Aware of Life Salad Dressing—Before and After Spot TV

	Before Spot TV				**After Spot TV**		
	Aware	**Not Aware**	**Total**	**Area**	**Aware**	**Not Aware**	**Total Area**
Test Area				**Test Area**			
Freq.	250	350	600	Freq.	330	170	550
Col %	61%	59%	60%	Col %	67%	44%	57%
Row %	42%	58%		Row %	66%	34%	
Control Area				**Control Area**			
Freq.	160	240	400	Freq.	160	220	380
Col %	39%	41%	40%	Col %	33%	56%	43%
Row %	40%	60%		Row %	42%	58%	
Total	410	590	1000	Total	490	390	880
Before TV Spot	41%	59%	100%	After TV Spot	56%	44%	100%

Interpretation of the Percentage Change

A second problem that arises in the use of percentages in cross-tabulations is the choice of which method to use in measuring differences in percentages. There are three principal ways to portray percentage change:

1. The absolute difference in percentages
2. The relative difference in percentages
3. The percentage of possible change in percentages

The same example can be used to illustrate the three methods.

Absolute Percentage Increase

Table 12.2 shows the percentage of respondents who are aware of Life before and after the spot TV campaign in the test and control areas. First, we note that the test-area respondents displayed a greater absolute increase in awareness. The increase for the test-area respondents was 24 percentage points, whereas the control-area awareness increased by only 2 percentage points.

Table 12.2 Aware of Life—Percentages Before and After the Spot TV Campaign

	Before the Campaign	*After the Campaign*
Test area	42%	66%
Control area	40%	42%

Relative Percentage Increase

The relative increase in percentage is [(66 – 42)]/42] × 100 = 57 percent and [(42 – 40)/40] × 100 = 5 percent, respectively, for test- and control-area respondents.

Percentage Possible Increase

The percentage of possible increase for the test area is computed by first noting that the maximum percentage-point increase that could have occurred is 100 – 42 = 58 points. The increase actually registered is 24 percentage points, or 100(24/58) = 41 percent of the maximum possible. That of the control area is 100(2/60) = 3 percent of the maximum possible.

In terms of the illustrative problem, all three methods give consistent results in the sense that the awareness level in the test area undergoes greater change than that in the control area. However, in other situations conflicts among the measures may arise.

The absolute-difference method is simple to use and requires only that the distinction between percentage and percentage points be understood. The relative-difference method can be misleading, particularly if the base for computing the percentage change is small. The percentage-of-possible-difference method takes cognizance of the greater difficulty associated with obtaining increases in awareness as the difference between potential-level and realized-level decreases. In some studies all three measures are used, inasmuch as they emphasize different aspects of the relationship.

Introducing a Third Variable into the Analysis

Cross-tabulation analysis to investigate relationships need not stop with two variables. Often much can be learned about the original two-variable association through the introduction of a third variable that may refine or explain the original relationship. In some cases, it may show that the two variables are related even though no apparent relationship exists before the third variable is introduced. These ideas are most easily explained by example.

Consider the hypothetical situation facing MCX Company, a specialist in telecommunications equipment for the residential market. The company has recently test-marketed a new service for online recording of cable or satellite programs without a storage box. Several months after the introduction, a telephone survey was taken in which respondents in the test area were asked whether they had adopted the innovation. The total number of respondents interviewed was 600 (Table 11.3).

Table 12.3 Adoption—Percentage by Gender and Age

	Men			*Women*		
Frequency	*<35 Yrs*	*≥35 Yrs*	*Total %*	*<35 Yrs*	*≥35 Yrs*	*Total %*
Adopters						
Number of Cases	100	60	160	11	9	20
Column %	50%	30%	40%	11%	9%	10%
Row %	62.5%	37.5%		55%	45%	
Non Adopters						
Number of Cases	100	140	240	89	91	180
Column %	50%	70%	60%	89%	91%	90%
Row %	41.7%	58.3%		49.4%	50.6%	
Total	200	200	400	100	100	200
Percentage	50%	50%	100%	50%	50%	100%

One of the variables of major interest in this study was the age of the respondent. Based on earlier studies of the residential market, it appeared that adopters of the firm's new products tended to be less than 35 years old. Accordingly, the market analyst decides to cross-tabulate adoption and respondent age. Respondents are classified into the categories "under 35 years" (<35) and "equal to or greater than 35 years" (≥35) and then cross-classified by adoption or not. Table 12.3 shows the full three-variable cross-tabulation. It seems that the total sample of 600 is split evenly between those who are under 35 years of age and those who are 35 years of age or older. Younger respondents display a higher percentage of adoption (37 percent = 100 + 11)/300) than older respondents (23 percent = (60 + 9)/300).

Analysis and Interpretation

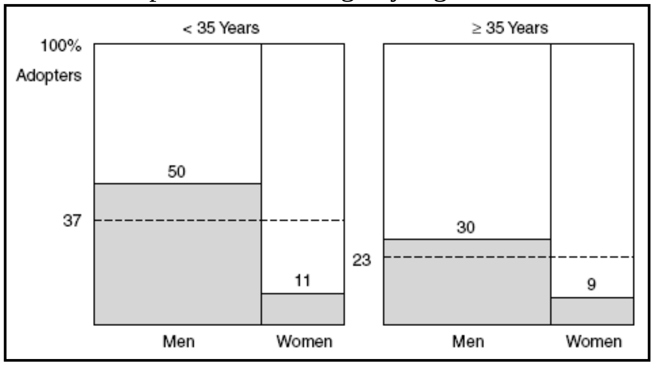
The researcher is primarily interested in whether this finding differs when gender of the respondent is introduced into the analysis. As it turned out, 400 respondents in the total sample were men, whereas 200 were women.

Table 12.3 shows the results of introducing gender as a third classificatory variable. In the case of men, 50 percent of the younger men adopt compared with only 30 percent of the older men. In the case of women, the percentages of adoption are much closer. Even here, however, younger women show a slightly higher percentage of adoption (11 percent) than older women (9 percent).

The effect of gender on the original association between adoption and age is to refine that association without changing its basic character; younger respondents show a higher incidence of adoption than older respondents. However, what can now be said is the following: If the respondent is a man, the differential effect of age on adoption is much more pronounced than if the respondent is a woman.

This pattern is even easier to identify when we show this information graphically (Figure 12.1). The height of the bars within each rectangle represents the percentage of respondents who are adopters. The relative width of the bars denotes the relative size of the categories—men versus women—representing the third variable, gender. The shaded portions of the bars denote the percentage adopting by gender, the dashed line represents the weighted average percentage adopting by gender, and the dashed line represents the weighted average percentage across the genders.

Figure 12.1 Adoption—Percentage by Age and Gender



It is easy to see from the figure that adoption differs by age group (37 percent versus 23 percent). Furthermore, the size of the difference depends on the gender of the respondent: Men display a relatively higher rate of adoption, compared with women, in the younger age category.

Recapitulation

Representatives of three-variable association can involve many possibilities that could be illustrated by the preceding adoption-age-gender example:

1. In the example presented, adoption and age exhibit initial association; this association is still maintained in the aggregate but is refined by the introduction of the third variable, gender.

2. Adoption and age do not appear to be associated. However, adding and controlling on the third variable, gender, reveals suppressed association between the first two variables within the separate categories of men and women. In the two-variable cases, men and women exhibit opposite patterns, canceling each other out.

Although the preceding example was contrived to illustrate the concepts, the results are not unusual in practice. It goes almost without saying that the introduction of a third variable can often be useful in the interpretation of two-variable cross-tabulations.

However, the reader should be aware of the fact that we have deliberately used the phrase *associated with* rather than *caused by.* Association of two or more variables does not imply causation, and this statement is true regardless of our preceding efforts to refine some observed two-variable association through the introduction of a third variable.

In principle, of course, we could cross-tabulate four or even more variables with the possibility of obtaining further insight into lower-order (e.g., two-variable) associations. However, somewhere along the line, a problem arises in maintaining an adequate cell size for all categories. Unless sample sizes are extremely large in the aggregate and the number of categories per variable is relatively small, cross-tabulations rarely can deal with more than three variables at a time. A further problem, independent of sample size, concerns the high degree of complexity of interpretation that is introduced by cross-tabulations involving four or more variables. In practice, most routine applications of cross-tabulation involve only two variables at a time.

As noted in Table 12.3, there are definite advantages associated with having a two category criterion variable, such as adoption versus non-adoption. In many applications, however, the criterion variable will have more than two categories. Cross-tabulations can still be prepared in the usual manner, although they become somewhat more tedious to examine.

BIVARIATE ANALYSIS: DIFFERENCE BETWEEN SAMPLE GROUPS

Marketing activities largely focus on the identification and description of *market segments*. These segments may be defined demographically, attitudinally, by the quantity of the product used, by activities participated in or interests, by opinions, or by a multitude of other measures. The key component of each of these variables is the ability to group respondents into market segments. Often this segmentation analysis involves the identification of differences and the asking of questions about the marketing implications of those differences: Do differences in satisfaction exist for the two or more groups that are defined by age categories?

Bivariate statistical analysis refers to the analysis of relationships between two variables. These analyses are often differences between respondent groups. In the following discussion, we explore bivariate statistical analysis and focus on the two-variable case as a bridge between the comparatively simple analyses already discussed and the more sophisticated techniques that will command our attention in later chapters. We begin with what is perhaps the most used test of market researchers: cross-tabulation. Next, we consider analysis of differences in group means. First, we discuss the *t*-test of differences in means of two independent samples, and then we look at one-way analysis of variance (ANOVA) that tests for differences in means for *k* groups. Finally, we provide a discussion of some of the more widely used nonparametric techniques. These are but a few of the possible parametric and nonparametric analyses that could be discussed (see Table 12.4).

Bivariate Cross-Tabulation

The chi-square statistic in a contingency table analysis is used to determine if the observed associations between the variables in the cross-tabulation are statistically significant.

Often called *chi-square analysis*, this technique is used when the data consist of counts or frequencies within categories of a tabulation or cross-tabulation table. In conjunction with the cross-tabulation we will introduce the chi-square statistic, χ^2, to determine the significance of observed association in cross-tabulations involving two or more variables. This is typically called a X^2 test of independence.

Cross-tabulation represents the simplest form of associative data analysis. At the minimum we can start out with a bivariate cross-tabulation of two variables, such as occupation and education, each of which identifies a set of exclusive and exhaustive categories. We know that such data are called *categorical* because each variable is assumed to be only nominal-scaled. Bivariate cross-tabulation is widely used in marketing applications to analyze variables at all levels of measurement. In fact, it is the single most widely used bivariate technique in applied settings.

Table 12.4 Selected Nonparametric Statistical Tests for Two-Sample Cases

What type of analysis do you want?	*Level of Measurement*		
	Nominal	*Ordinal*	*Interval / Ratio*
Two-sample related samples	McNemar test for the significance of changes	Sign test Wilcoxon matched-pairs signed-ranks test	
Independent samples	Fischer exact probability test X^2 test for two independent samples	Median test Mann-Whitney U test Kolmogorov-Smirnov two-sample test Wald-Wolfowitz runs test	t-test One-way ANOVA

In marketing research, observations may be cross-classified, as when we are interested in testing whether occupational status is associated with brand loyalty. Suppose, for illustrative purposes, that a marketing researcher has assembled data on brand loyalty and occupational status—white collar, blue collar, and unemployed or retired—that describes consumers of a particular product class. The data for our hypothetical problem appear in Table 12.5.

A total of four columns, known as *banner points*, are shown. Four rows or *stubs* are also shown. Professional cross-tabulation software will output multiple side by side tables that join multiple variables on the column banner points such that loyalty and another variable such as usage occasion could be analyzed simultaneously.

In a study of 230 customers, we are interested in determining if occupational category may be associated with the characteristic loyalty status. The data suggests that a relationship exists, but is the observed association a reflection of sampling variation, or can we conclude that a true relationship exists?

Table 12.5 Contingency Table of Observed versus Theoretical Frequencies

N, Expected n *X^2 Contribution* *Occupation*	*Highly Loyal*	*Moderately Loyal*	*Brand Switchers*	*Total Number (percent distribution)*
White-collar	30 (30.5) .01	42 (34.1) 1.86	18 (25.4) 2.17	90 (39.1%)
Blue Collar	14 (22.1) 2.93	20(24.5) .86	31 (18.4) 8.68	65 (28.3%)
Unemployed/ retired	34 (25.4) 2.88	25 (28.4) .40	16 (21.2) 1.27	75 (32.6%)
Total	78 (33.9%)	87 (37.8%)	65 (28.3%)	230 x^2=21.08

In analyzing the problem by means of chi-square, we make use of the marginal totals (column and row totals) in computing theoretical frequencies given that we hypothesize independence (no relationship) between the attributes loyalty status and occupational status. For example, we note from Table 12.5 that 33.9 percent (78/320) of the respondents are highly loyal. If possession of this characteristic is independent of occupational status, we would expect that 33.9 percent (78/320) of the 90 respondents classified as white-collar workers (i.e., 30.5) would be highly loyal. Similarly, 37.8

percent (87/320) of the 90 = 34.1 would be moderately loyal, and 28.3 percent (65/320) of the 90 = 25.4 would be brand switchers. In a similar fashion we can compute theoretical frequencies for each cell on the null hypothesis that loyalty status is statistically independent of occupational status. (It should be noted that the frequencies are the same, whether we start with the percentage of the row or the percentage of the column.)

The theoretical frequencies (under the null hypothesis) are computed and appear in parentheses in Table 12.5. The chi-square statistic is then calculated (and shown in the table) for each of the data cells in the table using the observed and theoretical frequencies:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - F_i)^2}{F_i}$$

where f_i = actual observed frequency, F_i = theoretical expected frequency, and k = number of cells (r x c).

The appropriate number of degrees of freedom to use in this example is four. In general, if we have R rows and C columns, the degrees of freedom associated with the chi-square statistic are equal to the product

$$(R - 1)(C - 1)$$

If we use a significance level of 0.05, the probability table value of chi-square is 11.488. Hence, because the computed χ^2 of 21.08 is greater than the table value of 11.488, we reject the null hypothesis of independence between the characteristics loyalty status and occupational status.

A correction factor must be applied to the formula for chi-square when the number of observations in a cell is less than 10 (or where a 2 × 2 contingency table is involved),. The numerator within the summation sign becomes $(\,|\,f_i - F_i\,| - 1/2)^2$ where the value 1/2 is the *Yates continuity correction*. This correction factor adjusts for the use of a continuous distribution to estimate probability in a discrete distribution.

Chi-square analysis of independence can be extended to deal with more than two variables. No new principles are involved. Three characteristics of the technique should be borne in mind, however. First, chi-square analysis deals with counts (frequencies) of data. If the data are expressed in percentage form, they should be converted to absolute frequencies. Second, the technique assumes that the observations are drawn independently. Third, the chi-square statistic cannot describe the relationship; it only gauges its statistical significance, regardless of logic or sense (Semon, 1999).

To assess the nature of the relationship, the researcher must look at the table and indicate how the variables appear to be related—a type of eyeball approach. This may involve examining any of the following combinations: (a) the variable combinations that produce large χ^2 values in the cells; (b) those with a large difference between the observed and expected frequencies; or (c) those where the cell frequency count, expressed as a percentage of the row total, is most different from the total column percentage (marginal column %). When variables are ordered or loosely ordered, a pattern can sometimes be observed by marking cells with higher than expected observed frequencies with a (+) and those with lower than expected observed frequencies with a (–), or even graphing the deviations from expected values, or cell X^2 values using a 3-dimensional contour map.

Bivariate Analysis: Differences in Means and Proportions

A great amount of marketing research is concerned with estimating parameters of one or more populations. In addition, many studies go beyond estimation and compare such population parameters by testing hypotheses about differences between them. Means, proportions, and variances are often the summary measures of concern. Our concern at this point is with differences in means and proportions. Direct comparisons of variances are a special case of the more general technique of analysis of variance, which is covered later in this chapter.

Standard Error of Differences

Here we extend the topic of sampling distributions and standard errors as they apply to a single statistic - to cover differences in statistics and show a traditional hypothesis for differences.

Standard Error of Difference of Means

For two samples, A and B, that are independent and randomly selected, the standard error of the differences in means is calculated by

$$\sigma_{\bar{x}_A-\bar{x}_B} = \sqrt{\frac{\sigma_{A^2}}{n_A} + \frac{\sigma_{B^2}}{n_B}}$$

This estimate of the standard error is appropriate for use in the denominator of the *z* test formula. If the population standard deviations, σ_i, are not known, then the estimated standard error becomes

$$est.S_{\bar{x}_A-\bar{x}_B} = \sqrt{\frac{S_{A^2}}{n_A} + \frac{S_{B^2}}{n_B}}$$

For relatively small samples the correction factor $N_i/n_i–1$ is used and the resulting formula for the estimated standard error is

$$est.S_{\bar{x}_A-\bar{x}_B} = \sqrt{\frac{S_{A^2}}{n_A - 1} + \frac{S_{B^2}}{n_B - 1}}$$

Of course, these formulas would be appropriate for use in the denominator of the *t*-test.

Standard Error of Differences of Proportions

Turning now to proportions, the derivation of the standard error of the differences is somewhat similar. Specifically, for large samples,

$$est.\sigma_{pA-pB} = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

For small samples, the correction factor is applied, resulting in

$$est.\sigma_{pA-pB} = \sqrt{\frac{p_A(1-p_A)}{n_A - 1} + \frac{p_B(1-p_B)}{n_B - 1}}$$

This estimate would again be appropriate for use in the denominator of the *Z*-test of proportions.

Testing of Hypotheses

When applying the standard error formulas for hypotheses testing concerning parameters, the following conditions must be met:

1. Samples must be independent.
2. Individual items in samples must be drawn in a random manner.
3. The population being sampled must be normally distributed (or the sample must be of sufficiently large size).
4. For small samples, the population variances must be equal.
5. The data must be at least interval scaled.

When these five conditions are met, or can at least be reasonably assumed to exist, the traditional approach is as follows.

1. The null hypothesis (H_0) is specified such that there is no difference between the parameters of interest in the two populations (e.g., H_0: $\mu_A - \mu_B = 0$); any observed difference is assumed to occur solely because of sampling variation.
2. The alpha risk is established ($\alpha = .05$ or other value).
3. A *Z* value is calculated by the appropriate adaptation of the *Z* formula. For testing the difference between two means, *Z* is calculated in the following way:

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sigma_{\bar{x}_A - \bar{x}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 0}{\sigma_{\bar{x}_A - \bar{x}_B}}$$

and for proportions

$$Z = \frac{(p_A - p_B) - (\pi_A - \pi_B)}{\sigma_{pA - pB}} = \frac{(p_A - p_B) - 0}{\sigma_{pA - pB}}$$

For unknown population variance and small samples, the student *t* distribution must be used, and for a test of the differences in means, *t* is calculated from

$$t = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{S_{\bar{x}_A - \bar{x}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 0}{S_{\bar{x}_A - \bar{x}_B}}$$

4. The probability of the observed difference of the two sample statistics having occurred by chance is determined from a table of the normal distribution (or the *t* distribution, interpreted with $[n_A + n_B - 2]$ degrees of freedom).
5. If the probability of the observed difference having occurred by chance is *greater* than the alpha risk, the null hypothesis of no difference is not rejected; it is concluded that the parameters of the two universes are not significantly different. If the probability of the observed difference having occurred by chance is *less* than the alpha risk, the null hypothesis is rejected; it is concluded that the parameters of the two populations differ significantly. In an applied setting, there are times when the level at which significance occurs (the alpha level) is reported and management decides whether to accept or reject.

An example will illustrate the application of this procedure. Let us assume we have conducted a survey of detergent and paper goods purchases from supermarkets among urban (population A) and rural (population B) families (see Table 12.6).

Table 12.6 Sample Group and Average Expenditure

Family Type	*Sample*	*Average Amount Spent*	*Standard Deviation*
Urban (Sample A)	$N_A = 400$	$\bar{X}_A$ - $32.00	$S_A = \$10.00$
Rural (Sample B)	$N_B = 225$	$\bar{X}_B = \$30.80$	$S_B = \$9.00$

The question facing the researcher is the following: "Do urban families spend more on these items, or is the $1.20 difference in means caused by sampling variations?" We proceed as follows. The hypothesis of no difference in means is established. We assume the alpha risk is set at .05. Since a large sample test is called for, the *Z* value is calculated using the separate variances estimate of the standard error of differences in means:

$$S_{\bar{x}_A-\bar{x}_B} = \sqrt{\frac{(10.0)^2}{400} + \frac{(9.0)^2}{225}} = \$0.78$$

The *Z* value is then determined to be

$$Z = \frac{(32.0 - 30.8) - 0}{0.78} = +1.54$$

The probability of the observed difference in the sample means having been due to sampling is specified by finding the area under the normal curve that falls to the right of the point $Z = +1.54$. Consulting a table of the Cumulative Normal Distribution, we find this area to be $1.0 - .9382 = 0.0618$. Since this probability associated with the observed difference ($p = 0.06$) is greater than the preset alpha, a strict interpretation would be that there is no difference between the two types of families concerning the average expenditure on nonfood items. In a decision setting, however, the manager would have to determine whether this probability (0.06) is low enough to conclude, on pragmatic grounds, that the families do not differ in their behavior.

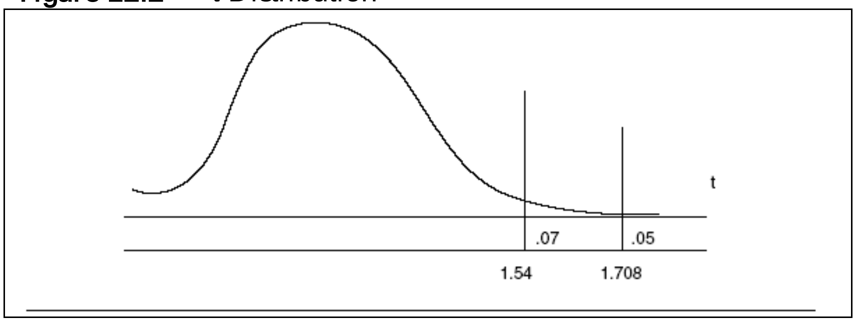
As stated previously, and discussed in the previous chapter, often there is no preset alpha and decisional considerations require that the manager determines the meaning of the reported alpha.

To illustrate the small sample case, let us assume that we obtain the same mean values and get values for s_A and s_B such that s_{xA} - s_{xB} = $0.78 from samples $n_A = 15$ and $n_B = 12$. With these data we calculate *t* as follows:

$$t = \frac{(32.0 - 30.8) - 0}{0.78} = 1.54$$

The critical value of t is obtained from a table of percentiles for the t distribution. For, say, $\alpha = .05$, we determine the critical value of t for $(n_A\ n_B - 2) = 25$ degrees of freedom to be 1.708 (one-tailed test).
Since the calculated t of $1.54 < 1.708$, we cannot reject the hypothesis of no difference in average amount spent by the two types of families. This result is shown in Figure 12.4. When samples are not independent, the same general procedure is followed. The formulas for calculating the test statistics differ, however.

Figure 12.2 *t*-Distribution



Testing the Means of Two Groups: The Independent Samples t-Test

The *t*-distribution revolutionized statistics and the ability to work with small samples. Prior statistical work was based largely on the value of *Z*, which was used to designate a point on the normal distribution where population parameters were known. For most market research applications it is difficult to justify the *Z*-test's assumed knowledge of μ and σ. The *t*-test relaxes the rigid assumptions of the *Z*-test by focusing on sample means and variances (X and s). The *t*-test is a widely used market research statistic to test for differences between two groups of respondents.

In the previous section, the t statistic was described. Most computer programs recognize two versions of this statistic. In the previous section, we presented what is called the separate variance estimate formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

This formula is appropriate where differences in large samples are tested.

The second method, called the pooled variance estimate, computes an average for the samples that is used in the denominator of the statistic:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_1^2(n_1 - 1) + S_2^2(n_2 - 1)/n_1 - 1 + n_2 - 1}}$$

The pooling of variances is a simple averaging of variances that is required when (1) testing for the same population proportion in two populations or (2) testing the difference in means between two small samples.

Table 12.7 shows a sample output of *t*-tests from SPSS. Two respondent groups were identified for the supermarket study:

1. Respondents who were males
2. Respondents who were females

The analysis by gender is shown in Part A of the output for the attitudinal dimensions, friendliness of clerks and helpfulness of clerks. It will be noted that the sample sizes are approximately equal. To show a contrast we have included in Part B of the output a *t*-test from another study where the number of males and females varies widely. In all three examples, the differences between separate-variance and pooled-variance analysis are very small. Table 12.9 also shows an *F* statistic. This is *Levene's test for equality of variances*. Using this test the researcher knows which *t*-test output to use—equal variances assumed or equal variances not assumed.

Table 12.7 Selected Output From PASW (SPSS) *t*-Test

					Equal Variances Not Not Assumed (Separate)			*Equal Variances Assumed (Pooled)*		
	Mean	*n*	*F*	*Sig*	*t*	*df*	*Sig (2-tailed)*	*t*	*df*	*Sig (2-tailed)*
Part A										
"Friendliness of clerks"			2713	.102	1.301	126.213	.19	1.298	128	.197
Male	4.55	64								
Female	4.82	66								
"Helpfulness of clerks"			15,401	.000	.308	105.177	.758	.308	126	.758
Male	4.50	64								
Female	4.56	64								
Part B										
"Shopping Experience"										
Male	1.977	44	.089	.765	.569	66,993	.571	.590	189	.556
Female	1.857	147								

Testing of Multiple Group Means: Analysis of Variance

Analysis of variance (ANOVA) is a logical extension of the independent groups *t*-test methodology. Rather than test differences between two group means, we test the overall difference in *k* group means, where the *k* groups are thought of as levels of a treatment or control variable(s) or factor(s). ANOVA is a general set of methodologies that handle many different types of research and experimental designs. Traditional use of analysis of variance has been to isolate the effects of experimental variables that are related in various experimental situations. The respondent receives a combination of the treatment levels from the different factors and the response is measured. More specifically, ANOVA is used to test the statistical significance of differences in mean responses given the introduction of one or more treatment effects.

Much experimental work has been conducted in medicine and agriculture. In pharmaceutical drug testing, positive and negative effects of dosage and formulation are measured over time and for different types of ailments and patient illness levels. The variables influencing the results are called *experimental factors*. In agriculture, crop yields are measured for plots of land, each of which receives a different treatment level that is defined by a factor or control variable. Control variables in this application might include seed type, fertilizer type, fertilizer dosage, temperature, moisture, and many other variables thought to influence production. In each of these plots, the average yield is measured and analyzed to determine the effect of the specific measured levels of the factors being evaluated. Marketing research experiments have control variables that are certainly different from agricultural experiments, but the principles are the same.

The proper planning and design of the relationships between the experimental factors may require the use of experimental design methodology such as the completely randomized, randomized block, Latin square, and factorial designs. Here we discuss the two most basic forms of the ANOVA methodology: the one-way ANOVA and the two-factor ANOVA designs (see Exhibit 12.1).

EXHIBIT 12.1 ANOVA Designs

Example: It is well known that interest ratings for TV ads are related to the advertising message. A simple one-way ANOVA to investigate this relationship might compare three messages:

Advertising message A	Advertising message B	Advertising message C
$\bar{X}_A$	$\bar{X}_B$	$\bar{X}_C$

Two-factor ANOVA includes a second factor, possibly the type of advertisement (magazine or TV):

	Message (A)	Message (B)	Message (C)
Magazine Ad (M)	$\bar{X}_{MA}$	$\bar{X}_{MB}$	$\bar{X}_{MC}$
TV Ad (T)	$\bar{X}_{TA}$	$\bar{X}_{TB}$	$\bar{X}_{TC}$

Each of the cells in this matrix would contain an average interest rating for the measures taken for the particular combination of message and media.

This brief introduction to the idea behind an ANOVA will hopefully reduce the impression that the technique is used to test for significant differences among the variances of two or more sample universes. This is not strictly the case. ANOVA is used to test the statistical significance of differences in mean responses given the introduction of one or more treatments effects.

The ANOVA Methodology

The appropriateness of the title analysis of variance comes from the method's ability to explain the variation in responses to the various treatment combinations. The methodology for explaining this variation is explained in Exhibit 12.2, which presents an example regarding responses to messages.

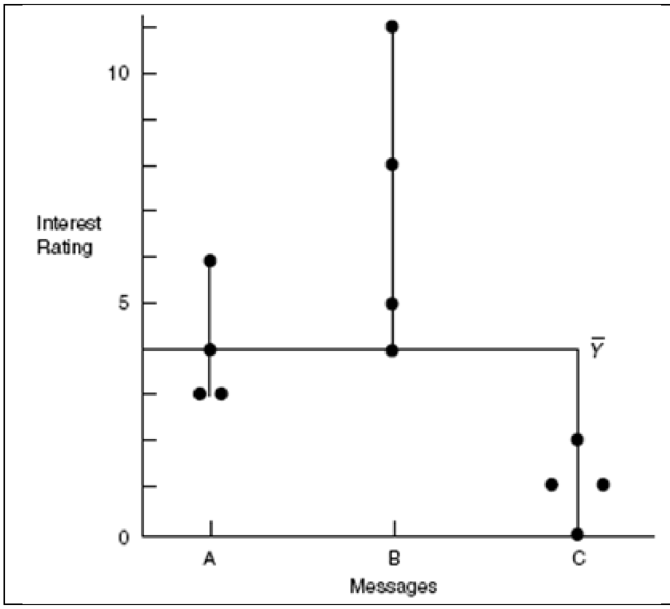
EXHIBIT 12.2 Example of ANOVA Methodology

Using the single-factor ANOVA design for the problem described in Exhibit 12.1, the actual interest rating data for 12 respondents (12 = 4 respondents for each of the 3 treatments) appears in tabular and graphical format as follows:

Obs #	*Msg. A*	*Msg. B*	*Msg. C*
1	6	8	0
2	4	11	2
3	3	4	1
4	3	5	1
Mean	4	7	1

It is apparent that the messages differ in terms of their interest ratings, but how do we perform ANOVA to determine if these differences are different statistically? There are three values that must be computed in order to analyze this pattern of values.

1. Compute the total sum of squares: The grand mean of the 12 observations is computed, followed by the variance of the individual observations from this mean.



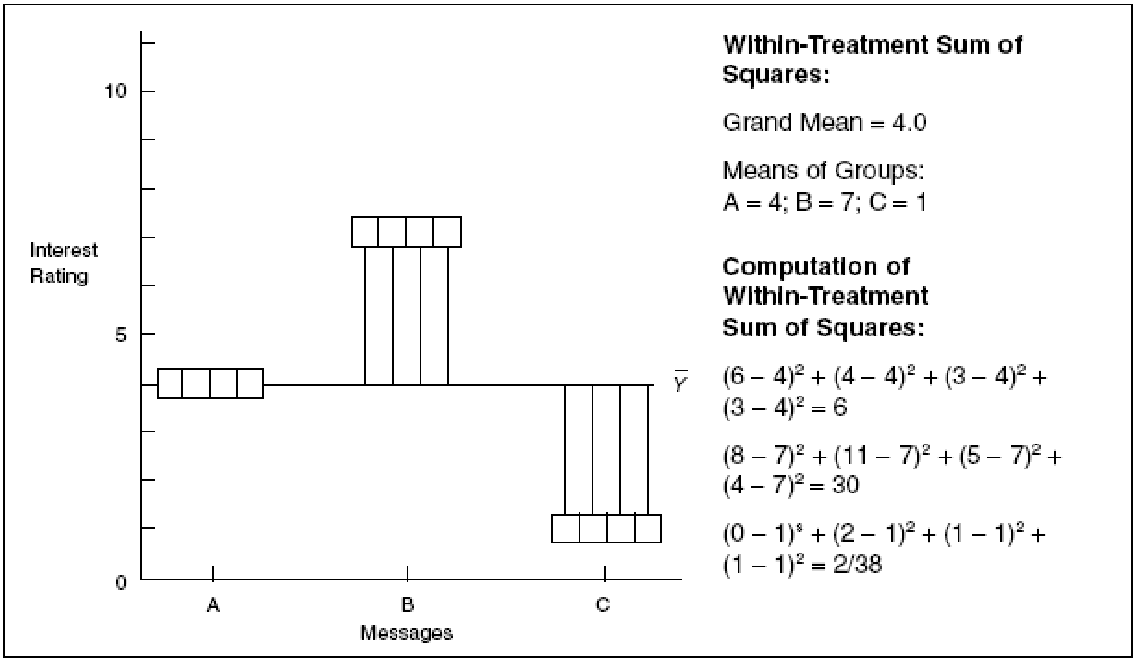
Total Sum of Squares:
Grand mean = 4.0 = $\bar{X}_G$

Computation of
Total Sum of Squares

$\Sigma(\bar{X}_{ij} - \bar{X}_G) =$
$(6\text{-}4)^2 + (4\text{-}4)^2 + \ldots + (1\text{-}4)^2 = 110$

2. Between-treatment sum of squares: The means of the factor levels (messages A, B, and C) are computed, followed by the deviation of the factor level means from the overall mean, weighted by the number of observations $\Sigma\, n(\bar{X}_j - \bar{X})^2$

3. Compute the within-treatment sum of squares: The means of the factor levels are computed, followed by the deviation of the observations within each factor level from that factor level mean.



Thus, an observation may be decomposed into three terms that are additive, and each explains a particular type of variance:

Observation = Overall mean + Deviation of the group mean from the overall mean
+ Deviation of the observation from the group mean

The overall mean is constant and common to all observations: The deviation of a group mean from the overall mean represents the effect on each observation of belonging to that particular group; the deviation of an observation from its group mean represents the effect on that observation of all variables other than the group variable.

The basic idea of ANOVA is to compare the between-treatment-groups sum of squares (after dividing by degrees of freedom to get a mean square) with the within-treatment-group sum of squares (also divided by the appropriate number of degrees of freedom). This is the *F* statistic that indicates the strength of the grouping factor. Conceptually,

$$F = \frac{\text{Sampling variance} + \text{Variance due to effect of treatment}}{\text{Sampling variance}}$$

The larger the ratio of between to within, the more we are inclined to reject the null hypothesis that the group mean $\mu_1 = \mu_2 = \mu_3$. Conversely, if the three sample means were very close to each other, the between-samples sum of squares would be close to zero and we would conclude that the population means are the same, once we consider the variability of individual cases within each sample group.

However, to make this comparison, it is necessary to assume that the error-term distribution has constant variance over all observations. This is exactly the same assumption as was made for the *t*-test.

In the next section we shall (a) use more efficient computational techniques, (b) consider the adjustment for degrees of freedom to obtain mean squares, and (c) show the case of the *F* ratio in testing significance. Still, the foregoing remarks represent the basic ANOVA idea for comparing between- with within-sample variability.

One-Way (Single Factor) Analysis of Variance

One-way ANOVA is analysis of variance in its simplest (single-factor) form. Suppose a new product manager for the hypothetical Friskie Corp. is interested in the effect of shelf height on supermarket sales of canned dog food. The product manager has been able to secure the cooperation of a store manager to run an experiment involving three levels of shelf height (knee level, waist level, and eye level) on sales of a single brand of dog food, which we shall call Snoopy. Assume further that our experiment must be conducted in a single supermarket and that our response variable will be sales, in cans, of Snoopy dog food for some appropriate unit of time. But what shall we use for our unit of time? Sales of dog food in a single store may exhibit week-to-week variation, day-to-day variation, and even hour-to-hour variation. In addition, sales of this particular brand may be influenced by the price or special promotions of competitive brands, the store management's knowledge that an experiment is going on, and other variables that we cannot control at all or would find too costly to control.

Assume that we have agreed to change the shelf-height position of Snoopy three times per day and run the experiment over eight days. We shall fill the remaining sections of our gondola with a filler brand, which is not familiar to customers in the geographical area in which the test is being conducted. Furthermore, since our primary emphasis is on explaining the technique of analysis of variance in its simplest form (analysis of one factor: shelf-height), we shall assign the shelf heights at random over the three time periods per day and not design an experiment to explicitly control and test for within-day and between-day differences. Our experimental results are shown in Table 12.8.

Here, we let X_{ij} denote the sales (in units) of Snoopy during the *i*th day under the *j*th treatment level. If we look at mean sales by each level of shelf height, it appears as though the waist-level treatment, the average response to which is $X_2 = 90.9$, results in the highest mean sales over the experimental period. However, we note that the last observation (93) under the eye-level treatment exceeds the waist-level treatment mean. Is this a fluke observation? We know that these means are, after all, *sample* means, and our interest lies in whether the three population means from which the samples are drawn are equal or not.

Now we shall show what happens when one goes through a typical one-way analysis-of variance computation for this problem.

Table 12.8 Sales of Snoopy Dog Food (in Units) by Level of Shelf Height

Shelf Height						
Knee Level		*Waist Level*		*Eye Level*		*Grand Total*
X_{11}	77	X_{12}	88	X_{13}	85	
X_{21}	82	X_{22}	94	X_{23}	85	
X_{31}	86	X_{32}	93	X_{33}	87	
X_{41}	78	X_{42}	90	X_{43}	81	
X_{51}	81	X_{52}	91	X_{53}	80	
X_{61}	86	X_{62}	94	X_{63}	79	
X_{71}	77	X_{72}	90	X_{73}	87	
X_{81}	81	X_{82}	87	X_{83}	93	
$X_T 1 = 648$		$X_T 2 = 727$		$X_T 3 = 677$		$X_{TT} = 2{,}052$
$\bar{X}_1 = 81.0$		$\bar{X}_2 = 90.9$		$\bar{X}_3 = 84.6$		$\bar{X}_{TT} = 85.5$

Table 12.9 shows the calculations for the among-treatments, within-treatments, and total sums of squares, the mean squares, and the *F* ratio. Had the experimenter used an alpha risk of 0.01, the null hypothesis of no differences among treatment levels would have been rejected. A table of *F* ratios is found in most statistics and research texts.

Table 12.9 Analysis of Variance – Snoopy Dog Food Experiment

Source of Variation	*Degrees of freedom*	*Sum of squares*	*Mean square*	*F Ratio*
Among treatments	$t - 1 = 2$	399.3	199.7	14.6 ($p < 0.01$)
Within treatments	$n - t = 21$	288.7	13.7	
Total	$n - 1 = 23$	688.0	$MS_T = SSt / df_T$ $MS_W = SSw / df_W$	$F = \dfrac{MS_T}{MS_W}$

Correction factor

$$C = \frac{(X_{TT})^2}{n} = \frac{(2,052)^2}{24} = 175,446.0$$

Total sum of squares

$$\sum X_{ij}^2 - C = (77)^2 + (82)^2 + \cdots + (87)^2 + (93)^2 - 175,466.0 = 688.0$$

Treatment sum of squares

$$\frac{\sum X_{ij}^2}{n_j} - C = \frac{(648)^2 + (727)^2 + (677)^2}{8} - 175,466.0 = 399.3$$

Within treatment sum of squares

$$\sum X_y^2 \frac{\sum X_{Tj}^2}{n_j} = (77)^2 + (82)^2 + \cdots + (87)^2 + (93)^2 - \frac{(648)^2 + (727)^2 + (677)^2}{8} = 288.7$$

Note that Table 12.9 shows shortcut procedures for finding each sum of squares. For example, the total sum of squares is given by

$$\sum X_{ij}^2 - \frac{(X_{TT})^2}{n} = 688.0$$

This is the same quantity that would be obtained by subtracting the grand mean of 85.5 from each original observation, squaring the result, and adding up the 24 squared deviations. This mean-corrected sum of squares is equivalent to the type of formula used earlier in this chapter.

The interpretation of this analysis is, like the *t*-test, a process of comparing the *F* value of 14.6 (2, 21 df) with the table value of 4.32 ($p = .05$). Because ($14.6 > 4.32$), we reject the null hypothesis that treatments 1, 2, and 3 have equivalent appeals.

The important consideration to remember is that, aside from the statistical assumptions underlying the analysis of variance, the *variance of the error distribution* will markedly influence the significance of the results. That is, if the variance is *large* relative to differences among treatments, then the true effects may be swamped, leading to an acceptance of the null hypothesis when it is false. As we know, an increase in sample size can reduce this experimental error. Though beyond the scope of this chapter, *specialized* experimental designs are available, the objectives of which are to increase the efficiency of the experiment by reducing the error variance.

Follow-up Tests of Treatment Differences

The question that now must be answered is the following: Which treatments differ? The *F*-ratio only provides information that differences exist. The question of where differences exist is answered by follow-up analyses, usually a series of independent samples *t*-tests, that compare the treatment level combinations ((1,2), (1,3), and (2,3)). Because of our previous discussion of the *t*-test, we will not discuss these tests in detail. We will only allude to the fact that there are various forms of the *t*-statistic that may be used when conducting a series of two group tests. These test statistics (which include techniques known as the LSD (Least Significant Difference), Bonnferoni's test, Duncan's multiple range tests, Scheffe's test, and others) control the cumulative probability that a Type I error will occur when a *series* of statistical tests are conducted. Recall that if in a series of statistical tests, each test has a .05 probability of a Type I error, then in a series of 20 such tests we would expect one (20 * .05 = 1) of these tests would report a significant difference that did not exist (Type I error). These tests typically are options provided by the standard statistical packages, such as the PASW (SPSS) program *Oneway*.

BIVARIATE ANALYSIS: MEASURES OF ASSOCIATION

Bivariate measures of association include the two-variable case in which both variables are interval or ratio scaled. Our concern is with the nature of the associations between the two variables and the use of these associations in making predictions.

Correlation Analysis

When referring to a simple two-variable correlation, we refer to the strength and direction of the relationship between the two variables. As an initial step in studying the relationship between the *X* and *Y* variables, it is often helpful to graph this relationship in a *scatter diagram* (also known as an *X*-*Y* plot). Each point on the graph represents the appropriate combination of scale values for the associated *X* and *Y* variables, as shown in Figure 12.3.

The objective of correlation analysis, then, is to obtain a measure of the degree of linear association (correlation) that exists between the two variables. The Pearson correlation coefficient is commonly used this purpose and is defined by the formula

$$\rho_{XY} = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - \bar{Y}}{S_Y}\frac{X_i - \bar{X}}{S_X} = \sum_{i=1}^{n}\frac{Z_X Z_Y}{n}$$

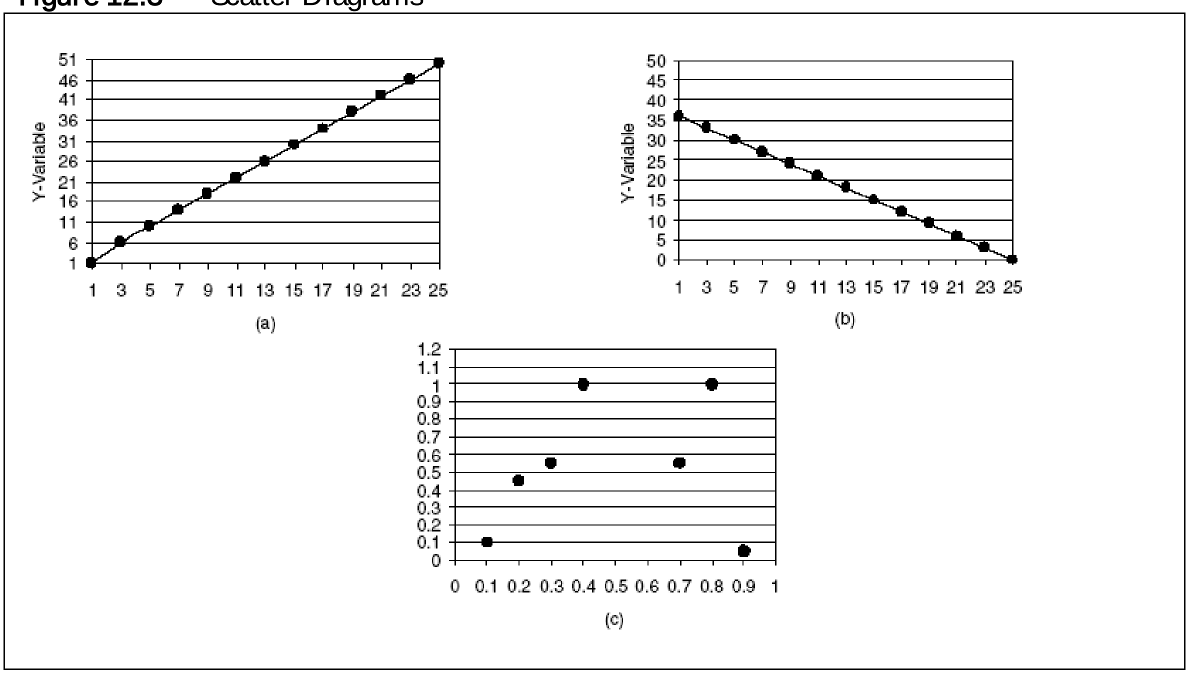
where n pairs of (X_i, Y_i) values provide a sample size n, and X, Y, s_X, and s_Y represent the sample means and sample standard deviations of the X and Y variables. The values of the correlation coefficient may range from +1 to –1.

These extreme values indicate perfect positive and negative linear correlations. Other relationships may appear to be curvilinear or even random plots and have coefficients between zero and the extreme values.

The alternate formulation shows the correlation coefficient to be the product of the Z scores for the X and Y variables. In this method of computing the correlation coefficient, the first step is to convert the raw data to a Z score by finding the deviation from the respective sample mean. The Z scores will be centered as a normally distributed variable (mean of zero and standard deviation of one).

The transformation of the X and Y variables to Z scores means that the scale measuring the original variable is no longer relevant, as a Z-score variable originally measured in dollars can be correlated with another Z-score variable originally measured on a satisfaction scale. The original metric scales are replaced by a new abstract scale (called correlation) that is the product of the two Z distributions.

Figure 12.3 Scatter Diagrams



Alternative formula:

$$\rho_{XY} = \frac{\sum_{i=1}^{n} Z_X Z_Y}{n}$$

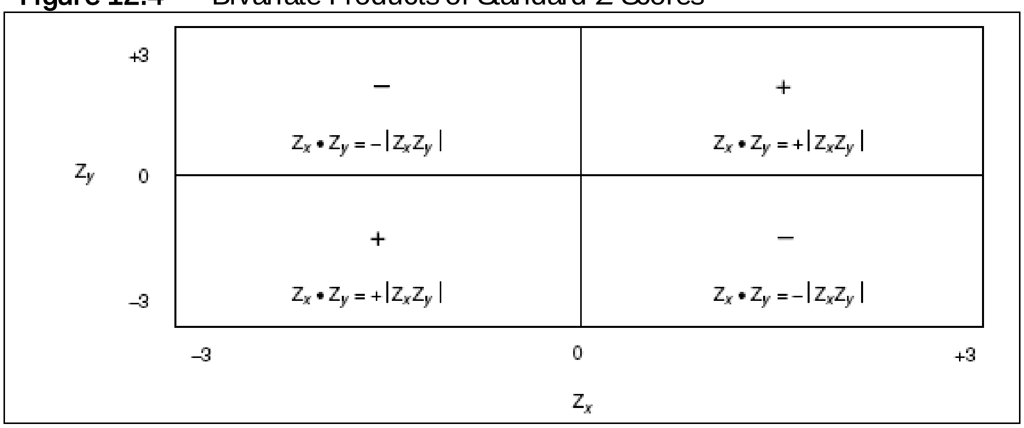
Where

$$Z_Y = \frac{Y_i - \bar{Y}}{S_Y} \quad Z_X = \frac{X_i - \bar{X}_1}{S_X}$$

By continuing our digression one step further, we can show how the correlation coefficient becomes positive or negative.

We know that the Z_X and Z_Y values will generally fall in the range –3 to +3. When both Z_X and Z_Y are positive or both are negative, r_{XY} is positive, as shown in Figure 12.6. When Z_X is positive and Z_Y negative (or the opposite), a negative correlation will exist. Of course, we are talking of the individual respondent's pairs of the Z_X and Z_Y variables, which when summed produce the overall correlation coefficient.

Figure 12.4 Bivariate Products of Standard *Z*-Scores



To summarize, the sign (+ or –) indicates the direction of the correlation and the absolute value of the coefficient indicates the degree of correlation (from 0 to 1). Thus, correlations of + .7 and – .7 are of exactly the same strength, but the relationships are in opposite directions.

We are cautioned that near-perfect positive or negative linear correlations do not mean causality. Often, other factors underlie and are even responsible for the relationship between the variables. For example, at one time Kraft Foods reported that sales of its macaroni and cheese product were highly correlated (negatively) with national indices of the health of the economy. We may not imply that Kraft sales directly caused fluctuations in the national economy, or vice versa. Consumer expectations and possibly personal income vary as a function of the national economy. In times of reduced family income, macaroni and cheese is a low-price dietary substitute for more expensive meals.

To demonstrate, we will consider a brief example of a correlation analysis that examines the relationships between (1) family income and (2) family consumption expenditures. The data and plot appear in Figure 12.5.

In order to calculate the correlation coefficient, we will reduce the previously identified alternative formula to the basic computational formula consisting of sums for the *X* and *Y* variables. This equation looks formidable but allows for easy computation by simply entering the appropriate summation values from the bottom of Table 12.10.

Figure 12.5 Consumption Expenditure and Income Data

X	39	43	21	64	57	47	28	75	34	52
Y	68	82	56	86	97	94	77	103	59	79

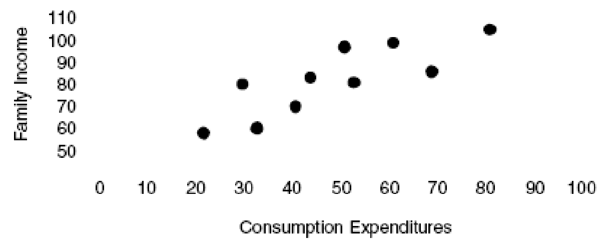


Table 12.10 Family Income and Family Consumption Expenditures

Resp.	*Y*	*X*	*XY*	Y^2	X^2
1	68	39	2652	4624	1521
2	82	43	3526	6724	1849
3	56	21	1176	3136	441
4	86	64	5504	7396	4096
5	97	57	5529	9409	3249
6	94	47	4418	8836	2209
7	77	28	2156	15929	784
8	103	75	7725	0609	5625
9	59	34	2006	3481	1156
10	79	52	4108	6241	2704
Sum	801	460	38800	66385	23634
Avg.	80.1	46	3880	6638.5	2363.4

Needless to say, hand computations such as this are rarely done today. Researchers routinely perform their analyses using Excel spreadsheets or computer packages. However, our brief discussion is included to provide understanding of underlying processes and the computational formula:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}} = \frac{\sum z_x z_y}{N}$$

Introduction to Bivariate Regression Analysis

In the analysis of associative data the marketing researcher is almost always interested in problems of predictions:

• Can we predict a person's weekly fast food and restaurant food purchases from that person's gender, age, income, or education level?
• Can we predict the dollar volume of purchase of our new product by industrial purchasing agents as a function of our relative price, delivery schedules, product quality, and technical service?

The list of such problems is almost endless. Not surprisingly, the linear regression model—as applied in either the bivariate (single predictor) or multivariate form (multiple predictors)—is one of the most popular methods in the marketing researcher's tool kit. The bivariate form is also known as *simple regression*.

The regression model has been applied to problems ranging from estimating sales quotas to predicting demand for new shopping centers. As an illustration, one of the leading ski resorts in the United States used a regression model to predict the weekend ticket sales, based on variables including the following:

- Highway driving conditions
- Average temperature in the three-day period preceding the weekend
- Local weather forecast for the weekend
- Amount of newspaper space devoted to the resort's advertisements in the surrounding city newspapers
- A moving average of the three preceding weekends' ticket sales

The model's accuracy was within ± 6 percent of actual attendance throughout the season.

Regression analysis in its simplest bivariate form involves a single dependent (criterion) and a single independent (predictor) variable. In its more advanced multiple-regression form, a set of predictors are used to form an additive linear combination that predicts the single dependent variable. In considering the use of either simple or multiple regressions, the researcher is interested in four main questions:

1. Can we find a predictor variable (or a linear composite of the predictor variables in the multiple case) that will parsimoniously express the relationship between a criterion variable and the predictor (set of predictors)?

2. If we can, how strong is the relationship; that is, how accurately can we predict values of the criterion variable from values of the predictor (linear composite)?

3. Is the overall relationship statistically significant?

4. Which predictor is most important in accounting for variation in the criterion variable? (Can the original model be reduced to fewer variables, but still provide adequate prediction of the criterion?)

The basic ideas of bivariate regression are most easily explained by a numerical example. We proceed one step at a time.

Suppose that a marketing researcher is interested in consumers' attitudes toward nutritional additives in ready-to-eat cereals. Specifically, a set of written concept descriptions of a children's cereal is prepared that vary on

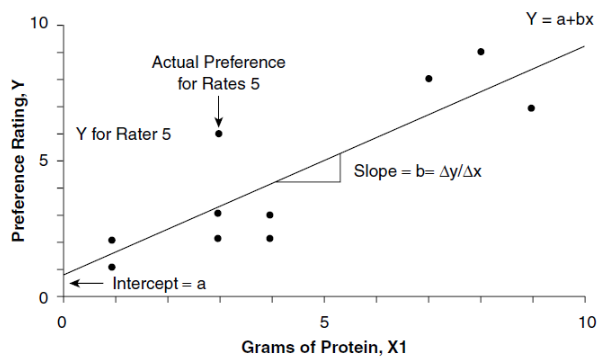
X_1: the amount of protein (in grams) per 2-ounce serving.

The researcher obtains consumers' interval-scaled evaluations of ten concept descriptions using a preference rating scale that ranges from 1, dislike extremely, up to 9, like extremely well.

TABLE 12.11 Preference Ratings of Ten Cereal Concepts Varying in Protein

(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Consumer	*Preference Rating, Y*	*Protein, X_1*	*Y^2*	*X*	*YX_1*
1	3	4	9	16	12
2	7	9	49	81	63
3	2	3	4	9	6
4	1	1	1	1	1
5	6	3	36	9	18
6	2	4	4	16	8
7	8	7	64	49	56
8	3	3	9	9	9
9	9	8	81	64	72
10	2	1	4	1	2
Total	43	43	261	255	247
Mean	4.3	4.3			
Std. Dev.	2.908	2.791			

Figure 12.6 Scatter Diagram and Least-Squares Regression Line--Preference Rating versus Grams of Protein



One of the first things that is usually done in examining two-variable relationships is to prepare a *scatter diagram* in which the ten values of Y are plotted against their X_1 counterparts. Figure 12.6 shows this plot. It appears that there is a direct relationship between Y and X_1. Moreover, it would seem that a linear or straight-line relationship might be an appropriate model for describing the functional form. A scatter diagram is a useful tool for aiding in *model specification*. The value of using a scatter diagram is illustrated in Exhibit 12.3.

Exhibit 12.3 Look at Your Data Before You Analyze

In deciding which type of regression approach to use, it is important that the researcher know the *shape* of the interrelationship. The shape of the interrelationship is easy to see on a scatter diagram. Looking at this visually helps decide whether the relationship is, or approximates being, linear or whether it has some other shape which would require a transformation of the data—by converting to square roots or logarithms--or treatment as nonlinear regression (Semon, 1993).

Examination of the data by scatter diagrams also allows the researcher to see if there are any "outliers"—i.e., cases where the relationship is unusual or extreme as compared to the majority of the data points. A decision has to be made whether to retain such outliers in the data set for analysis.

When the regression line itself is included on the scatter diagram, comparisons between actual values and the values estimated by the regression formula can be compared and used to assess the estimating error. Of course, what the analyst is seeking is a regression function that has the *best fit* for the data, and this is typically based on minimizing the squares of the distances between the actual and estimated values—the so-called *least-squares* criterion.

The equation for a linear model can be written $\hat{Y} = a + bx$, where $\hat{Y}$ denotes values of the criterion that are predicted by the linear model; *a* denotes the intercept, or value of $\hat{Y}$ when *X* is zero; and *b* denotes the slope of the line, or change in $\hat{Y}$ per unit change in *X*.

But how do we find the numerical values of *a* and *b*? The method used in this chapter is known as *least squares* as discussed in Exhibit 12.3. As the reader will recall from introductory statistics, the method of least squares finds the line whose sum of squared differences between the observed values Y_i and their estimated counterparts $\hat{Y}_i$ (on the regression line) is a minimum.

Parameter Estimation

To compute the estimated parameters (*a* and *b*) of the linear model, we return to the data of Table 12.11. In the two-variable case, the formulas are relatively simple:

$$b = \frac{\sum YX_1 - n\overline{Y}\,\overline{X}}{\sum X^2 - n\overline{X}^2}$$

$$= \frac{247 - 10(4.3)(4.3)}{255 - 10(4.3)^2}$$

$$= 0.886$$

where *n* is the sample size and Y and X denote the mean of *Y* and *X*, respectively. Having found the slope *b*, the intercept *a* is found from

$$a = \overline{Y} - b\overline{X}$$
$$= 4.3 - 0.886(4.3)$$
$$= 0.491$$

leading to the linear function $\hat{Y} = 0.491 + 0.886X$

This function is drawn with the scatter plot of points in Figure 12.6. It appears to fit the plotted points rather well, and the model seems to be well specified (a linear rather than arc linear or other form seems to fit).

Assumptions of the Model

Underlying least-squares computations is a set of assumptions. Although least-squares regression models do not need to assume normality in the (conditional) distributions of the criterion variable, this assumption is made when we test the statistical significance of the contribution of the predictor variable in explaining the variance in the criterion (does it differ from zero?) With this in mind the assumptions of the regression model are as follows (the symbols α and β are used to denote population counterparts of *a* and *b*:

1. For each fixed value of X we assume a normal distribution of Y values exists. Our particular sample we assume that each y value is drawn independently of all others. What is being described is the "classical" regression model. Modern versions of the model permit the predictors to be random variables, but their distribution is not allowed to depend on the parameters of the regression equation.

2. The means of all of these normal distributions of Y lie on a straight line with slope β.

3. The normal distributions of Y all have equal variances. This (common) variance does not depend on values assumed by the variable X.

Because all values rarely fall on the regression line (only when the correlation is 1.0), there is unexplained error in predicting Y. This error is shown in Figure 12.3 as the difference between the values Y_i and the regression line $\hat{Y}$. For the population, our model is expressed algebraically as:

$$Y = \alpha + \beta X_1 + \varepsilon$$

where α = mean of Y population when $X_1 = 0$
β = change in Y population mean per unit change in X_1
ε = error term drawn independently from a normally distributed universe with mean μ(ε); the error term is independent of X_1

The nature of these assumptions is apparent in Figure 12.7. The reader should note that each value of X has associated with it a normal curve for Y (assumption 1). The means of all these normal distributions lie on the straight line shown in the figure (assumption 2).

What if the dependent variable is not continuous? Exhibit 12.4 gives an alternative when the dependent variable can be viewed as a categorical dichotomous variable—use *logistic regression* (also known as *logit*). The analysis proceeds generally as we are discussing it—the major change is that a transformation has been applied to the dependent variable values.

Exhibit 12.4 When to Use Logistic Regression

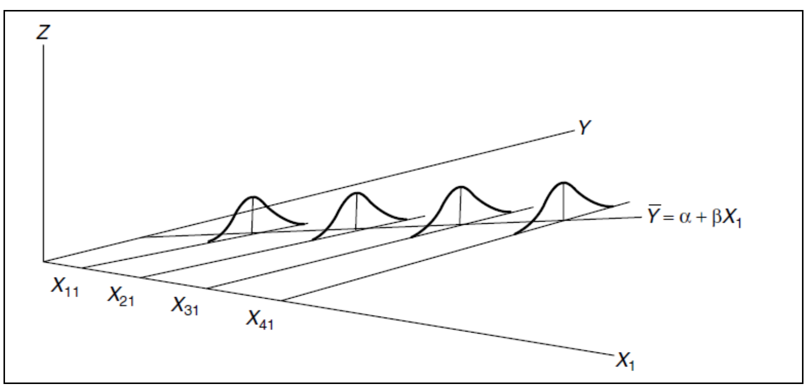
Data collected for customer satisfaction research provides a good illustration of when the researcher should consider transformation of data. Typically multi-point rating scales are used to obtain customer satisfaction data. Many believe that customer satisfaction ratings obtained on rating scales are not normally distributed, but are skewed toward higher scale values (Dispensa, 1997). Thus, in practice customers do not really view customer satisfaction ratings as continuous.

Ultimately, overall a customer is either satisfied or not satisfied. This creates a dichotomous dependent variable. Typically those customers who rate at the upper end of the scale, say 9 or 10 on a 10-point scale, are considered satisfied while all others are considered to be not satisfied. If this is so, normal regression analysis is not the proper technique to use as the dependent variable is binary not continuous.

A binary overall customer satisfaction variable follows the *logistic distribution*, thus allowing for the use of logistic regression. With one or more independent variables, this technique allows a researcher to determine the extent to which an independent variable affects the prediction of a satisfied customer through the logistic regression coefficients and their associated *log-odds* (Dispensa, 1997). Log-odds specify the direct association between the independent variable and the dependent variable. In addition, logistic regression calculates the probability of each customer being satisfied or not.

Typically, logistic regression is used for multiple regression situations where there are two or more independent variables. But, it is suitable for bivariate situations as well. The key to its being of value is the nature of the dependent variable, not the independent variable(s).

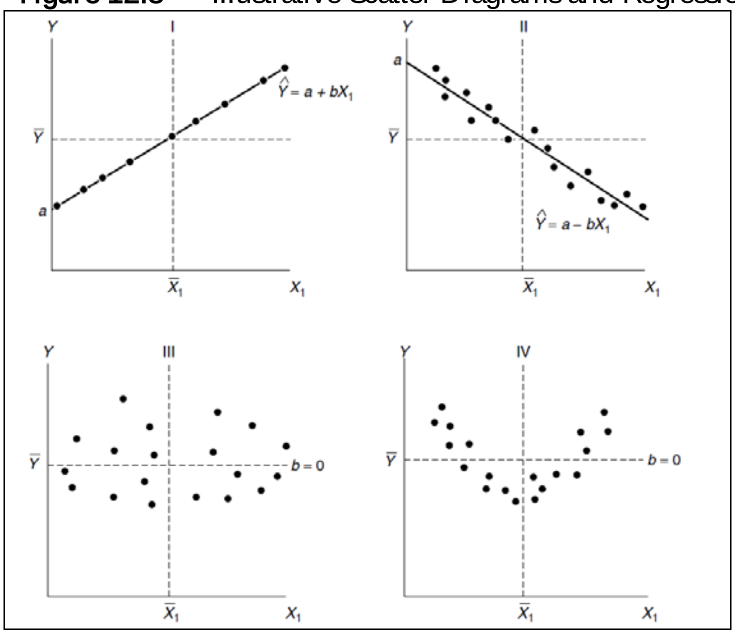
Figure 12.7 Two-Variable Regression Model--Theoretical



In constructing the estimating equation by least squares we have computed a regression line for a sample, and not for the population: $\hat{Y} = a + bX_1$ where $\hat{Y}$ is the estimated mean of Y, given X, and a, b are the sample estimates of α and β in the theoretical model. As already noted, this line appears in Figure 12.6 for the specific bivariate problem of Table 12.11.

However, functional forms other than linear may be suggested by the preliminary scatter plot. Figure 12.8 shows various types of scatter diagrams and regression lines for the two-variable case. Panel I shows the ideal case in which *all* the variation in Y is accounted for by variation in X_1. We note that the regression line passes through the mean of each variable and that the slope b happens to be positive. The intercept a represents the predicted value of Y when $X_1 = 0$. In Panel II we note that there is residual variation in Y, and, furthermore, that the slope b is negative. Panel III demonstrates the case in which no association between Y and X_1 is found. In this case the mean of Y is as good a predictor as the variable X_1 (the slope b is zero). Panel IV emphasizes that a linear model is being fitted. That is, no *linear* association is found (b = 0), even though a curvilinear relationship is apparent from the scatter diagram. Figure 12.8 illustrates the desirability of plotting one's data *before* proceeding to formulate a specific regression model.

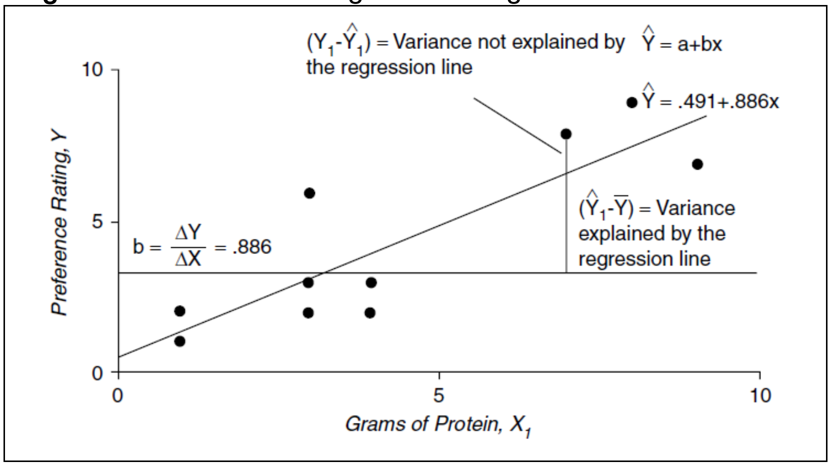
Figure 12.8 Illustrative Scatter Diagrams and Regression Lines



Strength of Association

It is one thing to find the regression equation (as shown in Figure 12.6), but at this point we still do not know how strong the association is well. Does the regression line $\hat{Y} = a + bx$ (which uses X as a predictor) explain the variation in Y (predict Y)? To answer this question, consider that the total variation in the Y variable may be divided into two component parts: (1) variance that is explained by the regression line, and (2) unexplained variation (residual). This may be expressed as and is viewed graphically in Figure 12.9.

Figure 12.9 Scatter Diagram and Regression Line for Cereal Problem



$\Sigma(Y_i-\bar{Y})^2$ =	$\Sigma(\hat{Y}_i-\bar{Y})^2$ +	$\Sigma(Y_i-\hat{Y})^2$
Total variation or Total Sum of Squares Deviation from the Mean	Variation Explained by the Regression or Sum of Squares Due to Regression	Variation Unexplained by the Regression. Sum of Squares Deviation from Regression (Error)

The measure of strength of association in bivariate regression is denoted by r^2 and is called the *coefficient of determination*. This coefficient varies between 0 and 1 and represents *the proportion of total variation in Y (as measured about its own mean Y) that is accounted for by variation in X_1*. For regression analyses it can also be interpreted as a measure of substantive significance, as we have previously defined this concept.

If we were to use the average of the Y values (Y) to estimate each separate value of Y, then a measure of our *inability* to predict Y would be given by the sum of the squared deviations $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$. On the other hand, if we tried to predict *Y* by employing a linear regression based on X, we could use each $\hat{Y}_i$ to predict its counterpart Y_i. In this case a measure of our *inability* to predict Y_i is given by $\sum_{i=1}^{n}(Y_i - \hat{Y}_I)^2$. We can define r^2 as a function of these two qualities:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

If each $\hat{Y}_i$ predicts its counterpart Y_i perfectly, then $r^2 = 1$, since the numerator of the variance unaccounted for is zero. However, if using the regression equation does no better than $\bar{Y}$ alone, then the total variance equals the variance unaccounted for and $r^2 = 0$, indicating no ability to predict Y_i (beyond the use of $\bar{Y}$ itself). The use of x_1 in a linear regression can do no worse than Y. Even if b turns out to be zero, the predictions are $\hat{Y}_i = a = Y$, which are the same as using the mean of criterion values in the first place.

Table 12.12 shows the residuals obtained after using the regression equation to predict each value of Y_t via its counterpart $\hat{Y}_i$. We then find r_x (where we now show the explicit subscripts) by computing from the table:

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = (-1.034)^2 + (-1.464)^2 + ... + (0.623)^2$$
$$= 21.09$$

Table 12.12 Actual Y_i, Predicted $\hat{Y}$, and Residuals $Y_i - \hat{Y}$

Rater	Actual, Y_i	Predicted,* $\hat{Y}_i$	Residuals $Y_i - \hat{Y}_i$
1	3	4.034	-1.034
2	7	8.464	-1.464
3	2	3.148	-1.148
4	1	1.377	-0.377
5	6	3.148	2.852
6	2	4.034	-2.034
7	8	6.692	1.308
8	3	3.148	-0.148
9	9	7.579	1.422
10	2	1.377	0.623
Mean	4.3	4.3	0

*From the equation $\hat{Y}_i = 0.491 + 0.886X_{i1}$.

This is the sum of squared errors in predicting Y_i from $\hat{Y}_i$. Next, we find:

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = (3 - 4.3)^2 + (7 - 4.3)^2 + ... + (2 - 4.3)^2$$
$$= 76.10$$

This is the sum of squared errors in predicting Y_i from $\overline{Y}$. Hence,

$$r^2_{yx_1} = 1 - \frac{21.09}{76.10}$$
$$= 0.723$$

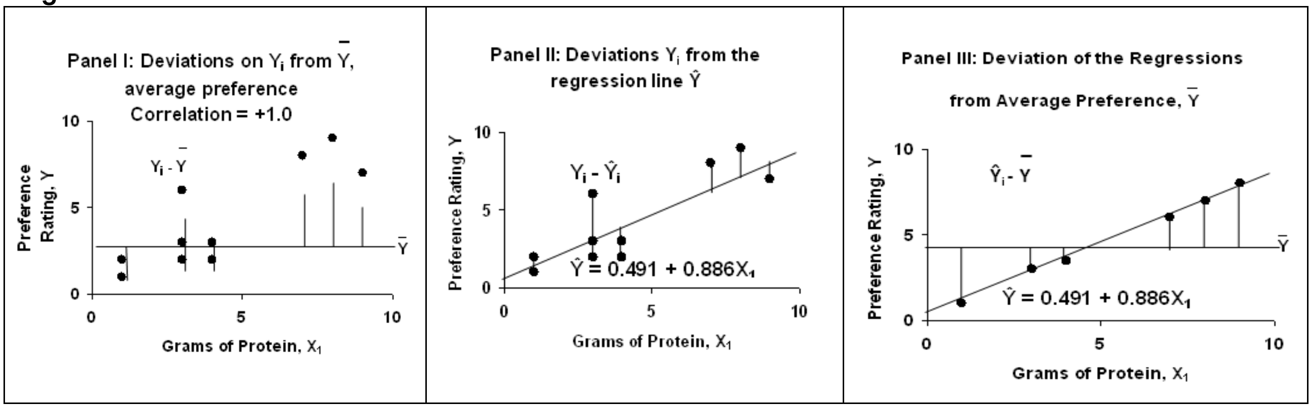
and we say that 72% of the variation in Y has been accounted for by variation in X_1. As might also be surmised, there is one more quantity of interest:

$$\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = (4.034 - 4.3)^2 + (8.464 - 4.3)^2 + ... + (1.377 - 4.3)^2$$
$$= 55.01$$

which is the accounted-for sum of squares due to the regression of Y on X_1.

Figure 12.10 (and 12.9 as well) put all these quantities in perspective by first showing deviation of $Y_i - \overline{Y}$. As noted above, the sum of these squared deviations is 76.10. Panel II shows the counterpart deviations of Y_i from $\hat{Y}_i$; the sum of these squared deviations is 21.09. Panel III shows the deviations of $\hat{Y}_i$ from $\overline{Y}_i$; the sum of these squared deviations is 55.01. We note that the results are additive: 21.09 + 55.01 = 76.10.

Figure 12.10 Breakdown of Deviations $Y_i - \bar{Y}$ into Two Additive Parts



Interpretation of Bivariate Regression

The sample data of Table 12.11 were analyzed using a standard linear regression analysis routine. Table 12.13 shows the output provided. The coefficients shown earlier appear here, along with some other measures as well.

Table 12.13 Summary Output of Regression Analysis of Table 12.11 Sample data

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	PROTEIN[a]		Enter

a. All requested variables entered.
b. Dependent Variable: PREFEREN

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.850[a]	.723	.688	1.62354

a. Predictors: (Constant), PROTEIN

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	55.013	1	55.013	20.871	.002[a]
	Residual	21.087	8	2.636		
	Total	76.100	9			

a. Predictors: (Constant), PROTEIN
b. Dependent Variable: PREFEREN

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model		B	Std. Error			
1	(Constant)	.491	.979		.501	.630
	PROTEIN	.886	.194	.850	4.568	.002

a. Dependent Variable: PREFEREN

Overall, the linear regression equation is:

$$Y = 0.491 + 0.886\ X_1$$

The standardized coefficient for the independent variable (beta or β) is more meaningful for multiple regression as it is a measure of the change in Y due to a unit change in X_i when all independent variables have been "transformed" to the same units.

The F ratio is the appropriate test for the hypotheses that the regression coefficient, $b_1 = 0$ and $r^2 = 0$. This test was discussed in more detail in Chapter 11, in the context of analysis of variance. However, as recalled from basic statistics, the F distribution is the distribution followed by the ratio of two independent, unbiased estimates of the normal population variance σ^2. If r^2 is zero, then the sample r^2 reflects only sampling error and the F ratio will tend to be equal to 1.0. We first obtain the mean squares, 55.013 and 2.636, by dividing the corresponding sums of squares by their respective degrees of freedom. Then we find the F ratio of 20.871. This value, with 1 degree of freedom for numerator and 8 degrees of freedom for denominator, is compared with a tabular F (see Table A-4 in Appendix A) of 3.46 with an (illustrative) significance level of 0.1. We reject the preceding null hypotheses and conclude that the overall regression equation is statistically significant.

Rank Correlation

The previous discussion of correlation and bivariate regression were based on the premise that the dependent variable is at least interval scaled or can be treated as such with little error. There are marketing problems, however, where the dependent and independent variables are rank orders or are best transformed into such rankings. In this situation the Spearman r's *rank correlation* can be used to estimate the association between sets of data.

We show the use of this measure by an example. Suppose a sales manager ranks salespersons by two different methods (performance index and a new subjective method). Since the new method is easier to use, the manager wants to know if it will yield the same relative results as the proven existing method. The scores have been transformed into rankings so that each salesperson has two rankings. Table 12.14 shows the rankings.

Table 12.14 Ranks on Two Methods of Salesperson Evaluation

	Rank			
Salesperson	*Performance Index (X)*	*New Method (Y)*	d_i	d
A	8	6	2	4
B	4	7	−3	9
C	1	2	−1	1
D	6	3	3	9
E	2	1	1	1
F	10	8	2	4
G	5	5	0	0
H	3	9	−6	36
I	7	4	3	9
J	9	10	1	1
			$\sum d_i^2 =$	74

To measure the extent of rank correlation we use the statistic

$$r_S = 1 - \frac{6 \sum_{i=1}^{N} d^2}{N(N^2 - 1)}$$

where N is the number of pairs of ranks and d is the difference between the two rankings for an individual (that is, $X - Y$). Applying this formula to our example, we get

$$r_S = 1 - \frac{6(74)}{10(100 - 1)} = .55$$

If the subjects whose scores were used in computing r_S were randomly drawn from a population, we can test the significance of the obtained value. The null hypothesis is that the two variables are not associated, and thus the true value of ρ is zero. Under H_0 any observed value would be due to chance. When $N \geq 10$, significance can be tested using the statistic

$$t = r_S \sqrt{\frac{N-2}{1-r_S^2}}$$

which is interpreted from a table of t values with (*N* - 2) degrees of freedom. For our example, we calculate

$$t = .55\sqrt{\frac{10-2}{1-(.55)^2}} = 1.870$$

Looking at the table of critical values of *t*, we find that $p > .10$ (two-tailed test) for (10 – 2 = 8) degrees of freedom. Thus, if a strict α level is to be adhered to (.10 or less), we tentatively accept H_0 and conclude that it is unlikely that a correlation exists between the scores from the two evaluation methods.

One final point concerning the use of the Spearman rank correlation coefficient is warranted. At times, tied observations will exist. When this happens, each of them is assigned the average of the ranks λ minute. If large, however, a correction factor must be applied (see Siegel, 1956, pp. 206–10).

The Spearman rank correlation coefficient is equivalent to the Pearson product-moment correlation coefficient with ranks substituted for the measurement observations, *X* and *Y*. The Spearman and other measures are discussed more fully by Siegel (1956, Chapter 9) and Gibbons (1993).

Finally, when the variables are nominally scaled, ordinal measures such as tau and r_S are not appropriate measures. Nominal variables lack the ordering property. One measure that can be used is Goodman and Kruskal's lambda (λ). Lambda is a measure of association whose calculation and interpretation are straightforward. Lambda tells us how much we can reduce our error in predicting *Y* once we know *X*, and is shown as

$$\lambda = \frac{\text{reduction in prediction errors knowing } X}{\text{prediction errors in not knowing}}$$

This measure, and others as well, are discussed by Lewis-Beck (1995, Chap. 4). Lambda is an option provided by most *Crosstab* programs. We discuss lambda further in the next section.

NONPARAMETRIC ANALYSIS

One reason for the widespread use of chi-square in cross-tabulation analysis is that most computer computational routines show the statistic as part of the output, or at least it is an option that the analyst can choose. Sometimes ordinal data are available and as such are stronger than simple nominal measurement. In this situation other tests are more powerful than chi-square. Three regularly used tests are the *Wilcoxon Rank Sum* (T), the *Mann-Whitney U*, and the *Kolmogorov-Smirnov* test. Siegel (1956) and Gibbons (1993) provide more detailed discussions of these techniques.

The Wilcoxon T test is used for dependent samples in which the data are collected in matched pairs. This test takes into account both the direction of differences within pairs of observations and the relative magnitude of the differences. The Wilcox matched-pairs signed-ranks test gives more weight to pairs showing large differences between the two measurements than to a pair showing a small difference. Again, to use this test, measurements must at least be ordinal scaled within pairs. In addition, ordinal measurement must hold for the differences between pairs.

This test has many practical applications in marketing research. For instance, an ordinal scaling device, such as a semantic differential, can be used to measure attitudes toward, say, a bank. Then, after a special promotional campaign, the same sample would be given the same scaling device. Changes in values of each scale could be analyzed by this Wilcoxon test.

With ordinal measurement and two independent samples, the Mann-Whitney U test may be used to test whether the two groups are from the same population. This is a relatively powerful nonparametric test, and is an alternative to the Student *t* test when the analyst cannot meet the assumptions of the *t* test or when measurement is at best ordinal. Both one- and two tailed tests can be conducted. As indicated earlier, results of U and t tests often are similar, leading to the same conclusion.

The Kolmogorov-Smirnov two-sample test is a test of whether two independent samples come from the same population or from populations with the same distribution. This test is sensitive to any kind of difference in the distributions from which the two samples were drawn—differences in location (central tendency), dispersion, skewness, and so on. This characteristic of the test makes it a very versatile test. Unfortunately, the test does not by itself show what kind of difference exists. There is a Kolmogorov-Smirnov one-sample test that is concerned with the agreement between an observed distribution of a set of sample values and some specified theoretical distribution. In this case it is a goodness-of-fit test similar to single-classification chi-square analysis.

Indexes of Agreement

Chi-square is appropriate for making statistical tests of independence in cross-tabulations. Usually, however, we are interested in the *strength* of association as well as the *statistical significance* of association. This concern is for what is known as *substantive* or *practical* significance. An association is substantively significant when it is statistically significant and of sufficient strength. Unlike statistical significance, however, there is no simple numerical value to compare with and considerable experimental research judgment is necessary. Although such judgment is subjective, it need not be completely arbitrary. The nature of the problem can offer some basis for judgment, and common sense can indicate that the degree of association is too low in some cases and high enough in others (Gold, 1969, p. 44).

Statisticians have devised a large number of indexes—often called *indexes of agreement*—for measuring the strength of association between two variables in a cross-tabulation. The main descriptors for classifying the various indexes are

1. Whether the table is 2 x 2 or larger, *R* x *C*
2. Whether one, both, or neither of the variables has categories that obey some natural order (e.g., age, income level, family size)

3. Whether association is to be treated symmetrically or whether we want to predict membership in one variable's categories from (assumed known) membership in the other variable's categories

Space does not permit coverage of even an appreciable fraction of the dozens of agreement indexes that have been proposed. Rather, we shall illustrate one commonly used index for 2 x 2 tables and two indexes that deal with different aspects of the larger *R* x *C* (row-by-column) tables.

The 2 × 2 Case

The *phi correlation coefficient* is a useful agreement index for the special case of 2 x 2 tables in which both variables are dichotomous. Moreover, an added bonus is the fact that phi equals the product-moment correlation—a cornerstone of multivariate methods—that one would obtain if he or she correlated the two variables expressed in coded 0 – 1 form.

To illustrate, consider the 2 x 2 cross-tabulation in Table 12.15, taken from a study of shampoos. We wish to see if inclusion of the shampoo benefit "body" in the respondent's ideal set is associated with the respondent's indication that her hair lacks natural "body." We first note from the table that high frequencies appear in the cells: (a) "body" included in ideal set and "no" to the question of whether her hair has enough (natural) body; and (b) "body" excluded from the ideal set and "yes" to the same question.

Table 12.15 Does Hair Have Enough Body Verses Body Inclusion in Ideal Set

	Hair Have Enough Body?		
	No	*Yes*	*Total*
Body included in ideal set	26 (A)	8 (B)	34
Body excluded from ideal set	17 (C)	33 (D)	50
Total	43	41	84

Before computing the phi coefficient, first note the labels, A, B, C, and D assigned to the four cells in Table 12.15. The phi coefficient is defined as

$$\begin{aligned}\phi &= \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}\\ &= \frac{26(33) - 8(17)}{\sqrt{(26+8)(17+33)(26+17)(8+33)}}\\ &= 0.417\end{aligned}$$

The value 0.417 is also what would be found if an ordinary product-moment correlation, were computed across the 84 pairs of numbers where the following dichotomous code values are used to identify the responses:

- Body in ideal set: Included $\Rightarrow$ 1, Excluded $\Rightarrow$ 0
- Hair have enough body? No $\Rightarrow$ 1, Yes $\Rightarrow$ 0

The phi coefficient can vary from −1 to 1 (just like the ordinary product-moment correlation). However, in any given problem the upper limit of phi depends on the relationships among the marginals. Specifically, a phi coefficient of –1 (perfect negative association) or 1 (perfect positive association) assumes that the marginal totals of the first variable are identical to those of the second. Looking at the letters (A, B, C, D) of Table 12.15, assume that the row marginals equaled the column marginals: then, $\Phi = 1$ if $B = C = 0$; similarly, $\Phi = -1$ if $A = D = 0$. The more different the marginals, the lower the upper limit that the (absolute) value of phi can assume.

The phi coefficient assumes the value of zero if the two variables are statistically independent (as would be shown by a chi-square value that is also zero). Indeed, the absolute value of phi is related to chi-square by the expression

$$\phi = \sqrt{\frac{\chi^2}{n}} \quad 1$$

where n is the total frequency (sample size). This is a nice feature of phi, in the sense that it can be computed quite easily after chi-square has been computed. Note, however, that phi, unlike chi-square, is *not* affected by total sample size because we have the divisor n in the above formula to adjust for differences in sample size.

The R × C Case

One of the most popular agreement indexes for summarizing the degree of association between two variables in a cross-tabulation of R rows and C columns is the *contingency coefficient*. This index is also related to chi-square and is defined as

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad 2$$

where n is again the total sample size. From Table 12.15 we can first determine that chi-square is equal to 14.61, which, with 1 degree of freedom, is significant beyond the 0.01 level.

We can then find the contingency coefficient C as the following:

$$C = \sqrt{\frac{14.61}{14.61 + 84}}$$
$$= 0.385$$

As may be surmised, the contingency coefficient lies between zero and 1, with zero reserved for the case of statistical independence (a chi-square value of zero). However, unlike the phi coefficient, the contingency can never attain a maximum value of unity. For example, in a 2 × 2 table, C cannot exceed 0.707. As might be noticed by the reader, there is an algebraic relationship between phi and the contingency coefficient (if the latter is applied to the 2 × 2 table):

$$\phi^2 = \frac{C^2}{1 - C^2}$$

In a 4 X 4 table its upper limit is 0.87. Therefore, contingency coefficients computed from different-sized tables are not easily comparable.

However, like phi, the contingency coefficient is easy to compute from chi-square; moreover, like phi, its significance has already been tested in the course of running the chi-square test.

Both phi and the contingency coefficient are symmetric measures of association. Occasions often arise in the analysis of *R X C* tables (or the special case of 2 · 2 tables) where we desire to compute an *asymmetric* measure of the extent to which we can reduce errors in predicting categories of one variable from knowledge of the categories of some other variable. Goodman and Kruskal's *lambda-asymmetric coefficient* can be used for this purpose (Goodman & Kruskal, 1954).

To illustrate the lambda-asymmetric coefficient, let us return to the cross-tabulation of Table 12.15. Suppose that we wished to predict what category—no versus yes—a randomly selected person would fall in when asked the question, "Does your hair have enough body?" If we had no knowledge of the row variable (whether that person included "body" in her ideal set or not), we would have only the *column* marginal frequencies to rely on.

Our best bet, given no knowledge of the row variable, is always to predict "no," the *higher* of the column marginal frequencies. As a consequence, we shall be wrong in 41 of the 84 cases, a probability error of 41/84 = 0.49 Can we do better, in the sense of lower prediction errors, if we utilize information provided by the row variable?

If we know that "body" is included in the ideal set, we shall predict "no" and be wrong in only 8 cases. If we know that "body" is not included in the ideal set, we shall predict "yes" and be wrong in 17 cases. Therefore, we have reduced our number of prediction errors from 41 to 8 + 17 = 25, a decrease of 16 errors. We can consider this error reduction relatively:

$$\lambda_{C|R} = \frac{\text{(number of errors in first case)} - \text{(number of errors in second case)}}{\text{numbers of errors in first case}}$$
$$= \frac{41 - 25}{41} = 0.39$$

In other words, 39 percent of the errors in predicting the column variable are eliminated by knowing the individual's row variable.

A less cumbersome (but also less transparent) formula for lambda-asymmetric is

$$\lambda_{C|R} = \frac{\sum_{k=1}^{K} f_{k_R^*} - F_c^*}{n - F_c^*} = \frac{(26 + 33) - 43}{84 - 43} = 0.39$$

Where f_{kr} *is* the *maximum* frequency found within each subclass of the row variable, F_c is the *maximum* frequency among the marginal totals of the column variable, and n is the total number of cases.

Lambda-asymmetric varies between zero, indicating no ability at all to eliminate errors in predicting the column variable on the basis of the row variable, and 1, indicating an ability to eliminate all errors in the column variable predictions, given knowledge of the row variable. Not surprisingly, we could reverse the role of criterion and predictor variables and find lambda-asymmetric for the row variable, given the column variable. In the case of Table 12.15, this results in λ=0.26.

Note that in this case we simply reverse the roles of row and column variables.

Finally, if desired, we could find a *lambda-symmetric index* via a weighted averaging of $\lambda_{C|R}$ and $\lambda_{R|C}$. However, in the authors' opinion, lambda-asymmetric is of particular usefulness to the analysis of cross-tabulations because we often want to consider one variable as a predictor and the other as a criterion. Furthermore, lambda-asymmetric *h* is a natural and useful interpretation as the percentage of total prediction errors that are eliminated in predicting one variable (e.g., the column variable) from another (e.g., the row variable).

SUMMARY

We began by stating that data can be viewed as recorded information useful in making decisions. In the initial sections of this chapter, we introduced the basic concepts of transforming raw data into data of quality. The introduction was followed by a discussion of elementary descriptive analyses through tabulation and cross-tabulation. The focus of this discussion was heavily oriented toward how to read the data and how to interpret the results. The competent analysis of research-obtained data requires a blending of art and science, of intuition and informal insight, and of judgment and statistical treatment, combined with a thorough knowledge of the context of the problem being investigated.

The first section of the chapter dealt with cross-tabulation and chi-square analysis. This was followed by discussing bivariate analysis of differences in means and proportions. We next focused on the necessary statistical machinery to analyze differences between groups: *t*-test, and one-factor and two-factor analysis of variance. These techniques are useful for both experimental and nonexperimentally obtained data. We then looked at the process of analysis of variance. A simple numerical example was used to demonstrate the partitioning of variance into among- and within-components. The assumptions underlying various models were pointed out and a hypothetical data experiment was analyzed to show how the ANOVA models operate.

We concluded by examining bivariate analyses of associations for interval- or ratio scaled data. The concept of associations between two variables was introduced through simple two-variable correlation. We examined the strength and direction of relationships using the scatter diagram and Pearson correlation coefficient. Several alternative (but equivalent) mathematical expressions were presented and a correlation coefficient was computed for a sample data set.

Investigations of the relationships between variables almost always involve the making of predictions. Bivariate (two-variable) regression was discussed as the foundation for the discussion of multivariate regression in the next chapter.

We ended the chapter with a discussion of the Spearman rank correlation as an alternative to the Pearson correlation coefficient when the data is of ordinal measurement and does not meet the assumptions of parametric methods. Also, the Goodman and Kruskal lambda measure for nominal measurement was briefly introduced, as were other nonparametric analyses.

There is a wide array of statistical techniques (parametric and non parametric) that focus on describing and making inferences about the variables being analyzed. Some of these were shown in Table 12.4. Although somewhat dated, a useful reference for selecting an appropriate statistical is the guide published by the Institute for Social

Research at the University of Michigan (Andrews, et. al., 1981 and its corresponding software *Statistical Consultant*. Fing (2003, pp. 78-80) presents a summary table of which technique to use under which condition.

REFERENCES

Andrews, F. M., Klem, L., Davidson, T. N., O'Malley, P. M., & Rodgers, W. L. (1981). *A guide for selecting statistical techniques for analyzing social science data* (2nd ed.). Ann Arbor: Institute for Social Research, University of Michigan.

Feick, L. F. (1984, November). Analyzing marketing research data with associated models. *Journal of Marketing Research, 21*, 376–386.

Fink, A. (2003). *How to manage, analyze, and interpret survey data*. Thousand Oaks, CA: Sage.

Gibbons, J. D. (1993). *Nonparametric statistics: An introduction*. Newbury Park, CA: Sage.

Gold, D. (1969, February). Statistical tests and substantive significance. *American Sociologist*, *4*, 44.

Goodman, L. A., & Kruskal, W. H. (1954, December). Measures of association for cross classification. *Journal of the American Statistical Association, 49*, 732–764.

Hellevik, O. (1984). *Introduction to causal analysis: Exploring survey data*. Beverly Hills, CA: Sage.

Lewis-Beck, M. S. (1995). *Data analysis: An introduction*. Thousand Oaks, CA: Sage.

Semon, T. T. (1999, August 2). Use your brain when using a chi-square. *Marketing News*, *33*, 6.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Smith, S. and Albaum, G. (2005). *Fundamentals of marketing research*. Thousand Oaks, CA: Sage Publications.

Zeisel, H. (1957). *Say it with figures* (4th ed.). New York: Harper & Row.

Chapter 13
MULTIVARIATE STATISTICAL ANALYSIS I
Analyzing Criterion-Predictor Association

In the previous chapters that discussed analysis of differences and association for two variables, we determined that selecting the appropriate method of analysis requires evaluating the number of dependent and independent variables and their level of measurement. The same process applies to multivariate analysis, as is illustrated in Figure 13.1.

Association implies only that two or more variables tend to change together to a greater or lesser degree, depending upon the degree of association involved. If we measure the amount of mutual change and find it to be persistent in both direction and degree, we may *not* conclude that there is necessarily a *causal* relationship, such that one variable is dependent (the effect) and the other variable (or variables) is independent (the deterministic or probabilistic cause). It should be understood, however, that *association does not imply causation*. However, if a set of variables are causally related, they will be associated in some way. Analytical techniques such as causal-path analysis and structural equation models have been developed to aid in examining possible causal relations in correlational data. These are discussed in depth by Blalock (1962), Bagozzi (1980), and Monroe and Petroshius (n.d.).

In this and the next chapter, we present a series of brief discussions that focus on individual multivariate techniques. Multivariate analyses arise when more than two variables are to be analyzed at the same time. We begin by extending simple bivariate regression discussed in the last chapter to include a second predictor variable. This use of multiple predictors allows the researcher to conduct both multiple and partial regression.

We next focus on two-group discriminant analysis. Like multiple regression, discriminant analysis is a predictor model. However, the dependent variable is categorical in measurement and defines two or more groups. In discriminant analysis, group membership (such as purchaser and non-purchaser groups) is predicted using the multiple independent variables.

Finally, a brief conceptual introduction is given to a diverse group of techniques, including canonical correlation, correspondence analysis, Chi-Square Automatic Interaction Detection (CHAID), and probit and logit analyses. We start with a broad overview of multivariate procedures.

AN OVERVIEW OF MULTIVARIATE PROCEDURES
The Data Matrix

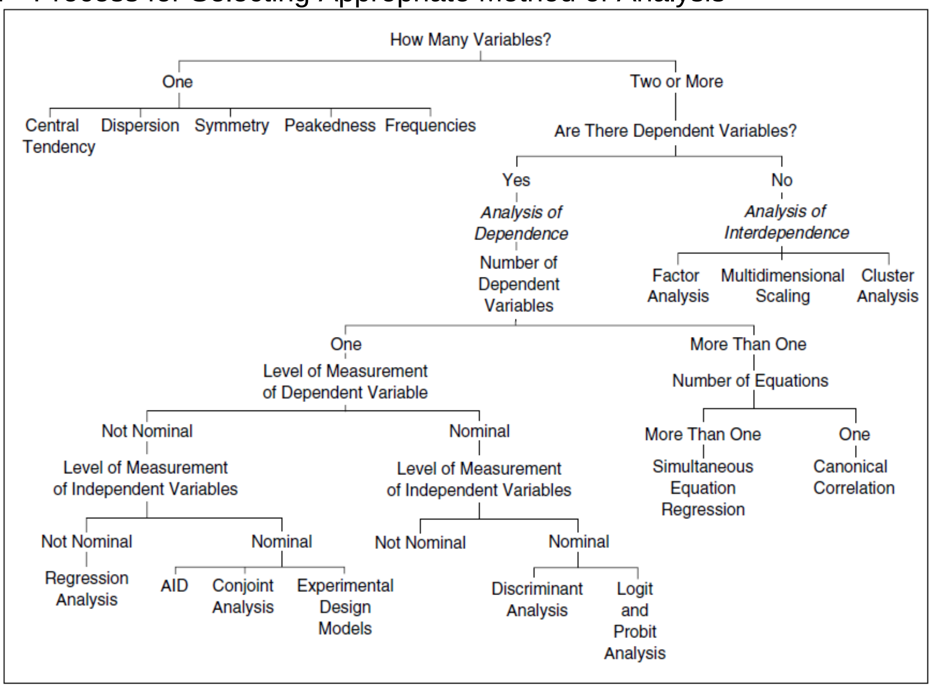
The raw input to any analysis of associative data consists of the *data matrix*, whose informational content is to be summarized and portrayed in some way. For example, the computation of the mean and standard deviation of a single variable is often done simply to obtain a summary of the meaning of the entire set of values. In so doing, we choose to forgo the full information provided by the data in order to understand some of its basic characteristics, such as central tendency and dispersion.

In virtually all marketing research studies we are concerned with variation in some characteristic, be it per capita consumption of soft drinks, TV viewing frequency, or customer intention to repurchase. Our objective now, however, is to concentrate on explaining the variation in one variable or group of variables in terms of *covariation* with other variables. When we analyze associative data, we hope to explain variation, achieving one or more of the following:

1. Determine the overall strength of association between the *criterion* and *predictor* variables (often called *dependent* and *independent* variables, respectively)
2. Determine a function or formula by which we can estimate values of the criterion variable(s) from values of the predictor variable(s)
3. Determine the statistical confidence in either or both of the above

In some cases of interest, however, we may have no criterion (dependent) variable. We may still be interested in the *interdependence* of a group of variables as a whole and the possibility of summarizing information provided by this interdependence.

Figure 13.1 Process for Selecting Appropriate Method of Analysis



A Classification of Techniques for Analyzing Associative Data

The field of associative data analysis is vast; hence, it seems useful to enumerate various descriptors by which the field can be classified. A conceptual illustration is shown in Table 13.1. The key notion underlying our classification is the data matrix. We note that a data matrix is often thought of as being in spreadsheet form (referred as a flat data matrix) and consists of a set of objects (the *n* rows) and a set of measurements on those objects (the *m* columns). The row objects may be people, things, concepts, or events. The column variables are characteristics of the objects. The cell values represent the state of object *i* with respect to variable *j*. Cell values may consist of nominal-, ordinal-, interval-, or ratio-scaled measurements, or various combinations of these as we go across columns. When using data analysis software, the data file created is structured exactly like the basic data matrix, which is, in effect, a type of spreadsheet. Note however that some analysis software also has the ability to read relational databases that are not in the traditional row x column spreadsheet format. However for relational database format data, the concepts of objects and variables still apply.

Table 13.1 Illustrative Basic Data Matrix

	Variable						
Object	*1*	*2*	*3*	. . .	*j*	. . .	*m*
1	X_{11}	X_{12}	X_{12}	. . .	X_{1j}	. . .	X_{1m}
2	X_{21}	X_{22}	X_{23}	. . .	X_{2j}	. . .	X_{2m}
3	X_{31}	X_{32}	X_{33}	. . .	X_{3j}	. . .	X_{3m}
.	.	.	.		.		.
.	.	.	.		.		.
.	.	.	.		.		.
i	X_{i1}	X_{i2}	X_{i3}	. . .	X_{ij}	. . .	X_{im}
.	.	.	.		.		.
.	.	.	.		.		.
.	.	.	.		.		.
n	X_{n1}	X_{n2}	X_{n3}	. . .	X_{nj}	. . .	X_{nm}

By reflecting on the steps in the research process discussed in Chapter 2, we can gain a perspective on the many available approaches to analysis of associate data. Although not exhaustive (or exclusive), the following represent the more common perspectives by which analysis activities can be directed:

1. Purpose of the study and types of assertions desired by the researcher: What kinds of statements—descriptive or inferential—does the researcher wish to make?
2. Focus of research: Is the emphasis on the objects (the whole profile or “bundle” of variables), the variables, or both?
3. Nature of the prior judgments assumed by the researcher as to how the data matrix should be partitioned (subdivided) into a number of subsets of variables.
4. Number of variables in each of the partitioned subsets: How many criterion versus predictor variables?
5. Type of association under study: Is the association linear, transformable to linear, or nonlinear?
6. Scales by which variables are measured: Are the scales nominal, ordinal, interval, ratio, or mixed?

All of these descriptors require certain decisions of the researcher, three of which are of immediate interest in determining the form of analysis:

1. Whether the objects or variables of the data matrix are partitioned into subsets or kept intact
2. If the matrix is partitioned into subsets of criterion and predictor variables, the number of variables in each subset
3. Whether the variables are nominal-scaled or interval-scaled
4. The relationships between the variables (linear, transformation to linear, or non-linear).

Most decisions about associative data analysis are based on the researcher’s private model of how the data are interrelated and what features are useful for study. The choice of various public models for analysis (multiple regression, discriminant analysis, etc.) is predicated upon prior knowledge of both the characteristics of the statistical universe from which the data were obtained and the assumption structure and objectives associated with each statistical technique.

Fortunately, we can make a few simplifications of the preceding descriptors. First, concerning types of scales, all the multivariate techniques of this book require no stronger measurement than interval scaling. Second, except for the ordinal scaling methods of multidimensional scaling and conjoint analysis (to be discussed in Chapter 13), we shall assume that (a) the variables are either nominal-scaled or interval-scaled, and (b) the functional form is linear in the parameters. While the original data may have been transformed by some nonlinear transformation (e.g., logarithmic), *linear in the parameters* means that all computed parameters *after* the transformation are of the first degree. For example, the function $Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_mX_m$ is linear in the parameters. Note that b_0 is a constant, while the other parameters, $b_1, b_2, \ldots, b_m$ are all of the first degree. Moreover, none of the b_j's depends on the value of either its own X_j or any other X_k ($k \neq j$). Even with these simplifying assumptions, we shall be able to describe a wide variety of possible techniques, and these may vary from being relatively simple, (as shown in Exhibit 13.1), to complex.

Analysis of Dependence

If we elect to partition the data matrix into criterion and predictor variables, the problem becomes one of analyzing *dependence structures*. This, in turn, can be broken down into subcategories:

1. Single criterion/multiple predictor association
2. Multiple criterion/multiple predictor association

Multivariate techniques that deal with single criterion/multiple predictor association include multiple regression, analysis of variance and covariance, two-group discriminant analysis, and CHAID (chi-square automatic interaction detection).

Multivariate techniques that deal with multiple criterion/multiple predictor association include canonical correlation, multivariate analysis of variance and covariance, and multiple discriminant analysis.

Analysis of Interdependence

In some cases, we may not wish to partition the data matrix into criterion and predictor subsets. If so, we refer to this case as the analysis of *interdependence structures*. Techniques such as factor analysis are used if the focus of interest is on the variables of the (intact) data matrix. Cluster analysis is relevant if we wish to focus on the grouping of objects in the data matrix, as based on their profile similarities.

Selecting a Technique

How does a researcher select the multivariate technique to use in a given situation? In addition to the research question underlying the analysis, we suggested in Figure 13.1 that the number of dependent and independent variables and their levels of measurement are two of the major criteria. Four major sets of factors influence the choice of an appropriate statistical technique:

- Research parameters (analysis purpose and specifications)
- Known characteristics of the data
- Properties of the statistical technique
- Researcher characteristics

Research parameters include the purpose of analysis and the kind of analysis required. Before the analysis begins, certain characteristics of the data are known, including the level of measurement, distribution of the data, and the nature and size of the sample. This information is matched with the properties of a particular statistical technique (robustness, assumptions required, and design purpose). Finally, a researcher's individual philosophy, expertise, and experience will influence the choice of statistical technique.

Operationally, the researcher can be assisted by any of numerous printed or online guides, such as that provided at http://www.qualtrics.com/wiki/index.php/Analysis.

MULTIPLE AND PARTIAL REGRESSION

Multiple regression extends the concepts of bivariate regression by simply adding multiple predictor (independent) variables. To illustrate this extension, we will introduce the second predictor $X2$. Table 13.2 presents a set of example data with two predictor variables.

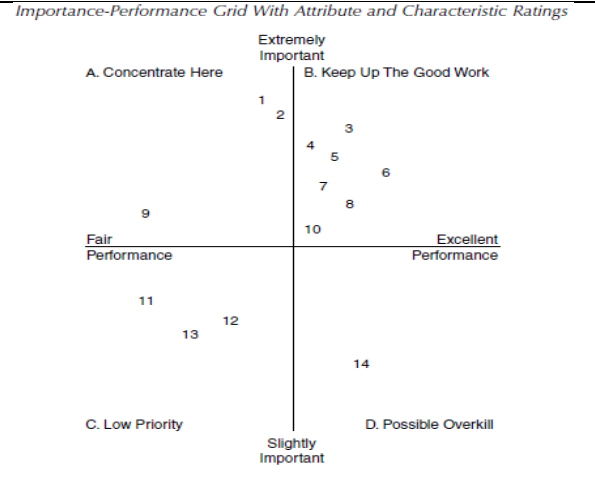
Table 13.2 Consumer Preference Ratings of Ten Cereals Varying in Nutritional Level

Rater	*Preference*	*Protein, X_1*	*Vitamin D, X_2*
1	3	4	2
2	7	9	7
3	2	3	1
4	1	1	2
5	6	3	3
6	2	4	4
7	8	7	9
8	3	3	2
9	9	8	7
10	2	1	3
Total	43	43	40
Mean	4.3	4.3	4.0
Standard deviation	2.908	2.791	2.708

EXHIBIT 13.1 Multivariate Analysis Can Be Simple

Not all multivariate analyses need to involve complex and sophisticated statistical techniques. A relatively simple analysis is importance-performance analysis. In the B2B area, for example, this could be done by asking customers to rate a supplier's performance on a set of attributes and characteristics, and then asking those customers how important each is to them. A central tendency measure would be calculated for each attribute's importance and performance measure and then plotted in two-dimensional space, as illustrated in the following figure, which shows a hypothetical grid for 14 attributes/characteristics.

Importance-Performance Grid With Attribute and Characteristic Ratings



The theoretical model now becomes

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

with parameters estimated by

$$\hat{Y} = a + b_1 X_1 + b_2 X_2$$

All assumptions previously discussed for the bivariate regression model continue to hold in the present case. However, it is important to remember that, in general, the current b_1— now called a *partial* regression coefficient with respect to X_1—will *not* equal its counterpart coefficient (b) obtained from the bivariate regression. This is because X_1 itself will usually be correlated with X_2. In the bivariate case, X_2 was not entered into the analysis and any variation in Y that was shared by X_1 and X_2 was credited solely to X_1. Such will no longer be the case. When there are only two predictors, X_1 and X_2, it is still possible to prepare a scatter plot.

However, now the plot appears in terms of a three-dimensional space. In this space we fit a plane whose sum of squared deviations of the Yi from their $\widehat{Y_i}$ counterparts (on the plane) is a minimum. Figure 13.3 shows the original scatter plot and the fitted plane for the sample problem of Table 13.2. Using regression analysis programs, the estimated (least-squares) equation is
$Y = 0.247 + 0.493\ X_1 + 0.484\ X_2$

Figure 13.3 Scatter Plot and Fitted Regression Plane

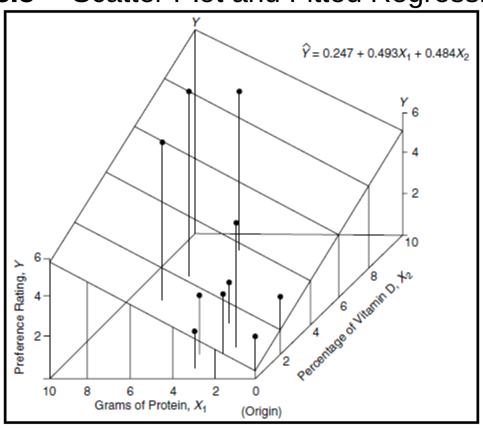


Table 13.3 Single Correlation Between *X*1 (Protein) and *X*2 (Vitamin)

		Protein	Vitamin
Protein	*Pearson Correlation*	1	.838**
	Sig. (2-tailed)	.	.002
	N	10	10
Vitamin	*Pearson Correlation*	.838**	1
	Sig. (2-tailed)	.002	.
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

The first slope coefficient of 0.493 denotes the change in $\widehat{Y_i}$ per unit change in X_1 when X_2 is held constant. Similarly, 0.484 denotes the change in $\widehat{Y_i}$ per unit change in X_2 when X_1 is held constant. We note that $0.493 \neq 0.886$, the slope obtained earlier in the bivariate case discussed in Chapter 11. This is because X_1 and X_2 are themselves correlated, and X_1 is now forced to share some of its Y-variable association with the second predictor, X_2.

Parameter Estimation

Finding the partial regression coefficients $b_1 = 0.493$ and $b_2 = 0.484$ and the intercept $a = 0.247$ is considerably more complicated than finding a single regression coefficient in bivariate regression. Exhibit 13.2 shows the statistical analysis for a regression analysis. This result can be replicated with any statistical package or a spreadsheet multiple regression function.

EXHIBIT 13.2 Summary Output of Regression Analysis (All Variables): Sample Data

Variables Entered/Removed

Model	*Variables Entered*	*Variables Removed*	*Method*
1	VITAMIN, PROTEIN[a]	.	Enter

All requested variables entered.
Dependent Variable: PREFERENCE

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.885[a]	.783	.721	1.53482

Model	Change Statistics R Square Change	F Change	df1	df2	Sig. F Change
1	.783	12.652	2	7	.005

Predictors: (Constant), VITAMIN, PROTEIN

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	59.610	2	29.805	12.652	.005[a]
	Residual	16.490	7	2.356		
	Total	76.100	9			

Predictors: (Constant), VITAMIN, PROTEIN
Dependent Variable: PREFERENCE

Coefficients

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	.247	.942		0.262	.801
	PROTEIN	.493	.336	.473	1.466	.186
	VITAMIN	.484	.346	.450	1.397	.205

A. Dependent Variable: PREFERENCE

The Coefficient of Multiple Determination, R^2

Finding the *coefficient of multiple determination* proceeds in the same way as the bivariate case. Exhibit 13.2 shows this value as: $R^2_{Y,X_1,X_2} = 0.783$.

We note that the multiple regression equation does quite well at prediction, since 78 percent of the variation in Y is accounted for by variation in X_1 and X_2 (R^2_{Y,X_1,X_2}). However, we recall from Chapter 12 that if only X_1 is employed, the (simple) coefficient of determination is 72 percent. Apparently, X_2 (vitamin) and X_1 (protein) are so highly correlated—their simple correlation is 0.838 (see Table 13.3)—that once X_1 is in the regression equation there is little need for X_2 as well.

What about R_{Y,X_1,X_2} the *square root* of R^2_{Y,X_1,X_2}?

$$R_{Y,X_1,X_2} = .885 = \sqrt{.783} = \sqrt{R^2_{Y,X_1,X_2}}$$

This is called the *coefficient of multiple correlation:* and is interpreted as the *simple correlation* (i.e., bivariate correlation) between Y_i and $\widehat{Y}_i$ where, as we know, the $\widehat{Y}_i$ are the *predicted* values obtained by the best linear composite of X_1 and X_2 in the least-squares sense; this composite is given by the multiple regression equation.

Standard Errors and *t*-Tests

While the F-test indicates that the overall regression model is significant, it does not follow that *both* b_1 and b_2 contribute significantly to overall accounted-for variance. It may be the case that a simpler model involving only X_1 (or only X_2) would be sufficient. The standard error of each individual regression coefficient provides the basis for a *t*-test of this simpler model.

The standard error is a measure of dispersion about the average partial regression coefficient over repeated samplings of Y for a fixed set of values on each of the predictors. The larger the standard error is, the less reliable our estimate of b_1 is (across repeated samplings from the same universe). Conceptually, this statement means that the standard error of b_1 increases as X_1 becomes more completely accounted for by the remaining predictor variables.

The *t*-value of 1.466 in Exhibit 13.2 is simply the ratio of b_1 to its own standard error, SE(b_1) = 4.73/3.36. The test of significance of t_1 is carried out by finding the tabular value of the *t*-distribution (Table A.2 in Appendix A) for seven degrees of freedom. If we continue to use a significance level of 0.1, then the tabular value of t is 1.415 and b_1 is significant. However, b_2 (whose *t*-value is 1.397) is *not* significant at the 0.1 level. Note that in Exhibit 13.2, the two variable (protein and vitamin) regression analysis reports protein to be nonsignificant with a probability of .186. This is due to the degrees of freedom issue used in the two variable regression. When considered alone, the *t*-test for protein is significant at the .1 level.

The *t*-test for each individual partial regression coefficient tests whether the increment in R^2 produced by the predictor in question is significant when a model including the predictor (and all other predictors) is compared with a model including all predictors but the one being tested.

Contribution to Accounted-for Variation

The coefficient of multiple determination in this example indicates that the variation of both independent variables accounted for 78.3 percent of the variation in the dependent variable. Since b_2 is not significant, we can presume that X_2 contributes little to the dependent variable's variation. In fact, this is the case. When a stepwise procedure was used for the regression analysis, X_1 was the only variable included and $R^2 = 0.723$. Thus X_2's contribution to the coefficient of multiple determination is only 0.06 (0.783–0.723).

In summary, Exhibit 13.2 shows the major output measures of interest in applied multiple regression studies:

1. The regression equation
2. R^2—both the sample-determined and the population-adjusted values
3. An F-test for testing the significance of the overall regression (involving both X_1 and X_2)

4. Individual t-tests and standard errors for testing each specific partial regression coefficient
5. Partial correlation coefficients
6. The accounted-for variance contributed by each predictor, where any shared variance of each predictor (beyond the first) is credited to the predictors that precede it, based on the researcher's order for including the predictor variables

In the sample problem R^2 was highly significant, but only the first variable X_1 was needed to account for most of the variance in Y. That is, the addition of X_2 accounted for an incremental variance of only six percentage points; moreover, the t-test for b_2 was not significant at the 0.1 level. Therefore, in practice we would employ a simple regression of Y on X_1 alone. This has already been computed in Chapter 11 as $\hat{Y} = 0.491 + 0.886X_1$.

Other Forms of Regression

With the large variety of computer programs available, the computation of multiple regression analysis is both fast and easy for the user. However, decisions must still be made by the user. Within a package, individual programs may differ with respect to the criterion for including the independent variables in the regression, the ability to repeat the analysis on subgroups of respondents and compare the subgroups, the types of residual analysis available, and the linearity assumption used in the analysis.

The two most widely used approaches are all-variable regression (demonstrated above) and stepwise regression, as described further in Exhibit 13.3 and shown in Exhibit 13.4. Many items shown in an output are common to both procedures, though not necessarily in the same format or sequence. Similarly, the results of the regression analyses are often similar for a small number of variables, becoming increasingly disparate as more variables are introduced. This disparity is explained by the fact that the rationale for including variables differs greatly. For more detail contrasting the difference between the all-variable and stepwise regression methodologies, see Green, Tull, and Albaum (1988, pp. 443–55).

EXHIBIT 13.3 Stepwise Regression Can Be Hazardous

Avoid using stepwise regression whenever possible, particularly when dealing with sample data. The theoretical and statistical shortcomings are inherent in the process and difficult to overcome (Curtis, 1990). If you do use a stepwise approach, however, certain precautions should be taken:

1. Limit the number of variables initially included in the predictor pool, eliminating some with high intercorrelations.
2. Use as large a sample as possible, ideally with 40–50 cases for each variable in the pool.
3. Validate the model through a follow-up sample or by splitting the original sample.
4. When multicollinearity is a problem, consider using factor analysis as a data reduction technique.

Stepwise regression is a computer-driven statistical procedure in which a subset of independent variables is selected from a larger pool of variables through a series of repetitive steps to arrive at the best-fitting model for the data. "Best" typically refers to maximizing the explained variance (R^2) of the dependent variable.

Stepwise regression begins with a dependent variable and a pool of independent variables. It starts building a model by selecting the variable most highly correlated with the dependent variable. The program then looks for the next variable in the pool that offers the largest increment in R^2 when added to the model. As new variables are added, others may be dropped in a repetitive process until the best-fitting model is produced.

But the method has serious implications. First, from a *substantive* perspective, the best-fitting model may not be the best—or even good. There are other variable selection techniques in the stepwise family, such as forward entry and backward elimination, which also are designed to obtain the best-fitting model. It would not be unusual to obtain three completely different bestfitting models from the same data set using these methods. The only time this is unlikely to happen is in the unusual situation when all predictor variables are uncorrelated with one another.

Another problem occurs when high levels of multicollinearity exist in the pool of independent variables. A single variable may be included, while two or more others that combine to make a more accurate predictive model may be excluded. The greater the multicollinearity, the more likely an unreasonable and unstable best-fitting model is selected.

Stepwise methods also may cause a problem when used for explanatory model building, which focuses on the causal relationships between the dependent and independent variables. The problem is that stepwise methods are purely data-driven, after-the-fact techniques not based on a prior theoretical model. Because data are typically nothing more than a snapshot of reality at any given time, and subject to significant fluctuation, theory is ultimately important to understand the true, underlying relationships at work.

EXHIBIT 13.4 Summary Output of Stepwise Regression Analysis

Sample Data from Table 13.2

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	VAR00002	.	Stepwise (Criteria: Probability of F to enter <=.050, Probability of F to remove>=.100).

Dependent Variable:VAR00001

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.850[a]	.723	.688	1.62354

Model	Change Statistics R Square Change	F Change	df1	df2	Sig. F Change
1	.723	20.871	1	8	.002

Predictors: (Constant), VAR00002

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	55.013	1	55.013	20.871	.002[a]
	Residual	21.087	8	2.636		
	Total	76.100	9			

Predictors: (Constant), VAR00002
Dependent Variable: VAR00001

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.491	.979		.501	.630
	VAR00002	.886	.194	.850	4.568	.002

Dependent Variable: VAR00001

Excluded Variables

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	VAR00003	.450[a]	1.397	.205	.467	.298

Predictors in the Model: (Constant), VAR00002
Dependent Variable: VAR00001

Multicollinearity

Put simply, *multicollinearity* refers to the problem in applied regression studies in which the predictor variables exhibit very high correlation among themselves. This condition distorts the value of the estimated regression coefficients, inflates the standard error of beta, and thereby makes it more difficult to determine which predictor variable is having an effect.

Unless one is dealing with experimental design data, it is almost always the case that predictor variables in multiple regression will be correlated to some degree. The question is how much multicollinearity can be tolerated without seriously affecting the results? Unfortunately, there is no simple answer to this question.

The study of multicollinearity in data analysis revolves around two major problems: (a) How can it be detected; and (b) what can be done about it? These problems are particularly relevant to marketing research, where one often faces the dilemma of needing a large number of variables to achieve accuracy of predictors, and yet finds that as more predictors are added to the model, their intercorrelations become larger.

As indicated above, what constitutes serious multicollinearity is ambiguous. Some researchers have adopted various rules-of-thumb: For example, any pair of predictor variables must not correlate more than 0.9; if so, one of the predictors is discarded. While looking at simple correlations between pairs of predictors has merit, it can miss more subtle relationships involving three or more predictors. The above rule can be extended, of course, to the examination of *multiple* correlations between each predictor and all other predictors. Usually one would want to guard against having any of these multiple correlations exceed the multiple correlation of the criterion variable with the predictor set.

Essentially there are three procedures for dealing with multicollinearity:

1. Ignore it.
2. Delete one or more of the offending predictors.
3. Transform the set of predictor variables into a new set of predictor-variable combinations that are mutually uncorrelated.

Ignoring multicollinearity need not be as cavalier as it might sound. First, one can have multicollinearity in the predictor variables and still have strong enough effects that the estimating coefficients remain reasonably stable. Second, multicollinearity may be prominent in only a subset of the predictors, a subset that may not contribute much to accounted-for variance anyway. A prudent procedure in checking one's predictor set for multicollinearity is to examine the standard errors of the regression coefficients (which will tend to be large in the case of high multicollinearity). Second, one may randomly drop some subset of the cases (perhaps 20 percent or so), rerun the regression, and then check to see if the signs and relative sizes of the regression coefficients are stable. Third, most computer regression routines incorporate checks for serious multicollinearity (see SPSS Linear Regression Analysis Statistics Help for a list of diagnostic analyses); if the program does not indicate this condition, the researcher can generally assume that the problem is not acute.

If multicollinearity is severe, one rather simple procedure is to drop one or more predictor variables that represent the major offenders. Usually, because of their high intercorrelations with the retained predictors, the overall fit will not change markedly. Pragmatically, if a particular pair of predictors is highly collinear, one would retain that member of the pair whose measurement reliability or theoretical importance is higher in the substantive problem under study.

Methods also exist for transforming the original set of predictors to a mutually uncorrelated set of linear composites—for example, principal components analysis (which is discussed in Chapter 14). If these components (linear composites) are interpretable in themselves, the researcher may use these in the regression analysis rather than the original variables. If *all* components are retained, the predictive accuracy will be precisely the same as that obtained from the original set of predictors. However, the problem here is that the components may *not* be interpretable in their own right. Another possibility, of course, is to use only one of the variables for each component to represent that component in the regression analysis.

Cross-validation

Probably the safest procedure for dealing with the variety of problems in multiple regression, including multicollinearity, is to use *cross-validation*. We have commented in this chapter on the tendency of regression models (and the same is true of other multivariate techniques) to capitalize on chance variation in the sample data. Since these techniques are optimization methods, they find the best possible fit of the model to the *specific* data, though one can also almost invariably find a poorer fit.

Cross-validation is a simple procedure for examining whether the regression equation holds up beyond the data on which its parameters are based. The researcher simply takes part of the data (perhaps a quarter to a third) and puts it aside. The regression equation is then computed from the remaining data. Following this, the researcher takes the "hold-out" data and computes a set of Yi using the earlier-computed regression equation and the predictor-variable values of the hold-out sample. The researcher then finds the simple coefficient of determination between the Yi in the hold-out sample and their predicted Yi counterparts. This coefficient is compared with the R^2 obtained from the original analysis to find the degree of shrinkage. Doing this requires a larger sample size.

An even better procedure is to *double cross-validate*, using the following steps:

1. Split the cases randomly into halves.
2. Compute *separate* regression equations for each half.
3. Use the first-half equation to predict the second-half Yi values.
4. Use the second-half equation to predict the first-half Yi values.
5. Examine each partial regression coefficient across split halves to see if agreement is obtained in both directions (algebraic sign) and in magnitude.
6. Compute a regression equation for the entire sample, using only those variables that show stability in the preceding step.

Since high multicollinearity will make sample-to-sample regression coefficients unstable, double cross-validation can help the researcher determine which coefficients exhibit stability across split halves.

Depending on the sample size, of course, one could split the sample into thirds, quarters, and so on. Usually, however, there is sufficient constraint on sample size, relative to the number of predictors, that split-half testing is about all that gets done. Even so, single and double cross-validation are extremely useful undertakings, and, with a few clicks of a mouse, are easy to implement.

Steckel and Vanhonacker (1991) developed a formal test for the cross-validation of regression models under the simple random splitting framework. Their results indicated that splitting the data into halves is suboptimal. They recommend that more observations should be used for estimation than validation.

TWO-GROUP DISCRIMINANT ANALYSIS

When dealing with associative data, the marketing researcher may encounter cases where the criterion variable is categorical, but where the predictor variables involve interval-scaled data. For example, one may wish to predict whether sales potential in a given marketing territory will be good or bad, based on certain measurements regarding the territory's personal disposable income, population density, number of retail outlets, and the like. Other potential applications also come to mind:

Two Groups:

- How do consumers who are heavy users of my brand differ in their demographic profiles from those who are nonusers?
- How do respondents who show high interest in a new set of concept descriptions differ in their readership levels of certain magazines from those who show low interest?
- How do homeowners who select a variable rate mortgage differ in their demographic profiles, mortgage shopping behavior and attitudes, and preferences for mortgage features from homeowners selecting a conventional fixed-rate mortgage?

Multiple Groups:

- Are significant demographic differences observed among purchasers of Sears, Goodyear, Goodrich, and Michelin tires?
- How do doctors, lawyers, and bankers differ in terms of their preference ratings of eight different luxury automobiles?
- How can loyal shoppers of Nordstrom's, Macy's, and Neiman-Marcus be distinguished on the basis of their attitudes about each retailer?

Many other such problems could be added to the list. However, each one has a common structure in which we assume that some test object (usually a person) falls into one of a set of categories. It is also assumed we know that person's profile on a set of (assumed or otherwise) interval-scaled predictor variables, such as age, income, years of education, or other background variables.

The problem is to predict a person's category from some function of the predictor variables. Here we shall assume that the function is linear. If only two categories are involved, the problem is a *two-group* discriminant case; if three or more categories are involved, we are dealing with *multiple* (group) discriminant analysis.

Objectives of Two-Group Discriminant Analysis

Two-group discriminant analysis (and classification) involves four main objectives:

1. Finding linear composites of the predictor variables that enable the analyst to separate the groups by maximizing among-groups (relative to within-groups) variation
2. Establishing procedures for assigning new individuals, whose profiles (but not group identity) are known, to one of the two groups

3. Testing whether significant differences exist between the mean predictor-variable profiles of the two groups
4. Determining which variables account most for intergroup differences in mean profiles

These objectives are the bases for very distinct, purposes and procedures for conducting discriminant analysis (Smith, 2004). The first procedure, *discriminant predictive* (or *explanatory*) *analysis*, is used to optimize the predictive functions.

The second procedure, *discriminant classification analysis*, uses the predictive functions derived in the first procedure to either classify fresh sets of data of known group membership, thereby validating the predictive function, or if the function has previously been validated, to classify new sets of observations of unknown group membership. The differences in the application and requirements for each are summarized in Table 13.4.

Table 13.4 Stages of Discriminant Analysis

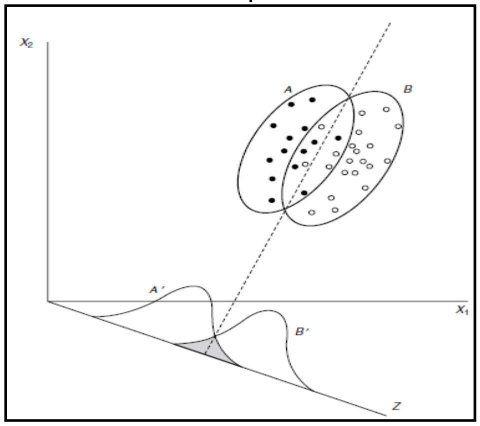
Stages	*Predictive discriminant analysis*	*Classification analysis of initial data set of known groupings*	*Classification analysis of new data set of known groupings*	*Classification analysis of new data set of unknown groupings*
Purpose	Derive discriminant function using initial data set. No classification involved	Determine how well discriminant function classifies (biased)	(1) Classify data using classification rule derived from predictive function (2) May be part of validation analysis of initial predictive function	(1) Classify data using classification rule derived from predictive function (2) May be part of validation analysis of initial predictive function
Assumptions of linear Discriminant model. Requirements	No validation required	No validation required	Validation required	Initial predictive function must have been previously validated

Geometric Representation

If we have *n* persons measured on *m* variables, the profile of each person can be portrayed as a point in *m* dimensions. If we *also* know the group to which each person belongs, and the groups differ in terms of average profiles (often called *centroids*), we might expect to find different groups occupying different regions of the space. The less overlap noted among intergroup profiles in that space, the more likely it is that discriminant analysis can help us separate the groups.

The scatter diagram and projection in Figure 13.4 provides one way to show what happens when a two-group discriminant function is computed. Suppose that we had two groups, *A* and *B*, and two measures, X_1 and X_2, on each member of the two groups. We could plot in the scatter diagram the association of variable X_2 with X_1 for each group, maintaining group identity by the use of filled-in dots or open circles. The resultant ellipses enclose some specified proportion of the points, say 95 percent, in each group. If a straight line is drawn through the two points where the ellipses intersect and is then projected to a new axis *Z*, we can say that the overlap between the *univariate* distributions A' and B' (represented by the shaded area) is smaller than would be obtained by any other line drawn through the ellipses representing the scatter plots. The shaded region under the curves can also be interpreted as representing the probabilities of mistakes when classification is done on the basis of likelihoods only.

Figure 13.4 Graphical Illustration of Two-Group Discriminant Analysis



The important thing to note about Figure 13.4 is that the Z-axis expresses the two-variable profiles of groups A and B as *single* numbers. That is, by finding a linear composite of the original profile scores we can portray each profile as a point on a line. Thus, the axis Z condenses the information about group separability (shown in the bivariate plot) into a set of points on a single axis. Here, Z is the *discriminant* axis.

In most problems, we have more than two predictor variables. If so, each predictor can represent a *separate* dimension (although we would be limited to three predictor variables if we wished to plot the data). In any case, the basic objective is still to find one axis in the m-dimensional space that maximally separates the centroids of the two groups after the points are projected onto this new axis.

In our discussion of multiple regression analysis we noted that one finds a linear composite that maximizes the coefficient of multiple determination, R^2. Analogously, in two-group discriminant analysis we try to find a linear composite of the original variables that maximizes the *ratio* of among-to-within groups variability. It should be noted that if m, the number of predictor variables, is quite large, we shall be parsimonious by portraying among-to-within groups variation in many fewer dimensions (actually a *single dimension* in the two group case) than found originally.

A Numerical Example

Let us return to the example involving ready-to-eat cereals that was first presented in Table 13.2 in the context of multiple regression. As recalled, we wished to see if amount of protein and vitamin D influenced consumers' evaluations of the cereals.

In the present case we shall assume that each of the ten consumer raters is simply asked to classify the cereal into one of two categories: *like* or *dislike*. The (hypothetical) data appear in Table 13.5, which has two predictor variables (again):

X_1: the amount of protein (in grams) per two-ounce serving
X_2: the percentage of minimum daily requirements of vitamin D per two-ounce serving

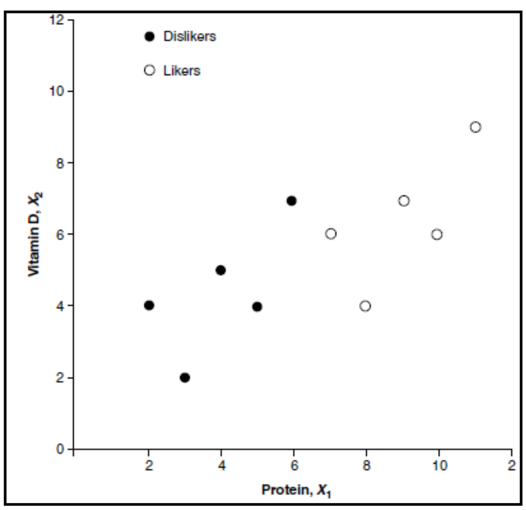
We first note in the table that the two groups are much more widely separated on X_1 (protein) than they are on X_2 (vitamin D). If we were forced to choose just one of the axes, it would seem that X_1 is a better bet than X_2. However, there is information provided by the group separation on X_2, so we wonder if some linear composite of both X_1 and X_2 could do better than X_1 alone.

Table 13.5 Consumer Evaluations (Like Versus Dislike) of Ten Cereals Varying in Nutritional Content

Person	*Evaluation*	*Protein* X_1	*Vitamin D,* X_2
1	Dislike	2	4
2	Dislike	3	2
3	Dislike	4	5
4	Dislike	5	4
5	Dislike	6	7
Mean		**4**	**4.4**
6	Like	7	6
7	Like	8	4
8	Like	9	7
9	Like	10	6
10	Like	11	9
Mean		**9**	**6.4**
Grand Mean		6.5	5.4
Std Deviation		3.028	2.011

Figure 13.5 shows a scatter plot of the X_1 and X_2 data of Table 13.5. We note that perfect discrimination can be achieved with X_1 if we erect a line perpendicular to the horizontal axis between the scale values of 6 and 7. On the other hand, there is no way that the use of X_2 alone would enable us to separate the groups. Given this picture, we would not be surprised if the best linear composite places a considerably larger weight on X_1 than on X_2.

Figure 13.5 Scatter Plot of Two-Group Sample Data of Table



Why not use X_1 alone, rather than a composite of X_1 and X_2? First, the data of Table 13.5 represent only a *sample*; it is quite possible that additional observations would show that X_1 alone would *not* effect perfect discrimination between the two groups. Second, we have not explicitly taken into consideration either the variability of X_1 versus X_2 or their correlation. One of the nice features of discriminant analysis is that all three aspects of the data— centroid, variance, and correlation—are considered when developing the linear composite that maximally separates the groups.

As noted earlier, the key problem of two-group discriminant analysis is to find a new axis so that projections of the points onto that axis exhibit the property of maximizing the separation between group means relative to their within-groups variability on the composite.

This discriminant axis can be defined in terms of a set of weights—one for each predictor variable axis—so that we have the following linear function:

$$Z = d_1X_1 + d_2X_2$$

where d_1 and d_2 are the weights that we seek.

Using a discriminant analysis program, we solve for the discriminant weights d_1 and d_2 that maximize the separation between the groups. This involves a procedure where two simultaneous equations are solved, leading to the desired discriminant function:

$$Z = 0.368X_1 - 0.147X_2$$

The weights d_1 and d_2 are known and labeled on computer output as *unstandardized discriminant function coefficients*. The general form for a linear discriminant function is $Z = d_0 + d_1X_1 + d_2X_2$, where d_0 is the intercept or constant term in the function. Since the example used deviations to calculate the weights d_1 and d_2, the function goes through the origin and $d_0 = 0$. Having found the discriminant function, it is a straightforward procedure to find discriminant scores for the centroids of the two groups and the grand mean:

$$\begin{aligned} \bar{Z}(\text{dislikers}) &= 0.368(4) - 0.147(4.4) = 0.824 \\ \bar{Z}(\text{likers}) &= 0.368(9) - 0.147(6.4) = 2.368 \\ \bar{Z}(\text{grand mean}) &= 0.368(6.5) - 0.147(5.4) = 1.596 \end{aligned}$$

We may also apply this procedure to each of the 10 pairs of X_1, X_2 values in Table 13.5 to get the linear composite. For example, the discriminant score for the first case in the disliker group is

$$Z = 0.368(2) - 0.147(4) = 0.148$$

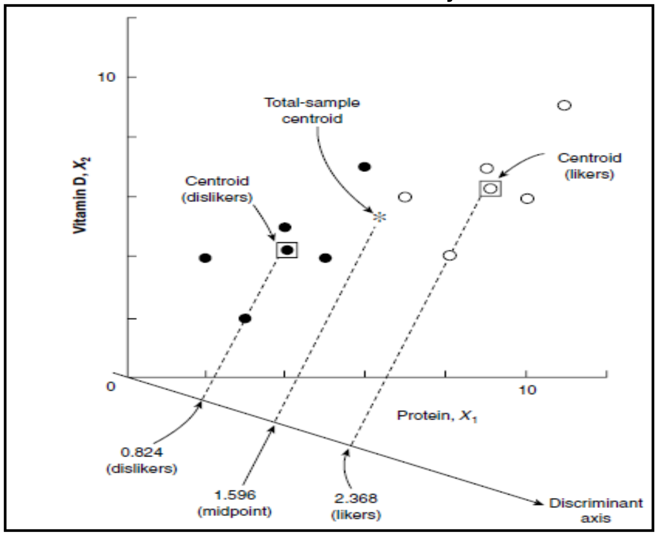
Plotting the Discriminant Function

The original scatter plot of the ten observations is reproduced in Figure 13.6. However, this time we also show the discriminant axis (linear composite) by passing a straight line through the point (0.368, – 0.147) and the intersection of the original axes (i.e., the origin).

Alternatively, one could use any point with coordinates proportional to (0.368, –.147). However, it should be noted that the *scale unit* on the discriminant axis in Figure 13.6 differs from the original unit in which X_1 and X_2 are expressed. To maintain the original unit, d_1 and d_2 would have to be normalized. The original points can then be projected onto this new axis.

To illustrate, we show projections of the grand centroid and the centroids of the dislikers and likers, respectively. (Similarly, all ten original points could be projected onto the discriminant axis as well.) We note that the discriminant axis favors X_1 (as we guessed it would) by giving about 2.5 times the (absolute-value) weight ($d_1 = 0.368$ versus $d_2 = -0.147$ to X_1) as is given to X_2.

Figure 13.6 Plot of the Discriminant Axis and Point Projections



The analysis output is shown in Exhibit 13.5. Note that the discriminant function coefficients are different from those shown earlier, meaning that the discriminant scores for each person also will differ. This is because the discriminant axis is *not run through the origin*, but rather there is an intercept term. The unstandardized discriminant function coefficients evaluated at group means sum to zero. The analysis is not affected by this, and all conclusions to be made will not differ whether or not the axis is through the origin.

EXHIBIT 13.5 Discriminant Analysis for Cereal Evaluations

Group Statistics

EVALUATE		Mean	Std. Deviation	Valid *N* (listwise) Unweighted	Weighted
GRP 1	PROTEIN	4.0000	1.58114	5	5.000
	VITAMIN	4.4000	1.81659	5	5.000
GRP 2	PROTEIN	9.0000	1.58114	5	5.000
	VITAMIN	6.4000	1.81759	5	5.000
Total	PROTEIN	6.5000	3.02765	10	10.000
	VITAMIN	5.4000	2.01108	10	10.000

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
PROTEIN	.242	25.000	1	8	.001
VITAMIN	.725	3.030	1	8	.120

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	3.860[a]	100.0	100.0	.891

First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-Square	df	Sig.
1	.206	11.068	2	.004

Standardized Canonical Discriminant Function Coefficients

	Function
	1
PROTEIN	1.323
VITAMIN	–.608

Structure Matrix

	Function
	1
PROTEIN	.900
VITAMIN	.313

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation within function.

Canonical Discriminant Function Coefficients

	Function
	1
PROTEIN	.837
VITAMIN	–.335
(Constant)	–3.632

Unstandardized coefficients

Functions at Group Centroids

	Function
EVALUATE	1
1.00	–1.757
2.00	1.757

Unstandardized canonical discriminant functions evaluated at group means

Prior Probabilities for Groups

		Cases Used in Analysis	
EVALUATE	Prior	Unweighted	Weighted
1.00	.500	5	5.000
2.00	.500	5	5.000
Total	1.000	10	10.000

Classification Function Coefficients

	EVALUATE	
	1.00	2.00
PROTEIN	1.035	3.976
VITAMIN	.706	–.471
(Constant)	–4.317	–17.081

Fisher's linear discriminant functions

Classification Tables[b,c]

EVALUATE			Predicted Group Membership		**Total**
			1.00	2.00	
Original	Count	1.00	5	0	5
		2.00	0	5	5
	%	1.00	100.0	.0	100.0
		2.00	.0	100.0	100.0
Cross-validated[a]	Count	1.00	5	0	5
		2.00	0	5	5
	%	1.00	100.0	.0	100.0
		2.00	.0	100.0	100.0

Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all case other than that case.
100.0% of original grouped cases correctly classified.
100.0% of cross-validated grouped cases correctly classified.

Classifying the Persons

It is well and good to find the discriminant function, but we are still interested in how well the function classifies the ten cases. The classification problem, in turn, involves two additional questions: (a) how well does the function assign the known cases in the sample; and (b) how well does it assign new cases *not* used in computing the function in the first place?

To answer these questions we need an *assignment rule*. One rule that seems intuitively plausible is based on Figure 13.6. A *classification boundary* between the two groups, Z_{crit}, can be identified as being midway between the means of the function for each of the two groups. To classify an individual, if $Z_i > Z_{\text{crit}}$ then the individual belongs in one group, while if $Z_i < Z_{\text{crit}}$ then the individual goes into the other group. As can be seen from Figure 13.6, no misassignments will be made if we adopt the rule:

- Assign all cases with discriminant scores that are on the left of the midpoint (1.596) to the *disliker* group.
- Assign all cases with discriminant scores that are on the right of the midpoint (1.596) to the *liker* group.

That is, all true dislikers will be correctly classified as such, as will all true likers. This can be shown by a 2 · 2 table, known as a classification or *confusion matrix*, as shown later in Table 13.6. We see that all entries fall along the main diagonal. For example, had any of the five true dislikers been called likers, the first row and second column would contain not a zero but, rather, the number of such misassignments.

The application of this rule can be stated in equivalent terms:

- Substitute the centroid of each group in the discriminant function and find the respective group scores (in our case, 0.824 for dislikers and 2.368 for likers).
- For any new case, compute the discriminant score and assign the case to that group whose group score is closer.

This rule makes two specific assumptions: (a) the prior probability of a new case falling into each of the groups is equal across groups; and (b) the cost of misclassification is equal across groups.

If a higher probability existed for likers, we could reduce the expected probability of misclassification by moving the cutting point of 1.596 to the left (closer to 0.824, the mean score for dislikers) so as to give a wider interval for the larger (likers) group. Similarly, if the cost of misclassifying a liker is higher than that for misclassifying a disliker, the cutting point would also be moved closer to 0.824.

A second point of interest concerns the tendency for classification tables (i.e., the confusion matrix) based on the sample used to develop the classification function to show better results than would be found upon cross-validation with new cases. That is, some capitalization on chance takes place in discriminant analysis, and one needs a way to measure this bias of an inflated percentage of correctly classified observations, just as cross-validation should be employed in multiple regression.

Procedures are available to develop a truer summary of the degree of correct assignments than would be obtained from fresh data (Lachenbruch, 1975; Klecka, 1980; Frank, Massy, & Morrison, 1965).

The most frequently suggested validation approach is the *holdout* method, in which the data set is randomly split into two subsamples. One subsample is used to develop the discriminant function and the other subsample is used as “fresh data” to test the function. It is suggested that this split-sample validation be replicated by using a different convention for the random assignment of the observations. Whether a researcher uses only one validation sample or replications, there may be problems associated with the sampling procedures used to arrive at the analysis and validation subsamples. These problems relate to the number of variables included in the discriminant function (when a stepwise procedure is used) and to inferences that the researcher might make concerning the classificatory power of resulting discriminant functions.

There may be differing numbers of variables included on any given replication. Also, there can be a wide range in the correct classification performance of the validation replications, although the analysis functions are quite stable.

There is another dimension to the validation process discussed above. For the analysis and validation samples there is a measure of the percentage of cases that would have been correctly classified based on *chance* alone. Obviously, the researcher is interested in whether the discriminant function is a significant improvement over chance classification.

Because our “Cereal Data” achieved 100 percent correct classification, we will now switch to another example to demonstrate these measures of chance classification. The classification or “confusion” matrix shown at the end of Table 13.6 was derived from a discriminant analysis where accident insurance purchasers and non-purchasers were predicted using demographic variables.

Overall, 74.16 percent of all subjects surveyed were correctly classified as purchasers and non-purchasers of accident insurance. This observed classification was found to be significant at the .001 level ($\chi 2 = 24.59$, df = 1) and ($Q = 69.58$, df = 1). Thus, observed classification is significantly different from expected chance classification.

Tests of Group Differences

Morrison (1969) considered the question of how well variables discriminate by formulating a likelihood ratio to estimate chance classification. This likelihood analysis provides a criterion that may be used to compare the proportion of correctly classified observations with the proportion expected by chance. This proportion, designated the proportional chance criteria, or Cpro (Morrison, 1969), is expressed as where

α = the proportion of customers in the sample categorized as purchasers

p = the true proportion of purchasers in the sample

$(1 - \alpha)$ = the proportion of the sample classified as non-purchasers

$(1 - p)$ = the true proportion of non-purchasers in the sample

This likelihood analysis states that 65.94 percent of the overall sample is expected to receive correct classification by chance alone. The proportional chance criterion, C_{pro}, has been used mainly as a point of reference for subjective evaluation (Morrison, 1969), rather than the basis of a statistical test to determine if the expected proportion differs from the observed proportion that is correctly classified.

This relationship between chance and observed proportions can be tested using a Z statistic of the form

$$\frac{P_{cc} - C_{pro}}{\sqrt{\frac{(C_{pro}) - (1 - C_{pro})}{n}}} = \frac{.7416 - .6594}{\sqrt{\frac{(.6594)(.3406)}{298}}} = 2.99$$

where P_{cc} is the percent of observations correctly classified, $C_{pro} = p\alpha + (1 - p)(1 - \alpha)$.

Thus for the example problem, the difference between expected and actual overall correct classification is significantly different at the .01 level. This overall test of significance suggests that further analysis should be conducted to determine the source of the deviation from chance expectations.

Table 13.6 Classification Table for Accident Insurance Purchasers

Fequency, Row % Chi-Square Contribution	*Predicted Purchase*	*Predicted Non-Purchase*	*Row Total Row Percentage*
Actual Purchase	n=22 66.7% 15.41	n=11 33.3% 6.46	n=33 11.1%
Actual Non-Purchase	N=66 24.9% 1.92	N=199 75.1% .80	n=265 88.9%
Column Totals Column Percentage	n=88 29.5%	n=21 70.5%	n=298 100%

Percent of Cases Correctly Classified = 221/298 = 74.16%
Chi-Square = 24.599 df=1, Significance < .001

Divergence may be present in any of the classification matrix cells (i.e., purchasers or non-purchasers, that are either correctly or incorrectly categorized), and thus each may be tested to determine whether its proportion differs from chance.

Classification and Misclassification Within Groups

The analysis to determine the source of deviation is conducted using the maximum chance criterion, designated C_{max} (Morrison, 1969). *C*max is the maximum expected correct classification for a selected group of interest. The computation of C_{max} is based on the assumption that all observations are categorized as coming from that group: for example, given that all 298 purchasers and non-purchasers were classified as purchasers, then the maximum correct classification, C_{max}, would be expressed

$$C_{max} = \frac{\text{Total purchasers}}{\text{Total customers}} = \frac{33}{298}$$

Because we are interested in the correct classification of insurance purchasers, the test of classification involves asking if the 66.67-correct insurance purchaser classification differs significantly from the 11.1-percent maximum expected chance classification. The Z statistic is used to test this relationship as shown for the example analysis.

$$Z_{11} = \frac{\text{Observed Correct Classification} - C_{max}}{\sqrt{\frac{(C_{max}) - (1 - C_{max})}{n}}} = \frac{.667 - .111}{\sqrt{\frac{(.111)(.889)}{33}}} = 10.17^*$$

* Significant at the .001 level.

This test may also be conducted for the other cells in the classification matrix:

$$Z_{12} = \frac{.333 - .111}{\sqrt{\frac{(.111)(.889)}{33}}} = 4.06$$

$$Z_{21} = \frac{.249 - .889}{\sqrt{\frac{(.889)(.111)}{265}}} = -33.16$$

$$Z_{22} = \frac{.751 - .889}{\sqrt{\frac{(.889)(.111)}{265}}} = -7.15$$

Thus cell Z_{11} shows that observed classification is significantly greater than is expected to occur by chance classification alone. The analysis of cells (1,2) and (2,1) shows that observed and expected misclassification results differ in that purchasers are misclassified into cell (1,2) less often than expected by chance, and non-purchasers are misclassified into cell (2,2) more often than expected by chance. Thus the discriminant functions appear to shift the classification of subjects toward the purchaser categories, as demonstrated by significantly greater than expected classification in the upper and left portions of the classification matrix.

At the most basic level, the validity of discriminant function analysis lies in the stability of the coefficients derived. These coefficients are the basis of classifying, profiling, and evaluating the underlying discriminant dimensions. It would be valuable to have a validation procedure that uses all the sample data for evaluating the stability of parameter estimates while allowing unbiased estimation of error rates. The *jackknife statistic* and the *U-method* have been proposed as such procedures. It is beyond the scope of this book to discuss these specialized techniques, and the reader is referred to the excellent references already cited.

When a discriminant analysis is run using a computer program, there is another option for validation, especially when a small sample is involved. The *leave-one-out* procedure operates by having each case (or observation) classified into a group according to classification functions computed from all the data except the case being classified. While we do not deal specifically with the classification of cases other than those from which the discriminant function is developed, it should be recognized that once the discriminant function is developed and validated for the population it may be used to classify other groups within the population.

Testing Statistical Significance

While the discriminant function perfectly classifies the ten cases of the calibration sample in our cereal likes and dislikes illustration, we still have not tested whether the group centroids differ significantly. This is analogous to testing for the significance of R^2 in multiple regression. Tests of the equality of group centroids can also proceed on the basis of an F-ratio that, in turn, is calculated from a variability measure known as *Mahalanobis squared distance*. We do not delve into the technical details of Mahalanobis squared distance, other than to say that it is like ordinary (Euclidian) squared distance that is computed between two centroids in a space with correlated axes and different measurement units.

Two other measures are widely used for testing overall statistical significance related to a discriminant function. The *canonical correlation coefficient* is a measure of the association that summarizes how related the discriminant function is to the groups. We discuss canonical correlation analysis later in this chapter. An indirect, and most widely used, approach to test for the statistical significance of the discriminant function examines the ability of the variables to discriminate among the group beyond the information that has been extracted by the previously computed functions. This is known as *residual discrimination,* and is measured by the statistic *Wilks' lambda* (also called the *U*-statistic). Wilks' lambda is a multivariate measure of group differences over discriminating variables and can be calculated in several ways. In general, it is calculated such that values of lambda near zero indicate high discrimination, and when it equals its maximum value of 1.0 the group centroids are equal and there is no discrimination (Klecka, 1980).

The statistical significance of a discriminant analysis merely provides answers to the questions:

- Is there a relationship?
- Do the predictor variables discriminate among the groups?
- Does this particular discriminant function contribute to the relationship?

Statistical significance says nothing about *how strong* a relationship is, *how much* difference exists between the groups, or to *what extent* a function contributes to overall discrimination. Moreover, tests of statistical significance are sensitive to sample size. For example, with a large sample size it is not difficult to get a significant *F*-ratio, even though classification accuracy is poor. Finally, there is a need to go beyond statistical significance and test for the *practical* (or substantive) significance. There is no single index of practical significance that is widely accepted.

Relative Importance of Predictor Variables

Because the original variables X_1 and X_2 were expressed in different units and display different standard deviations as well, the analyst generally *standardizes* the discriminant weights before examining their relative importance. A simple standardization procedure multiplies each discriminant weight (unstandardized) by the *total sample standard deviation* of that variable:

$$dj^{s(\omega)} = d_j \sigma_j$$

Standardized coefficients allow only an ordinal interpretation of variable importance. These coefficients are not appropriate to assess the relative discriminatory power of the variables included in the analysis. Mosteller and Wallace (1963) offer an appropriate measure of relative discriminating power:

$$I_j = \left| d_j(\bar{X}_{j1} - \bar{X}_{j2}) \right|$$

where

I_j = the importance value of the *j*th variable
d_j = unstandardized discriminant coefficient for the *j*th variable
X_{jk} = mean of the *j*th variable for the *k*th group

The relative importance weights may be interpreted as the portion of the discriminant score separation between the groups that is attributable to the *j*th variable. Since a relative importance value shows the value of a particular variable relative to the sum of the importance values of all variables, the relative importance of a variable (R_j) is given by Aw and Waters (1974):

$$R_j = \frac{I_j}{\sum_{j=1}^{n} I_j}$$

The end result of using this procedure is shown in Table 13.7, which is taken from a study of teenage smoking designed to see if consumption values (as represented by underlying had smoked and those who had never smoked (Albaum, Baker, Hozier, & Rogers, 2002). As these data show, although nine variables were included in the discriminant function, slightly more than 88 percent of the total discrimination was accounted for by only three variables.

Table 13.7 Relative Importance of Consumption Values for Teenage Smokers and Nonsmokers

Variables (Factors)	Standardized Coefficients	Unstandardized Coefficients (k_j)	Have Smoked Mean (X_1)	Never Smoked Mean (X_2)	$[k_j(X_{j1}-X_{j2})]$ (I_j)	Relative Importance Weight (R_j)
1	-0.298	-0.299	-0.0989	0.0908	0.0567	7.9%
2	0.091	0.091	0.0301	-0.0277	0.0053	0.7%
3	0.832	0.864	0.2861	-0.2627	0.4741	66.4%
4	0.344	0.345	0.1144	-0.1050	0.0757	10.6%
5	0.364	0.366	0.1211	-0.1112	0.0850	11.9%
6	-0.018	-0.018	-0.0061	0.0056	0.0002	0.0%
7	-0.004	-0,004	-0.0014	0.0013	0.0000	0.0%
8	-0.007	-0.007	-0.0023	0.0021	0.0000	0.0%
9	0.161	0.161	0.0533	-0.0490	0.0165	2.3%
					0.7135	100.0%

Correctly Classified 60.2% (cross-validated grouped cases); Cpro – 51.8%
Wilk's Lambda - .886 (prob. < .01)
Canonical Correlation - .337
Source: Albaum, Baker, Hozier and Rogers, 2002, p.68.

Determining relative importance of the predictor variables in discriminant analysis becomes increasingly difficult when more than two groups are involved. Although various coefficients and indices can be determined, interpretation becomes critical since more than one discriminant function may be involved.

Multiple Discriminant Analysis

All of the preceding discussion regarding objectives and assumption structure applies to multiple discriminant analysis as well. Accordingly, discussion of this section will be comparatively brief. What primarily distinguishes *multiple discriminant analysis* from the two-group case is that *more than one* discriminant function may be computed. For example, if we have three groups we can compute, in general, two nonredundant discriminant functions (as long as we also have at least two predictor variables). In general, with *G* groups and *m* predictors we can find up to the lesser of $G - 1$, or *m*, discriminant functions.

Not all the discriminant functions may be statistically significant, however. Moreover, it turns out to be a characteristic of multiple discriminant analysis that the first function accounts for the highest proportion of the among-to within-groups variability; the second function, the next highest; and so on. Accordingly, we may want to consider only the first few functions, particularly when the input data are rather noisy or unreliable to begin with. There remains the problem of interpretation of the functions.

Multiple discriminant analysis is considerably more detailed than might be surmised by this brief overview. Interested readers should consult more extensive and advanced discussions of the topic such as Johnson and Wichen (2002), Hair, Tatham, Anderson, and Black (1998), and Stevens (1996).

OTHER CRITERION-PREDICTOR ASSOCIATION MULTIVARIATE TECHNIQUES

Chi-Square Automatic Interaction Detection (CHAID) Analysis

One of today's key tools in database management and data mining is CHAID. This tool allows the researcher to explore relationships where both the number of respondents and the number of variables is very large. Inherent in these relationships, and in the resulting predictive models, is the phenomenon of interaction in which the response to changes in the level of one predictor variable depends on the level of some other predictor (or predictors). When interaction effects exist, the simple additive property of individual predictor-variable contributions to changes in the criterion no longer holds.

The major problem is that when mining data using exploratory analyses of observational and survey data, one does not ordinarily know *which* predictors are interactive (and how they interact). CHAID is insensitive to various forms of interaction.

CHAID is a sequential merging routine that merges nonsignificant predictor variable categories by analyzing a series of two-way cross-tabulations of the predictor and dependent variable.

While the predictor variables might originally be (a) nominal-, (b) ordinal-, or (c) interval scaled, these predictors are *all* recorded into categorical (nominal-scaled) variables with one of two characteristics:

1. Order across classes can be disregarded
2. Order across classes can be retained

For example, if one of the original predictors is employment status (white-collar, blue-collar, unemployed) one may wish to treat this variable as an unordered polytomy. Some other variable such as age may be recorded into 18–20; 21–23; 24–26 years, and so forth. In this case one would probably wish to maintain order across classes. Each predictor variable can be designated by the researcher as unordered or ordered, independently of the rest. It has been suggested that the primary goals of CHAID are segmentation analysis and an exploration of the relationships between variables. CHAID is often an appropriate technique in the following circumstances:

- We do not want to make strong assumptions about relationships.
- There are more than just a few potential explanatory variables.
- There are a large number of observations.
- AID also offers a variety of uses:
- Identification of demographic variables that discriminate and predict the dependent variable category, such as heavy users
- Identification of interaction effects for further analysis, such as logistic regression
- Data mining to find significance among the hundreds of possible crosstabulation combinations
- Identification of predictor variables that contribute little to explaining the dependent variable

Basically, CHAID performs a series of contingency table analyses of the data (similar in spirit to the bivariate cross-tabulation analyses discussed in Chapter 12). After splitting the initial sample on the basis of the "best" predictor, the process is repeated on each of the two subsamples, and so on. At each step of the sequential splitting and merging process, CHAID looks for the best *available* split, not the best set of some number of final groups. The main result of this is a tree structure that shows three statistics at each stage:

1. The predictor variable leading to the best binary split and how that predictor *is* split
2. The number of persons assigned to each of the two subgroups
3. The criterion-variable splits on each of the two subgroups

An Example of CHAID

Advertising executives have conducted research to identify profitable segments for a sweepstakes promotion.

Profit levels associated with responses are found to be $35 when the respondent pays for a subscription; –$7 when a respondent does not subscribe, but receives an introductory issue and follow-up sales materials; –$.15 for nonrespondents for mailing costs.

Of particular interest is the extent to which a respondent is profitable. That is, when asked to consider the subscription to magazines through the associated sweepstakes plan, which market segments are most likely to subscribe?

A large database (n = 81,040) was available that contained information about their response to the sweepstakes offer and known information about eight different demographic and credit card ownership predictors. All respondents had recently received a sweepstakes offer and had a known response that was recorded in the criterion variable to indicate that he or she was a subscriber (coded "Paid Respondent"), respondent-nonsubscriber (coded "Unpaid Respondent"), or a "nonrespondent."

CHAID was applied to this large data bank, and the results are shown in the tree diagram of Figure 13.7. At the top of the diagram we note that the total-sample probability of being a subscriber is 0.0059. That is, 478 out of the total of 81,040 recipients of the sweepstakes offer paid for a subscription. The first variable on which the total sample is split involves the size of the respondent's household, followed by age, income, and presence of a bankcard in the household.

Figure 13.8 shows a summary map of the CHAID tree that is numbered by the split order. These splits are detailed in Table 13.8, the gain summary, that shows the key segments for sweepstakes profitability based on the measure of profitability gain.

We observe that Node 8 provides the most gain (profitability) and is defined by the predictor variables as a two-person household, headed by an individual aged 55–64 with income of $15,000 or more. If we were to label this profile, it is the classic "empty-nester" whose children are no longer at home and now has a little extra money and time for their own pursuits, one of which is reading.

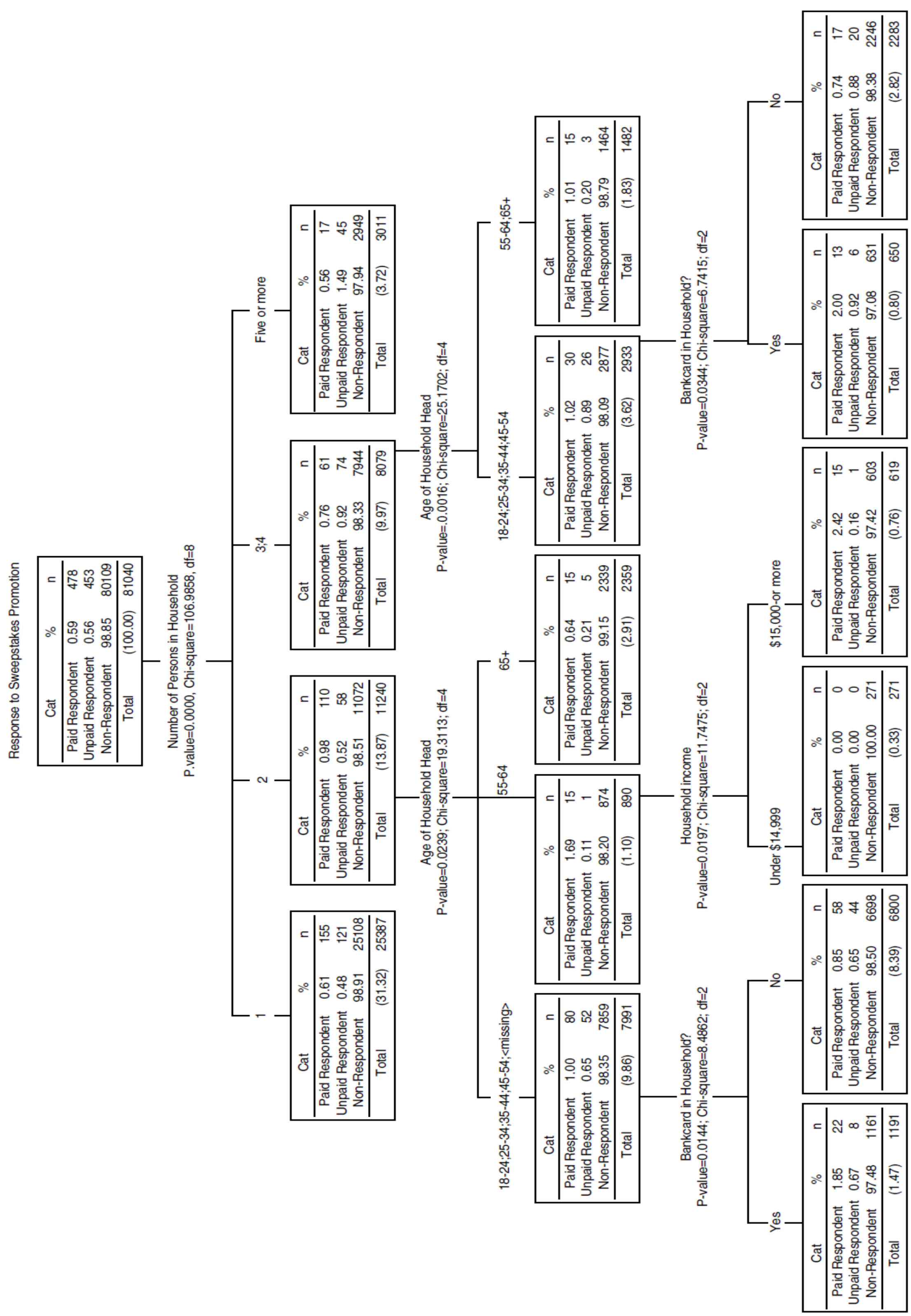


Figure 13.7 CHAID Sweepstakes Tree

Figure 13.8 CHAID Tree Summary Map

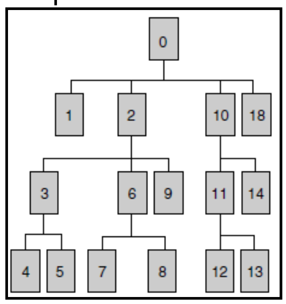


Table 13.8 CHAID Gain

Gain Summary *Target variable: Response to Sweepstakes Promotion*								
	Node-by-Node				*Cumulative*			
Node	*Node: n*	*Node:%*	*Gain*	*Index (%)*	*Node: n*	*Node:%*	*Gain*	*Index (%)*
8	619	0.76	0.69	3628.50972	619	0.76	0.69	3628.50972
12	650	0.80	0.49	2572.90367	1269	1.57	0.59	3087.81316
4	1191	1.47	0/45	2381.18629	2460	3.04	0.52	2745.70235
14	1482	1.83	0.19	1008.12275	3942	4.86	0.40	2092.45705
5	6800	8.39	0.11	554.14567	10742	13.26	0.21	1118.66098
9	2359	2.91	0.06	309.87554	13101	16.17	0.19	973.02898
13	2283	2.82	0.05	271.75404	15384	18.98	0.17	868.95912
1	25384	31.32	0.03	168.00454	40768	50.31	0.08	432.51311
18	3011	3.72	-0.05	-283.25238	79938	98.64	0.02	115.27446
7	271	0.33	-0.15	-787.99468	80209	98.97	0.02	122.22261

Node 12 is a 3–4 person household with a young to middle-age head of household who holds a bankcard.

The process of identifying profitable segments continues until we observe that Nodes 18 and 7 offer a negative gain (meaning they are unprofitable). We conclude that we should not target the large family segment having five or more family members (Node 18), or the poor empty-nester segment that is identical to our most profitable segment (Node 8), except without the income base.

The upshot of all of this is that by data mining with CHAID, we can identify 50.31 percent of our database that promises to be profitable and another 49.69 percent (39,441 individuals) that may be unprofitable. We further observe that the average gain (profit) drops from 0.17 to 0.08 with the inclusion of Node 1, which has an average gain of 0.03 per person for that segment. The successive targeting of the next most profitable segment can do much to maximize profits, as can eliminating the negative gain segments. Simply refining the database by deleting nonconforming profiles often does much to improve profitability of the marketing effort.

In short, this kind of information underscores the importance of certain types of demographic variables in this case, and leads to a more detailed understanding of the kinds of segments that responded by subscribing to this sweepstakes magazine offer.

While other summary statistics (e.g., probabilities, chi-square) can also be shown, the above outputs represent the principal ones. It should be emphasized that a basic assumption underlying the application of CHAID is that variables useful in explaining one part of the database are not necessarily those most effective for explaining another part.

A few other key considerations, in the nature of restrictions or criteria for the stopping of the splitting sequence, enter into the actual application of CHAID:

1. All partitionings of the sample may be subject to a requirement that the proportionate reduction in the criterion-variable probability exceed some level (specified by the researcher). This is to guard against partitionings that do not appreciably reduce variation in the criterion variable.
2. To qualify for further splitting, a group must have a probability greater than some level (again specified by the researcher). This is to guard against splits that, pragmatically speaking, are not worth the effort (for example, where the group is already quite homogeneous).
3. In addition to the control parameters above, the researcher may place an upper limit on the total number of groups formed and/or the minimum number of persons (or objects) in each group.

A few caveats regarding the application of CHAID should be mentioned at this point. First, CHAID is generally designed for really large samples, on the order of 1,000 or more. Since many versions of CHAID will take many predictors, the program has ample opportunity to capitalize on chance variation in the data. Moreover, no statistical inferential apparatus is associated with the approach. This suggests the value of cross-validating results on new data or, possibly, double cross-validating by applying CHAID to separate halves of the sample.

Second, CHAID, being a sequential search procedure, does not specify an explicit model in the way, for example, that Discriminant Analysis does. In this regard it is often useful to use CHAID as an initial screening device to find those predictors that appear to be most prominent in accounting for criterion dependence. This can then be followed by the formulation of an explicit logit regression model that includes main effect and interaction terms of specific interest to the research. The joint use of CHAID with other statistical techniques, such as CART models (classification and regression trees), provides additional useful features for exploratory data analysis.

Third, despite the explicit concern for interaction, CHAID is found to be insensitive to various forms of interaction. Since CHAID only examines the immediate effect of a predictor as a two-way table and not future possible splits, any interactions that are not one-stage will not be identified.

One of the nice characteristics of CHAID is its simplicity of output, which facilitates an ease of understanding by researchers and managers alike. The output takes the form of a tree diagram, in which one can follow the result of each split as it takes place. This is illustrated in the brief case study analyzed using the SPSS program *AnswerTree,* a suite of programs that includes CHAID and CART models. The data is that supplied by Magidson (1993) in the original SPSS CHAID 6.0 program. *AnswerTree* is compatible with SPSS and all current Windows operating systems.

Canonical Correlation Analysis

Canonical correlation is a generalization of multiple correlation analysis to two or more criterion (dependent) variables. To illustrate the technique of canonical correlation, let us consider a radial tire study which has three main predictor variables:

*X*1: General interest in the product class of steel-belted radials
*X*2: Whether the firm's old brand was chosen as the respondent's last purchase of replacement tires
*X*3: Pre-exposure (before seeing the commercials) interest in the firms' new brand of steel-belted radials

Now let us assume that *two* criterion variables are involved:

*Y*1: Believability of the claims made in the firm's new TV commercial;
*Y*2: Post-exposure interest in the claims made in the firm's new brand (as before).

What we would like to find out is how highly correlated the *group* of two criterion variables is with the *group* of three predictors. Moreover, we would like to find a linear composite of the *Y*-variable set and a (different) linear composite of the *X*-variable set that will produce a maximal correlation.

This is what canonical correlation is all about. Canonical correlation deals with (a) both description and statistical inference of (b) a data matrix partitioned into at least two predictors where (c) all variables are interval-scaled and (d) the relationships are assumed to be linear. Thompson (1984, p. 10) notes that, in more general terms, canonical correlation analysis can be used to investigate the following research questions:

- To what extent can one set of two or more variables be predicted or explained by another set of two or more variables?
- What contribution does a single variable make to the explanatory power of the set of variables to which the variable belongs?
- To what extent does a single variable contribute to predicting or explaining the composite of the variables in the variable set to which the variable does not belong?
- What different dynamics are involved in the ability of one variable set to explain, in different ways, different portions of the other variable set?
- What relative power do different canonical functions have to predict or explain relationships?
- How stable are canonical results across samples or sample subgroups?
- How closely do obtained canonical result conform to expected canonical results?

Table 13.9 shows the results of the radial tire problem. In general, with *p* criteria and *q* predictors, one can obtain more than a single pair of linear composites—up to a maximum of the smaller of *p* and *q*. Thus, in our case we would obtain two pairs of linear composites, uncorrelated across pairs, with the first pair exhibiting maximum correlation.

In general, the canonical correlation of successive pairs decreases; that is, the first pair displays the highest correlation, the second pair the next highest, and so on. All composites are mutually uncorrelated *across* pairs. However, as it turned out, only the first pair of linear composites was statistically significant at the 0.05 alpha level; hence, only this pair of weights is shown.

Table 13.9 Result of Canonical Correlation

	Canonical Weights	*Structure Correlations*
Criterion set		
Y_1	-0.110	0.594
Y_2	1.069	0.997
Predictor set		
X_1	0.346	0.539
X_2	0.141	0.382
X_3	0.817	0.930
Canonical correlation	0.582	

Input Correlations For Canonical Analysis

	Y_1	Y_2	X_1	X_2	X_3
Y_1	1.000				
Y_2	0.659	1.000			
X_1	0.202	0.315	1.000		
X_2	0.097	0.218	0.086	1.000	
X_3	0.321	0.539	0.226	0.258	1.000

Table 13.9 shows that the canonical correlation between the two batteries is 0.582. As is the case with multiple correlations, this measure varies between zero (no correlation) and one (perfect correlation). Since *mutual* association between the pair of batteries is involved, we can say that the pair of linear composites account for $(0.582)^2$, or 34 percent of the shared variation between the two batteries.

The canonical weights for the criterion set show that Y_2 (post-exposure interest) is the dominant variable in the criterion set; its canonical weight is 1.069. The dominant variable in the predictor set is *X*3; its weight is 0.817. Since all variables are standardized to zero mean and unit standard deviation *before* the analysis, the weights are already in standardized form.

Table 13.9 really says that if we formed a linear composite of the criterion variables using the canonical weights:

$$T_c = -0.110Y_{s1} + 1.069Y_{s2}$$

and another linear composite of the predictors, using the canonical weights:

$$T_p = 0.346X_{s1} + 0.141X_{s2} + 0.817X_{s3}$$

and took these two columns of numbers (the canonical weight scores) and correlated them, the result would be a simple correlation of 0.582 (the canonical correlation).

The structure correlations are also shown in Table 13.9. These are the simple correlations of each original variable with the canonical scores of its own battery's linear composite. Again we note that Y_2 is the most important variable (structure correlation of 0.997) in the criterion set and X_3 the most important variable (structure correlation of 0.930) in the predictor set. Indeed, as noted in the input correlation matrix, the *simple* correlation between Y_2 and X_3 is 0.539, almost as high as the correlation between the full batteries.

Since this is a rather sophisticated technique, we refer the reader to other sources for further discussion of the procedure, such as Thompson (1984) and Levine (1977).

Probit and Logit Analysis

At the beginning of this chapter we emphasized that multiple regression is a useful technique for estimating a dependent variable from a set of independent variables. A major assumption of standard regression analysis is that the dependent variable is continuous and at least interval-scaled. When the dependent variable is dichotomous, binary regression can be used.

What can the researcher do when the problem requires estimating relationships where the dependent variables that are nonmetric (i.e., nominal- or ordinal-scaled)? When regression analysis is misapplied, we often see results with an unnecessarily high proportion of unexplained variance (i.e., a lower R^2), misleading estimates of the effects of the predictor variables, and an inability to make statements about the probability of given responses. To overcome these problems, *probit* and *logit* have been developed.

Probit and logit deal with the same type of problem as regression, the problem of predicting a dependent variable that is nominal or ordinal scaled. They differ solely in the assumption made about the frequency distribution of the response:

- In probit, the response is assumed to be normally distributed.
- In logit, a logistic distribution is assumed.

Reported research about marketing applications seems to indicate that logit is the favored technique of the two (Malhotra, 1984). Logit analysis is often used in choice based conjoint analysis where the dependent variable is the brand chosen (discussed in Chapter 14).

Explanations of how these techniques work and the underlying theories behind them tend to be technical. For a more detailed explanation we refer the reader to the references already cited and to the excellent works of Aldrich and Nelson (1984), Liao (1994), and Menard (2001).

Path Analysis/Causal Modeling

In recent years, marketing researchers have become increasingly interested in the useful application of path analysis and structural equation modeling as an approach to *causal modeling*. Causal modeling provides the researcher with a systematic methodology for developing and testing theories in marketing. From all this interest has emerged the development of *structural equation modeling procedures*. These procedures blend two basic techniques:

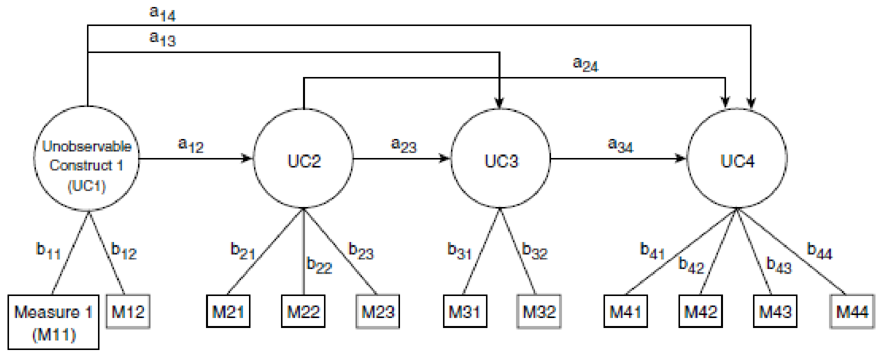
- Factor analysis
- Simultaneous equation regression

Both of these techniques use multiple measures and simultaneous equation models to estimate the path of causation.

The essence of what happens with this type of analysis can be seen from a simple illustration. In Figure 13.10 we show a system with four unobservable constructs, each of which has more than one observable measure. In some situations the measures themselves may be correlated, which precludes finding a strong reliability from either simple indices or regressions that relate the indices themselves as means of showing relations between the unobserved constructs.

To overcome this problem and better estimate the relevant parameters, some rather complicated procedures have been developed which estimate both between the constructs (*bij*) and their measures and the links between the constructs themselves (*aij*) simultaneously. One widely used technique to estimate these parameters is LISREL (structural equation software).

Figure 13.10 Hypothetical Causal Structure



SOURCE: From Albaum, G., Baker, K., Hozier, G., and Rogers, R., "Smoking Behavior, Information Sources, and Other Intervention Activities," in *Journal of Consumer Affairs, 36*/1:68. Copyright © 2002. Reprinted by permission of the University of Wisconsin Press.

Another technique is PLS (partial least squares). A third technique, available from SPSS, is AMOS (structural equation and path analysis software). These techniques, and structural equation modeling in general, are highly technical methods of analyses and are recommended only for experienced users.

SUMMARY

Chapter 13 has focused on a number of multivariate techniques. Our discussions of multiple regression and discriminant analysis have been more detailed than the descriptions of other techniques. We have deliberately focused on those techniques that are most often used in market research.

We started the chapter with a discussion of multiple regression and described such measures as regression coefficients, the coefficient of determination, and the product-moment coefficient.

Next we discussed discriminant analysis, one of the most widely used criterion-predictor methods. We described a sample problem and solved it numerically. Discriminant analysis was also discussed as a tool for prediction and classification.

This chapter concluded with a brief introduction to some less well-known techniques for analyzing between-set dependence: CHAID, canonical analysis, correspondence analysis, and probit and logit analysis. At this point, it might appear that an almost bewildering array of techniques has been paraded before the reader.

We have tried to discuss the principal assumption structure of each technique, appropriate problems for applying it, and sufficient numerical applications to give the reader a feel for the kind of output generated by each program. Further discussion of the use of multivariate analysis in marketing can be found in Hair, Anderson, Tatham.

Our coverage of so vast and complex a set of methods is limited in depth as well as breadth. The fact remains, however, that marketing researchers of the future will have to seek grounding in multivariate methodology, if current research trends are any indication.

This grounding will probably embrace three facets: (a) theoretical understanding of the techniques; (b) knowledge of the details of appropriate computer algorithms to implement the techniques; and (c) a grasp of the characteristics of substantive problems in marketing that are relevant for each of the methods. Fortunately, most statistical analysis packages have programs for using these techniques.

REFERENCES

Aaker, D. A. (1981). Multivariate analysis in marketing. In *Multivariate analysis in marketing* (2nd ed.). Palo Alto, CA: Scientific Press.

Albaum, G., Baker, K. G., Hozier, G. C., Jr., & Rogers, R. D. (2002). Smoking behavior, information sources, and consumption values of teenagers: Implications for public policy and other invention failures. *Journal of Consumer Affairs*, *36*(1), 50–76.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit, and probit models*. Beverly Hills, CA: Sage.

Aw, R. W., & Waters, D. (1974). A discriminant analysis of economic, demographic, and attitudinal characteristics of bank charge-card customers. *Journal of Finance*, *29*, 973–980.

Bagozzi, R. (1980). *Causal models in marketing*. New York: Wiley.

Berk, K. N., & Carey, P. (2000). *Data analysis with Microsoft Excel*. Pacific Grove, CA: Duxbury Press.

Blalock, H. M., Jr. (1962). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Carroll, J. D., Green, P. E., & Schaffer, C. M. (1987). Comparing interpoint distances in correspondence analysis. *Journal of Marketing Research*, *24*, 445–450.

Carroll, J. D., Green, P. E., & Schaffer, C. M. (1986). Interpoint distance comparisons in correspondence analysis. *Journal of Marketing Research*, *23*, 271–280.

Clausen, S. E. (1998). *Applied correspondence analysis: An introduction*. Thousand Oaks, CA: Sage.

Crask, M. R., & Perrault,W. D., Jr. (1979, February). Validation of discriminant analysis in marketing research. *Journal of Marketing Research*, *14*, 60–68.

Curtis, M. E. (1990, Sept. 3). Avoid hazards of stepwise regression. *Marketing News*, *24*, 26.

Frank, R. E., Massy,W. F., & Morrison, D. G. (1965, August). Bias in multiple discriminant analysis. *Journal of Marketing Research, 2*, 250–258.

Green, P. E., Tull, D. S., & Albaum, G. (1988). *Research for marketing decisions* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Greenacre, M. J. (1993). *Correspondence analysis in practice*. London: Academic Press.

Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Johnson, R. A., & Wichen, D. W. (2002). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Klecka,W. R. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage.

Lachenbruch, P. A. (1975). *Discriminant analysis*. New York: Hafner Press.

Levine, M. S. (1977). *Canonical analysis and factor comparison*. Beverly Hills, CA: Sage.

Liao, T. F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*. Thousand Oaks, CA: Sage.

Magidson, J., & SPSS, Inc. (1993). *SPSS for Windows CHAID Release 6.0.* Chicago: SPSS, Inc.
Malhotra, N. K. (1984, February). The use of linear logit models in marketing. *Journal of Marketing*

Research, 21, 20–31. Menard, S. (2001). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Monroe, K. B., & Petroshius, S. M. (n.d.). *Developing causal priorities in marketing.* Working Paper, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Morrison, D. G. (1969, May). On the interpretation of discriminant analysis. *Journal of Marketing Research, 6,* 156–163.

Mosteller, F., & Wallace, D. L. (1963, June). Influence in an authorship problem. *Journal of the American Statistical Association, 55,* 275–309.

Smith, S. M. (2004). *A note on the interpretation and analysis of the linear discriminant model for prediction and classification.* Retrieved May 18, 2004, from http://marketing.byu.edu/htmlpages/ tutorials/discriminant.htm

Steckel, J. H., & Vanhonacker, W. R. (1991). Crossvalidating regression models in marketing research. (Working Paper # MARK-91-10.) Stern School of Business, New York University.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretations*. Beverly Hills, CA: Sage.

Whitlark, D., & Smith, S. (2001, Summer). Using correspondence analysis to map relationships: It's time to think strategically about positioning and image data. *Marketing Research, 2,* 23–27.
Whitlark, D., & Smith, S. (2004). *Measuring brand performance using online questionnaires: Advantages and issues with "pick any" data.* Provo, UT: Marriott School, Brigham Young Univ.

Chapter 14
MULTIVARIATE ANALYSIS II
FACTOR ANALYSIS, CLUSTERING METHODS, MULTIDIMENSIONAL SCALING AND CONJOINT ANALYSIS

In this chapter we discuss two techniques that do not require data to be partitioned into criterion and predictor variables. Rather, it is the entire set of interdependent relationships that are of interest. We discuss factor analysis as a methodology that identifies the commonality existing in sets of variables. This methodology is useful to identify consumer lifestyle and personality types.

Continuing with analyses that do not partition the data, a second set of methods is effective in clustering respondents to identify market segments. The fundamentals of cluster analysis are described using examples of respondents and objects grouped because they have similar variable scores.

Third, we discuss two sets of multivariate techniques, multidimensional scaling and conjoint analysis, that are particularly well suited (and were originally developed) for measuring human perceptions and preferences. Multidimensional scaling methodology is closely related to factor analysis, while conjoint analysis uses a variety of techniques (including analysis of variance designs and regression analysis) to estimate parameters, and both techniques are related to psychological scaling (discussed in Chapter 9). The use of both multidimensional scaling and conjoint analysis in marketing is widespread.

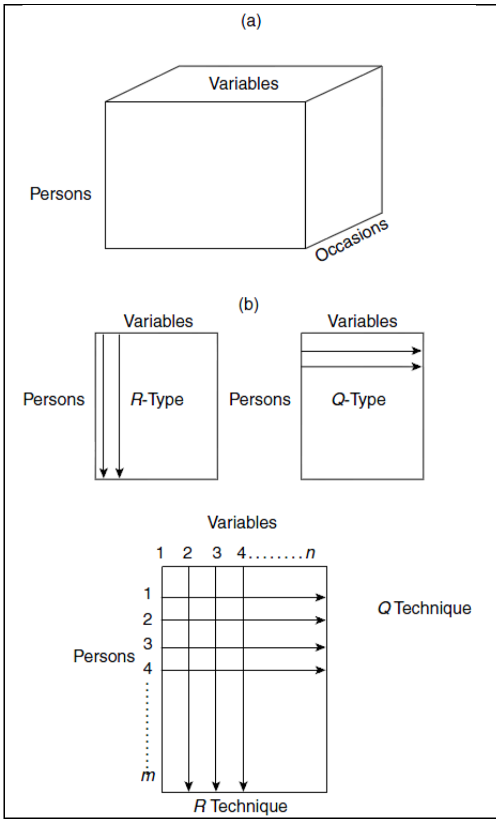
AN INTRODUCTION TO THE BASIC CONCEPTS OF FACTOR ANALYSIS

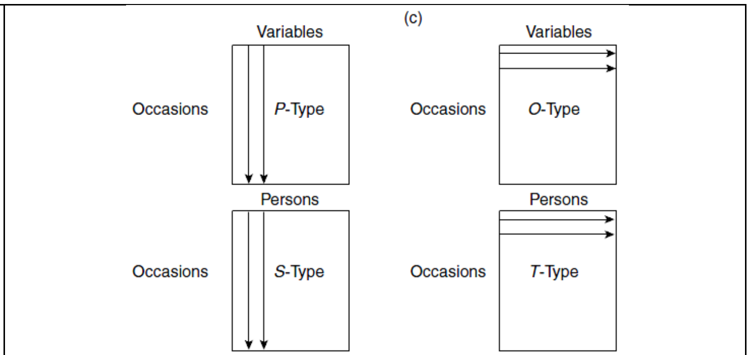
Factor analysis is a generic name given to a class of techniques whose purpose often consists of data reduction and summarization. Used in this way, the objective is to represent a set of observed variables, persons, or occasions in the form of a smaller number of hypothetical, underlying, and unknown dimensions called *factors*.

Factor analysis operates on the data matrix. The form of the data matrix can be flipped (transposed) or sliced to produce different types, or *modes*, of factor analysis. The most widely used mode of factor analysis is the *R*-technique (relationships among items or variables are examined), followed distantly by the *Q*-technique (persons or observations are examined). These, together with other modes, are identified in Exhibit 14.1. "Creative" marketing researchers may find *S*- and *T*-techniques helpful when analyzing purchasing behavior or advertising recall data. The *P*- and *O*-techniques might be appropriate for looking at the life cycle of a product class, or perhaps even changes in demographic characteristics of identified market segments.

EXHIBIT 14.1 Modes of Factor Analysis

Six distinct modes of factor analysis have been identified (Stewart, 1981, p. 53). The alternative modes of factor analysis can be portrayed graphically. The original data set is viewed as a *variables/persons/occasions* matrix (a). *R*-type and *Q*-type techniques deal with the variables/persons dichotomy (b). In contrast, *P*-type and *O*-type analyses are used for the occasions/ variables situation and *S*-type and *T*-type are used when the occasions/persons relationship is of interest (c).





Technique	*Factors Are Loaded By*	*Indices Of Association Are Computed Across*	*Data Are Collected On*
R	Variables	Persons	One occasion
Q	Persons	Variables	One occasion
S	Persons	Occasions	One variable
T	Occasions	Persons	One variable
P	Variables	Occasions	One person
O	Occasions	Variables	One person

Factor analysis does *not* entail making predictions using criterion and predictor variables; rather, interest is centered on summarizing the relationships involving a *whole* set of variables. Factor analysis has three main qualities:

1. The analyst is interested in examining the strength of the overall association among variables, in the sense that a smaller set of factors (linear composites of the original variable) may be able to preserve most of the information in the full data set. Often one's interest will stress description of the data rather than statistical inference.
2. No attempt is made to divide the variables into criterion versus predictor sets.
3. The models typically assume that the data are interval-scaled.

The major substantive purpose of factor analysis is to search for (and sometimes test) structure in the form of constructs, or *dimensions*, assumed to underlie the measured variables. This search for structure is accomplished by literally partitioning the total variance associated with each variable into two components: (a) common factors and (b) unique factors. *Common factors* are the underlying structure that contributes to explaining two or more variables. In addition, each variable is usually influenced by unique individual characteristics not shared with other variables, and by external forces that are systematic (non-random) and not measured (possibly business environment variables). This non-common factor variance is called a *unique factor* and is specific to an individual variable. Graphically, this may appear as diagrammed in Figure 14.1, in which four variables are reduced to two factors that summarize the majority of the underlying structure, and four unique factors containing information unique to each variable alone.

The structure of the factors identified by the analysis will, of course, differ for each data set analyzed. In some applications the researcher may find that the variables are so highly correlated that a single factor results. In other applications the variables may exhibit low correlations and result in weak or ill-defined factors. In response to these eventualities, the researcher may revise the variable list and add additional variables to the factor analysis. The process of adding and eliminating variables is common in factor analysis when the objective of the analysis is to identify those variables most central to the construct and to produce results that are both valid and reliable. Behavioral and consumer researchers have employed these methods to develop measurement instruments such as personality profiles, lifestyle indexes, or measures of consumer shopping involvement. Thus, in addition to serving as a data reduction tool, factor analysis may be used for to develop behavioral measurement scales.

We use a numerical example to illustrate the basic ideas of factor analysis. A grocery chain was interested in the attitudes (in the form of images) that customers and potential customers had of their stores. A survey of 169 customers was conducted to assess images. The information obtained included 14 items that were rated using a seven-category semantic differential scale. These items are shown in Table 14.1. The resulting data set would then be a matrix of 169 rows (respondents) by 14 columns (semantic differential scales). These data will be analyzed as *R*-type factor analysis.

Figure 14.1 The Concept of Factor Analysis

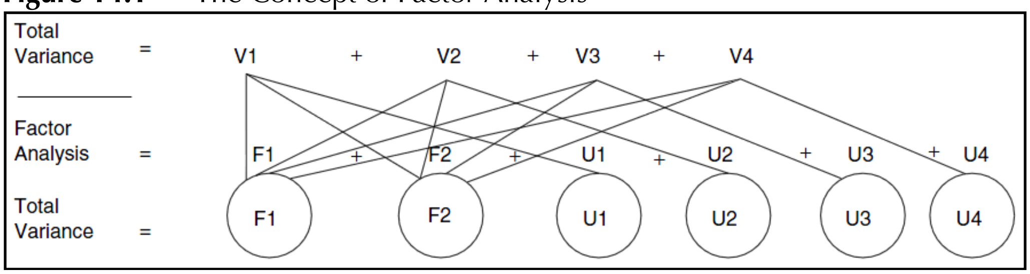


Table 14.1 Bipolar Dimensions Used in Semantic Differential Scales for Grocery Chain Study

Inconvenient location—Convenient location
Low-quality products—High-quality products
Modern—Old-fashioned
Unfriendly clerks—Friendly clerks
Sophisticated customers—Unsophisticated customers
Cluttered—Spacious
Fast check-out—Slow check-out
Unorganized layout—Organized layout
Enjoyable shopping experience—Unenjoyable shopping experience
Bad reputation—Good reputation
Good service—Bad service
Unhelpful clerks—Helpful clerks
Good selection of products—Bad selection of products
Dull—Exciting

IDENTIFYING THE FACTORS

If we now input the raw data into a factor analysis program, correlations between the variables are computed, as is the analysis. Some relevant concepts and definitions for this type of analysis are presented in Exhibit 14.2.

A factor analysis of the 14 grocery-chain observed variables produces a smaller number of underlying dimensions (factors) that account for most of the variance. It may be helpful to characterize each of the 14 original variables as having an equal single unit of variance that is redistributed to 14 underlying dimensions or factors. In every factor analysis solution, the number of input variables equals the number of common factors plus the number of unique factors to which the variance is redistributed. In factor analysis, the analysis first determines how many of the 14 underlying dimensions or factors are common, and then the common factors are interpreted.

EXHIBIT 14.2 Some Concepts and Definitions of *R*-Type Factor Analysis

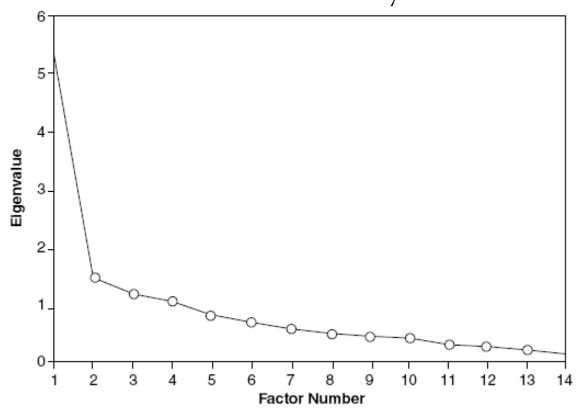
Factor Analysis:	A set of techniques for finding the underlying relationships between many variables and condensing the variables into a smaller number of dimensions called factors.
Factor:	A variable or construct that is not directly observable, but is developed as a linear combination of observed variables.
Factor Loading:	The correlation between a measured variable and a factor. It is computed by correlating factor scores with observed manifest variable scores.
Factor Score:	A value for each factor that is assigned to each individual person or object from which data was collected. It is derived from a summation of the derived weights applied to the original data variables.
Communality (h^2):	The common variance of each variable summarized by the factors, or the amount (percent) of each variable that is explained by the factors. The uniqueness component of a variable's variance is $1-h^2$.
Eigenvalue:	The sum of squares of variable loadings of each factor. It is a measure of the variance of each factor, and if divided by the number of variables (i.e., the total variance), it is the percent of variance summarized by the factor.

Table 14.2 identifies the proportion of variance associated with each of the 14 factors produced by the analysis where the factors were extracted by Principal Component analysis. *Principal Components*, one of the alternative methods of factor analysis, is a method of factoring which results in a linear combination of observed variables possessing such properties as being orthogonal to each other (i.e., independent of each other), and the first principal component represents the largest amount of variance in the data, the second representing the second largest, and so on. It is the most conservative method. For a more detailed discussion of the alternative methods, see Kim and Mueller (1978a, 1978b). In column two, the eigenvalues are reported.

Computed as the sum of the squared correlations between the variables and a factor, the eigenvalues are a measure of the variance associated with the factor. The eigenvalues reported in Table 14.2 are a measure of the redistribution of the 14 units of variance from the 14 original variables to the 14 factors. We observe that factors 1, 2, 3, and 4 account for the major portion

(66.5 percent) of the variance in the original variables. In Figure 14.2, a *scree plot* depicts the rapid decline in variance accounted for as the number of factors increase. This chart graphs the eigenvalues for each factor. It is a useful visual tool for determining the number of significant factors to retain. The shape of this curve suggests that little is added by recognizing more than four factors in the solution (the additional factors will be unique to a single variable).

Figure 14.2 Scree Plot for Grocery Chain Data Factors



An accepted rule-of-thumb states that if a factor has an associated eigenvalue greater than or equal to 1.0, then the factor is "common" and a part of the solution. This rule-of-thumb is closely aligned with the intuitive decision rules associated with the scree chart. When we observe an eigenvalue less than 1.0, the factor accounts for less variance than was input by a single input variable.

Table 14.2 Factor Eigenvalues and Variance Explained for Grocery Chain Study

Initial eigenvalues			
Factor	*Total*	*% Variance*	*Cumulative %*
1	5.448	38.917	38.917
2	1.523	11.882	49.799
3	1.245	8.89	58.689
4	1.096	7.827	66.516
5	0.875	6.247	72.763
6	0.717	5.12	77.883
7	0.62	4.429	82.312
8	0.525	3.747	86.059
9	0.478	3.413	89.472
10	0.438	3.131	92.603
11	0.326	2.329	94.932
12	0.278	1.985	96.917
13	0.242	1.729	98.646
14	0.19	1.354	100

Table 14.3 shows the matrix of factor loadings, or correlations of the variables with the factors. If each factor loading in each column were squared, the sum would equal the eigenvalue shown in Table 14.2. Squaring the loadings (h^2) and summing across the columns results in the amount of variance in the variables that is to be explained by the factors. These values are known as *communalities*.

Interpreting the Factors

The interpretation of the factors is subjectively based on the pattern of correlations between the variables and the factors. The factor loadings provide the basis for interpreting the factors; those variables having the highest loading contribute most to the factor and thereby should receive the most weight in interpreting of the factor.

In factor analysis two solutions typically are obtained. The *initial* solution is based on certain restrictions: (a) there are *k* common factors; (b) underlying factors are orthogonal (i.e., uncorrelated or independent) to each other; and (c) the first factor accounts for as much variance as possible, the second factor for as much of the residual variance as possible left unexplained by the first factor, and so on (Kim & Mueller, 1978a). The second solution is accomplished through rotation aimed at getting loadings for the variables that are either near one or near zero for each factor. The most widely used rotation is called *varimax,* a method of rotation which leaves the factors uncorrelated. This rotation maximizes the variance of a column of the factor loading matrix, thus simplifying the factor structure and making the factors more interpretable.

In Table 14.3, Factor 1 is identified by four variables. The major contributions are made by the variables "Quality of products," "Reputation," Selection of products," and "Modernism." We might interpret this factor as the construct *up-to-date quality products*.

Factor 2 is identified by three variables: "Sophistication of customers," "Speed of checkout," and "Dull/Exciting." This factor might be interpreted as the *fast and exciting for sophisticated customers*. Factor 3 is explained by the variables "Friendliness of clerks," Cluttered/ Spacious," and "Layout." One interpretation of this factor is that it represents the construct of *friendliness of store*. Finally, the last factor is defined by five variables. These all might be a reflection of *satisfaction with the shopping experience*.

Table 14.3 Varimax Rotated Factor Loading Matrix for Grocery Chain Data *Factor*

Factor Loadings (Sorted)					
Variable	*1*	*2*	*3*	*4*	*Communalities (h^2)*
Selection of products	0.799	-0.066	0.184	0.126	0.692
Quality of products	0.789	0.033	0.237	-0.081	0.687
Reputation	0.724	-0.283	0.096	0.296	0.702
Modern	-0.665	0.216	-0.071	-0.221	0.543
Customers	-0.235	0.781	-0.139	0.069	0.689
Check-out	0.170	0.720	-0.040	-0.326	0.655
Dull	-0.284	0.668	-0.227	0.046	0.581
Cluttered	0.070	-0.162	0.894	0.052	0.834
Layout	0.323	-0.058	0.742	0.150	0.681
Friendliness of clerks	0.199	-0.298	0.606	0.433	0.683
Location	0.013	0.218	0.011	0.735	0.587
Service	-0.257	0.339	-0.393	-0.588	0.680
Helpfulness of clerks	0.281	-0.338	0.290	0.597	0.634
Shopping experience	-0.353	0.448	-0.183	-0.552	0.664

The example of Table 14.3 depicts a set of factors with loadings that are generally high or low. However, the loadings are often in the .4 to .8 range, questioning at what level the variables make significant enough input to warrant interpretation in the factor solution. A definitive answer to this question cannot be given; it depends on sample size. If the sample size is small, correlations should be high (generally .6 and above) before the loadings are meaningful. But as the sample size increases, the meaning of correlations of lower value may be considered (generally .4 and above).

Overall, it should be obvious that *more than one interpretation may be possible for any given factor*. Moreover, it may be that a factor may not be interpretable in any substantive sense. This may or may not be a problem, depending upon the objective of the factor analysis. If done for data-reduction purposes, and the results will be used in a further analysis (such as multiple regression or discriminant analysis), being unable to interpret substantively may not be critical. One use of factor analysis is to identify those variables that reflect underlying dimensions or constructs. Once identified, the researcher can select one or more original variables for each underlying dimension to include in a subsequent multivariate analysis. This ensures that all underlying or latent dimensions are included in the analysis.

Factor Scores

Once the underlying factors are identified, the resulting factors or constructs are often interpreted with respect to the individual respondents. Simply stated, we would like to know how each respondent scores on each factor. Does the respondent have high scores on the *up-to-date-quality products* and *friendliness of store* constructs? In general, since a factor is a linear combination (or linear composite) of the original scores (variable values), it can be shown as

$$F_i = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

where F_i is the factor score for the ith factor, the a_n are weights (factor loadings for the n variables) and the X_n are respondent i's standardized variable scores.

Most factor analysis computer programs produce these factor score summates and merge them with the original data file. Augmenting the data set with factor scores enables the analyst to easily prepare descriptive or predictive analyses that segment respondents scoring high on a given factor. In short, factor scores (rather than original data values) can be used in subsequent analysis.

Correspondence Analysis

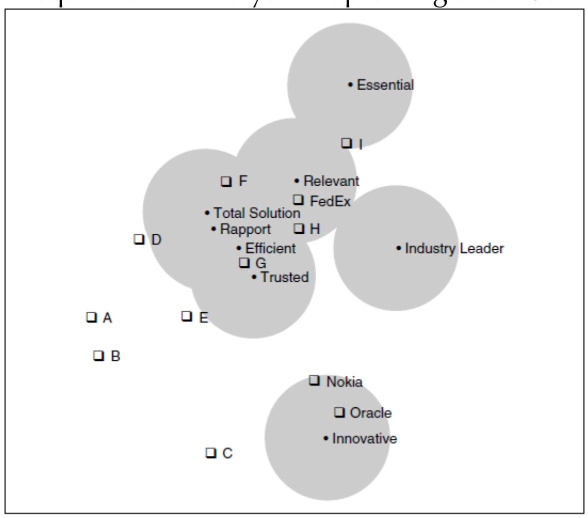
Correspondence analysis can be viewed as a special case of canonical correlation analysis that is analogous to a principal components factor analysis for nominal data. Canonical correlation, as we have just seen, examines the relations between two sets of continuous variables; correspondence analysis examines the relations between the *categories of two discrete variables*. Correspondence analysis can be applied to many forms of contingency table data, including frequency counts, associative data (pick *k* of *n*), or dummy variables. This analysis develops ratio-scaled interpoint distances between the row and column categories that depict accurate and useful positioning maps.

Correspondence analysis is often used in positioning and image studies where the researcher wants to explore the relationships between brands, between attributes, and between brands and attributes. In strategic terms, the marketing researcher may want to identify (a) closely competitive brands, (b) important attributes, (c) how attributes cluster together, (d) a brand's competitive strengths, and most importantly (e) ideas for improving a brand's competitive position (Whitlark & Smith, 2001).

According to Clausen (1998, p. 1), the main purpose of correspondence analysis is twofold:

1. To reveal the relationships in a complex set of variables by replacing the data with a simpler data matrix without losing essential information.
2. To visually display the points in space. This helps interpretation. Correspondence analysis analyzes the association between two or more categorical variables and represents the categorical marketing research data with a two- or three-dimensional map.

Figure 14.3 Example Correspondence Analysis Map of Logistical Services Providers



Another way of thinking about correspondence analysis is as a special type of cross-tabulation analysis of contingency tables. Categories with similar distributions will be placed in close proximity to each other, while those with dissimilar distributions will be farther apart. The technique is capable of handling large contingency tables.

An example from Whitlark and Smith (2004) will be used to illustrate this technique. When administering long and complicated surveys, it is sometimes impractical to collect attribute ratings for all brands and products the researcher is interested in. The respondent may have little or no knowledge of some brands, and providing many ratings can be an onerous task that produces poor quality data. Obviously the respondent should not be asked to evaluate unfamiliar brands. If the list of brands is large, the respondent will be asked to only evaluate a subset of the list that contains familiar brands. But more to the point of using correspondence analysis, in this situation the researcher will often simplify the data collection task by not using rating scales, but will instead give the respondent a list of attributes and ask them to check off the ones they feel best describe a particular brand that they are familiar with. This type of question produces “pick *k* of *n*” associative data, where *k* represents the number of attributes a respondent associates with a brand and *n* represents the total number of descriptive attributes included in the survey.

The correspondence analysis map shown in Figure 10.9 describes 12 companies providing communications, logistics consulting, and software support to a group of international freight handlers and shippers. Three well-known companies in the United States—Oracle, Nokia, and FedEx—are labeled using their names, and a series of nine less-familiar companies are labeled using letters of the alphabet. The 12 companies were evaluated by nearly 800 freight handlers, who indicated which attributes (pick *k* of *n*) best described the companies. The two-dimensional map of the companies accounts for nearly 90 percent of the variance in the data.

The results of this analysis show that Oracle and Nokia are perceived as being the most innovative and as industry leaders, while FedEx offers a relevant and total solution.

Correspondence analysis is a very helpful and interesting analysis tool that provides meaning and interpretation to large, complex data sets that contain this type of data. A more detailed explanation of this technique will be found in the excellent works by Clausen (1998), Greenacre (1993), and Carroll, Green, and Schaffer (1986; 1987).

BASIC CONCEPTS OF CLUSTER ANALYSIS

Like factor analysis, clustering methods are most often applied to object · variable matrices. However rather than focus on the similarity of variables as in factor analysis, the usual objective of *cluster analysis* is to separate objects (or people) into groups such that we maximize the similarity of objects within each group, while maximizing the differences between groups. Cluster analysis is thus concerned ultimately with classification, and its techniques are part of a field of study called *numerical taxonomy* (Sokal & Sneath, 1963; Sneath & Sokal, 1973). Cluster analysis can also be used to (a) investigate useful conceptual schemes derived from grouping entities; (b) generate a hypothesis through data exploration; and (c) attempt to determine if types defined through other procedures are present in a data set (Aldenderfer & Blashfield, 1984). Thus, cluster analysis can be viewed as *a set of techniques designed to identify objects, people, or variables that are similar with respect to some criteria or characteristics*. As such, it seeks to describe so-called *natural groupings*, as described in Exhibit 14.3.

EXHIBIT 14.3 Clustering for Segmentation

From a marketing perspective, it should be made clear that a major application of cluster analysis is for *segmentation*. To illustrate, consider a financial services company that wanted to do a segmentation study among its sales force of dealers/agents (Swint, 1994/1995). The objective was to identify the characteristics of "high producers" and "mediocre producers" of sales revenue. The desire was to *profile* the dealers/agents and segment them with respect to motivations, needs, work styles, beliefs, and behaviors. The data were analyzed using cluster analysis, and six cluster solutions emerged. The six clusters were then subject to discriminant analysis to how well the individual clustering attributes actually discriminated between the segments. The end result of all these analyses was six well defined clusters that identified the producer segments.

The type of clustering procedure that we shall discuss assigns each respondent (object) to one and only one class. Objects within a class are usually assumed to be indistinguishable from one another. Thus in cluster analysis, we assume here that the underlying structure of the data involves an unordered set of discrete classes. In some cases we may also view these classes as hierarchical in nature, where some classes are divided into subclasses.

Primary Questions

Clustering procedures can be viewed as preclassificatory in the sense that the analyst has *not* used prior information to partition the objects (rows of the data matrix) into groups. We note that partitioning is performed on the objects rather than the variables; thus, cluster analysis deals with intact data (in terms of the variables). Moreover, the partitioning is not performed *a priori* but is based on the object similarities themselves. Thus, the analyst *is* assuming that clusters exist. This type of presupposition is different from the case in discriminant analysis (discussed in Chapter 13), where groups of objects are predefined by a variable:

- Most cluster-analysis methods are relatively simple procedures that are usually not supported by an extensive body of statistical reasoning.
- Cluster-analysis methods have evolved from many disciplines, and the inbred biases of these disciplines can differ dramatically.
- Different clustering methods can (and do) generate different solutions from the same data set.
- The strategy of cluster analysis is structure-seeking, although its operation is structure imposing.

Given that no information on group definition in advance, we can identify four important considerations in selecting (or developing) cluster analysis algorithms. We must decide:

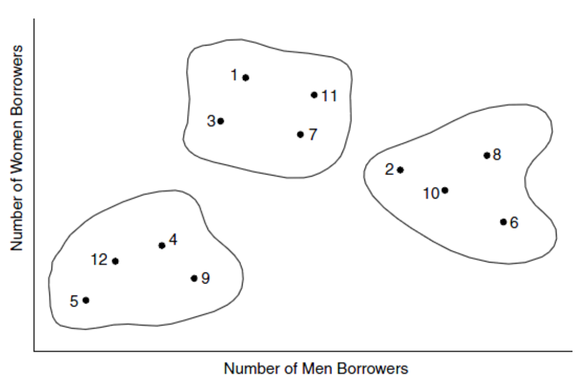
1. What measure of inter-object similarity is to be used, and how is each variable to be weighted in the construction of such a summary measure?
2. After inter-object similarities are obtained, how are the classes of objects to be formed?
3. After the classes have been formed, what summary measures of each cluster are appropriate in a descriptive sense—that is, how are the clusters to be defined?
4. Assuming that adequate descriptions of the clusters can be obtained, what inferences can be drawn regarding their statistical reliability?

Choice of Proximity Measure

The choice of a *proximity*, *similarity*, or *resemblance measure* (all three terms will be used synonymously here) is an interesting problem in cluster analysis. The concept of similarity always raises the question: Similarity with respect to what? Proximity measures are viewed in relative terms—two objects are similar, relative to the group, if their profiles across variables are close or if they share many aspects in common, relative to those which other pairs share in common.

Most clustering procedures use pairwise measures of proximity. The choice of which objects and variables should be included in the analysis, and how they should be scaled, is largely a matter of the researcher's judgment. The possible measures of pairwise proximity are many. Generally speaking, these measures fall into two classes: (a) distance-type measures (Euclidean distance); and (b) matching-type measures. A simple application illustrating the nature of cluster analysis using a distance measure is shown in Exhibit 14.4.

EXHIBIT 14.4 A Simple Example of Cluster Analysis



We can illustrate cluster analysis by a simple example. The problem is to group a set of twelve branches of a bank into three clusters of four branches each. Groups will be formed based on two variables, the number of men who have borrowed money (*X*1) and the number of women who have borrowed money (*X*2). The branches are plotted in two dimensions in the figure.

We use a proximity measure, based on Euclidean distances between any two branches. Branches 2 and 10 appear to be the closest together. The first cluster is formed by finding the midpoint between branches 2 and 10 and computing the distance of each branch from this midpoint (this is known as applying the *nearest-neighbor algorithm*). The two closest branches (6 and 8) are then added to give the desired-size cluster. The other clusters are formed in a similar manner. When more than two dimensions (that is, characteristics) are involved, the algorithms become more complex and a computer program must be used for measuring distances and for performing the clustering process.

Selecting the Clustering Methods

Once the analyst has settled on a pairwise measure of profile similarity, some type of computational routine must be used to cluster the profiles. A large variety of such computer programs already exist, and more are being developed as interest in this field increases. Each clustering program tends to maintain a certain individuality, although some common characteristics can be drawn out. The following categories of clustering methods are based, in part, on the classification of Ball and Hall (1964):

1. *Dimensionalizing the association matrix.* These approaches use principal-components or other factor-analytic methods to find a dimensional representation of points from *interobject* association measures. Clusters are then developed on the basis of grouping objects according to their pattern of component scores.

2. *Nonhierarchical methods.* The methods start right from the proximity matrix and can be characterized in three ways:

   a. *Sequential threshold.* In this case a cluster center is selected and all objects within a prespecified distance threshold value are grouped. Then a new cluster center is selected and the process is repeated for the unclustered points, and so on. (Once points enter a cluster, they are removed from further processing.)

   b. *Parallel threshold.* This method is similar to the preceding method, except that several cluster centers are selected simultaneously and points within a distance threshold level are assigned to the nearest center; thresholds can then be adjusted to admit fewer or more points to clusters.

   c. *Optimizing partitioning.* This method modifies categories (a) or (b) in that points can later be reassigned to clusters on the basis of optimizing some overall criterion measure, such as average within-cluster distance for a given number of clusters.

3. *Hierarchical methods.* These procedures are characterized by the construction of a hierarchy or tree-like structure. In some methods each point starts out as a unit (single point) cluster. At the next level the two closest points are placed in a cluster. At the following level a third point joins the first two, or else a second two-point cluster is formed based on various criterion functions for assignment. Eventually all points are grouped into one larger cluster. Variations on this procedure involve the development of a hierarchy from the top down. At the beginning the points are partitioned into two subsets based on some criterion measure related to average within-cluster distance. The subset with the highest average within-cluster distance is next partitioned into two subsets, and so on, until all points eventually become unit clusters.

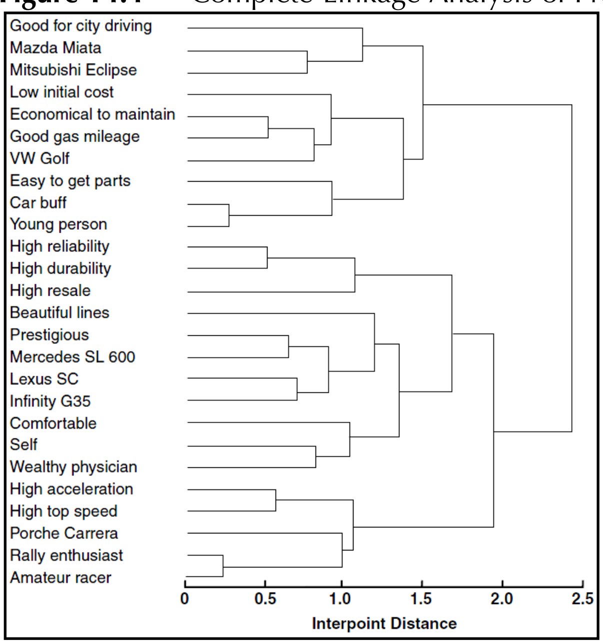
While the above classes of programs are not exhaustive of the field, most of the more widely used clustering routines can be classified as falling into one (or a combination) of the above categories. Criteria for grouping include such measures as average within-cluster distance and threshold cutoff values. The fact remains, however, that even the optimizing approaches achieve only conditional optima, since an unsettled question in this field is *how many* clusters to form in the first place.

A Product-Positioning Example of Cluster Analysis

Cluster analysis can be used in a variety of marketing research applications. For example, companies are often interested in determining how their products are positioned in terms of competitive offerings and consumers' views about the types of people most likely to own the product.

For illustrative purposes, Figure 14.4 shows the result of a hypothetical study conducted for seven sport cars, six types of stereotyped owners, and 13 attributes often used to describe cars.

Figure 14.4 Complete-Linkage Analysis of Product-Positioning Data



The inter-object distance data were based on respondents' degree-of-belief ratings about which attributes and owner types described each car. In this case, a complete-linkage algorithm was also used to cluster the objects (Johnson, 1967). The complete linkage algorithm starts by finding the two points with the minimum Euclidean distance. However, joining points to clusters is accomplished by maximizing the distance from a point in the first cluster to a point in the second cluster. Looking first at the four large clusters, we note the *car* groupings:

- Mazda Miata, Mitsubishi Eclipse
- VW Golf
- Mercedes 600 SL, Lexus SC, Infinity G35
- Porsche Carrera

In this example, the Porsche Carrera is seen as being in a class by itself, with the attributes *high acceleration* and *high top speed.* Its' perceived (stereotyped) owners are *rally enthusiast* and *amateur racer.*

Studies of this type enable the marketing researcher to observe the interrelationships among several types of entities—cars, attributes, and owners. This approach has several advantages. For example, it can be applied to alternative advertisements, package designs, or other kinds of communications stimuli. That is, the respondent could be shown blocks of advertising copy (brand unidentified) and asked to provide degree-of-belief ratings that the brand described in the copy possesses each of the *n* features.

Similarly, in the case of consumer packaged goods, the respondent could be shown alternative package designs and asked for degree-of-belief ratings that the contents of the package possess various features. In either case one would be adding an additional set (or sets) of ratings to the response sets described earlier. Hence, four (or more) classes of items could be represented as points in the cluster analysis.

Foreign Market Analysis

Companies considering entering foreign markets for the first time, as well as those considering expanding from existing to new foreign markets, have to do formal market analysis. Often a useful starting point is to work from a categorization schema of potential foreign markets. Cluster analysis can be useful in this process.

To illustrate, we use the study by Green and Larsen (1985). In this study, 71 nations were clustered on the basis of selected economic characteristics and economic change. The specific variables used were (a) growth in Gross Domestic Product; (b) literacy rate; (c) energy consumption per capita; (d) oil imports; and (e) international debt. Variables a, d, and e were measured as the change occurring during a specified time period.

Table 14.4 Composition of Foreign Market Clusters

CLUSTER 1		*CLUSTER 3*		*CLUSTER 5*	
Belgium	Finland	Ethiopia		Colombia	Venezuela
Canada	Norway	Ghana		Costa Rica	Yugoslavia
Denmark	Switzerland	India		Ecuador	El Salvador
Sweden	New Zealand	Liberia		Greece	Iran
USA	France	Libya		Hong Kong	Tunisia
Germany	Ireland	Madagascar		Jordan	Indonesia
Netherlands	Italy	Mali		Mexico	Nigeria
UK	Austria	Senegal		Paraguay	Malawi
Australia				Portugal	
CLUSTER 2		*CLUSTER 4*			
Cameroon	Honduras	Brazil	Uruguay		
Central African Republic	Nicaragua	Chile	Spain		
Egypt	Morocco	Israel	Sri Lanka		
Somalia	Ivory Coast	Japan	Thailand		
Togo	Tanzania	Korea	Turkey		
Zaire	Pakistan	Peru	Argentina		
Zambia	Philippines	Guatemala			
	Singapore	Kenya			

SOURCE: From Green, R. T. & Larsen, T., *Export Markets and Economic Change,* 1985 (working paper).Reprinted with permission.

Clustering was accomplished by use of a *K*-means clustering routine. This routine is a nonhierarchical method that allocates countries to the group whose centroid is closest, using a

Euclidean distance measure. A total of five clusters was derived based on the distance between countries and the centers of the clusters across the five predictor variables. The number of clusters selected was based on the criteria of total within-cluster distance and interpretability. A smaller number of clusters led to a substantial increase of within-cluster variability, while an increase in the number of clusters resulted in group splitting with a minimal reduction in distance. The composition of the clusters is shown in Table 14.4.

Computer Analyses

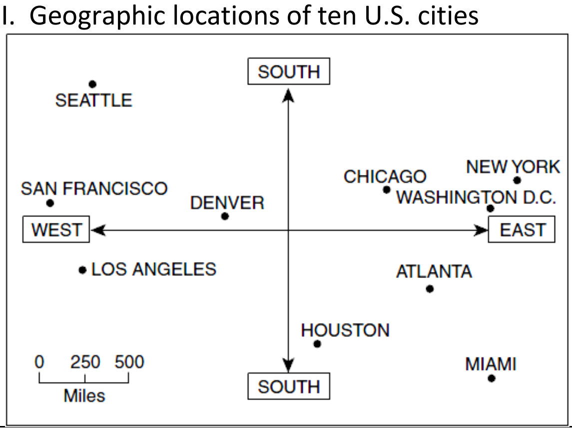
There are many computer programs available for conducting cluster analysis. Most analysis packages have one or more routines. Smaller, more specialized packages (such as PC-MDS) for cluster routines are also available. Finally, some academicians have developed their own cluster routines that they typically make available to other academicians for no charge.

MULTIDIMENSIONAL SCALING (MDS) ANALYSIS

Multidimensional scaling is concerned with portraying psychological relations among stimuli—either empirically-obtained similarities, preferences, or other kinds of matching or ordering—as geometric relationships among points in a multidimensional space. In this approach one represents *psychological dissimilarity as geometric distance*. The axes of the geometric space, or some transformation of them, are often (but not necessarily) assumed to represent the psychological bases or attributes along which the judge compares stimuli (represented as points or vectors in his or her psychological space).

Figure 14.5 Nonmetric MDS of 10 U.S. Cities

I. Geographic locations of ten U.S. cities



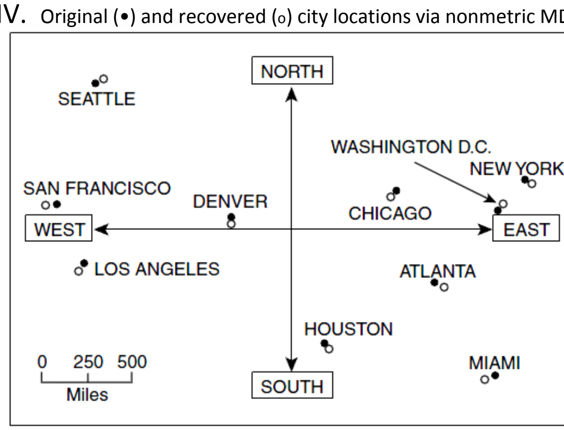
II. Airline distance between ten U.S. cities

CITIES	ATLA.	CHIC.	DENV.	HOUS.	L.A.	MIAMI	N.Y.	S.F.	SEAT.	WASH. D.C.
ATLANTA		587	1212	701	1936	604	748	2139	2182	543
CHICAGO			920	940	1745	1188	713	1858	1737	597
DENVER				879	831	1726	1631	949	1021	1494
HOUSTON					1374	968	1420	1645	1891	1220
LOS ANGELES						2339	2451	347	959	2300
MIAMI							1092	2594	2734	923
NEW YORK								2571	2408	205
SAN FRANCISCO									678	2442
SEATTLE										2329
WASHINGTON D.C.										

III. Nonmetric (Rank Order) Distance Data

CITIES	ATL	CHI	DEN	HOU	LA	MIA	NY	SF	SEA	D.C.
ATL		4	22	8	34	6	10	35	36	3
CHI			13	15	31	21	9	32	30	5
DEN				12	11	29	27	16	19	26
HOU					24	18	25	28	33	23
LA						39	42	2	17	37
MIA							20	44	45	14
NY								43	40	1
SF									7	41
SEA										38
D.C.										

IV. Original (•) and recovered (○) city locations via nonmetric MDS



In this section we start with an intuitive introduction to multidimensional scaling that uses a geographical example involving a set of intercity distances. In particular, we show how MDS takes a set of distance data and tries to find a spatial configuration or pattern of points in some number of dimensions whose distances best match the input data.

Let us start by looking at Panel I of Figure 14.5. Here we see a configuration of ten U.S. cities, whose locations have been taken from an airline map (Kruskal & Wish, 1978). The actual intercity distances are shown in Panel II. The Euclidean distance between a pair of points *i* and *j*, in any number of *r* dimensions, is given by

$$d_{ij} = \left[\sum_{k=1}^{r} (x_{ik} - x_{jk})^2 \right]^{1/2}$$

In the present case, $r = 2$ and only two dimensions are involved. For example, we could use the map to find the distance between Atlanta and Chicago by (a) projecting their points on axis 1 (East-West), finding the difference, and squaring it; (b) projecting their points on axis 2 (North-South) and doing the same; and then (c) taking the square root of the sum of the two squared differences.

In short, it is a relatively simple matter to go from the map in Panel I to the set of numerical distances in Panel II. However, the converse is *not* so easy. And that is what MDS is all about.

Suppose that we are shown Panel II of Figure 14.4 without the labels so that we do not even know that the objects are cities. The task is to work backward and develop a spatial map. That is, we wish to find, simultaneously, the number of dimensions and the configuration (or pattern) of points in that dimensionality so that their computed interpoint distances most closely match the input data of Panel II. This is the problem of *metric* MDS.

Next, suppose that instead of the more precise air mileage data, we have only rank order data. We can build such a data matrix by taking some order-preserving transformation in Panel II to produce Panel III. For example, we could take the smallest distance (205 miles between New York and Washington) and call it 1. Then we could apply the same rules and rank order the remaining 44 distances up to rank 45 for the distance (2,734 miles) between Miami and Seattle. We could use a nonmetric MDS program to find the number of dimensions and the configuration of points in that dimensionality such that the ranks of their computed interpoint distances most closely match the ranks of the input data.

In this example where the actual distance data is considered, it turns out that metric MDS methods can find, for all practical purposes, an exact solution (Panel I). However, what is rather surprising is that, even after downgrading the numerical data to ranks, nonmetric methods can also achieve a virtually perfect recovery as well.

Panel IV shows the results of applying a nonmetric algorithm to the ranks of the 45 numbers in Panel III. Thus, even with only rank-order input information, the recovery of the original locations is almost perfect.

We should quickly add, however, that neither the metric nor nonmetric MDS procedures will necessarily line up the configuration of points in a North-South direction; all that the methods try to preserve are *relative* distances. The configuration can be arbitrarily rotated, translated, reflected, or uniformly stretched or shrunk by so-called configuration congruence or matching programs, so as to best match the target configuration of Panel I. None of these operations will change the *relative* distances of the points.

Psychological Versus Physical Distance

The virtues of MDS methods are not in the scaling of physical distances but rather in their scaling of *psychological distances*, often called *dissimilarities*. In MDS we assume that individuals act as though they have a type of "mental map", (not necessarily visualized or verbalized), so that they view pairs of entities near each other as similar and those far from each other as dissimilar. Depending on the relative distances among pairs of points, varying *degrees* of dissimilarity could be imagined.

We assume that the respondent is able to provide either numerical measures of his or her perceived degree of dissimilarity for all pairs of entities, or, less stringently, ordinal measures of dissimilarity. If so, we can use the methodology of MDS to construct a *physical* map in one or more dimensions whose interpoint distances (or ranks of distances, as the case may be) are most consistent with the input data.

This model does not explain perception. Quite the contrary, it provides a useful *representation* of a set of subjective judgments about the extent to which a respondent views various pairs of entities as dissimilar. Thus, MDS models are representations of data rather than theories of perceptual processes.

Classifying MDS Techniques

Many different kinds of MDS procedures exist. Accordingly, it seems useful to provide a set of descriptors by which the methodology can be classified. These descriptors are only a subset of those described by Carroll and Arabie (1998); and Green, Carmone and Smith (1989)

1. *Mode:* A mode is a class of entities, such as respondents, brands, use occasions, or attributes of a multiattribute object.
2. *Data array:* The number of ways that modes are arranged. For example, in a two-way array of single mode dissimilarities, the entities could be brand-brand relationships, such as a respondent's rating of the *ij*th brand pair on a 1–9 point scale, ranging from 1 (very similar) to 9 (very different). Hence, in this case, we have one mode, two-way data on judged dissimilarities of pairs of brands.
3. *Type of geometric model:* Either a distance model or a vector or projection model (the latter represented by a combination of points and vectors).
4. *Number of different sets of plotted points (or vectors)*: One, two, or more than two.
5. *Scale type:* Nominal-, ordinal-, interval-, or ratio-scaled input data.

Data Mode/Way

In marketing research most applications of MDS entail either single mode, two-way data, or two-mode, two-way data. Single mode, two-way data are illustrated by input matrices that are square and symmetric, in which all distinct pairs of entities (e.g., brands) in a $I \cdot I$ matrix are often judgment data expressing relative similarity/dissimilarity on some type of rating scale. The data collection instructions can refer to the respondent making judgments that produce data representing pair-wise similarity, association, substitutability, closeness to, affinity for, congruence with, co-occurrence with, and so on. Typically, only $I(I - 1)/2$ pairs are evaluated and entered into the lower- or upper-half of the matrix. This is the case because of the symmetric relationship that is assumed to exist. MDS solutions based on single mode, two-way input data lead to what are often called *simple spaces*—that is, a map that portrays only the set of

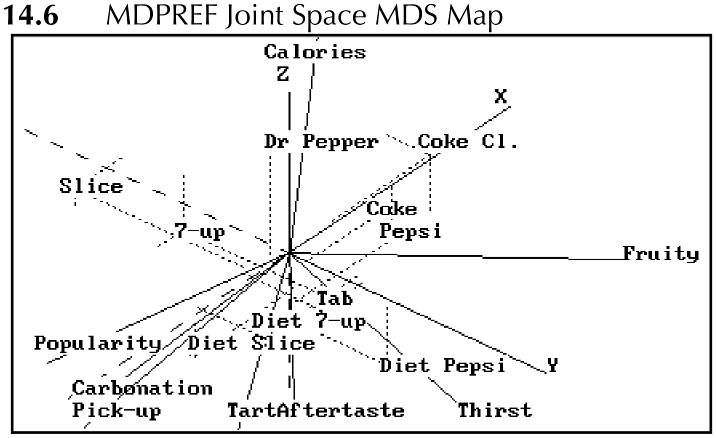
I points, as was shown in Figure 14.4. Pairs of points close together in this geometric space are presumed to exhibit high subjective similarity in the eyes of the respondent.

Another popular form of marketing research data entails input matrices that represent two-mode, two-way relationships, such as the following six examples:

1. A set of *I* judges provide preference ratings of *J* brands
2. Average scores (across respondents) of *J* brands rated on *I* attributes
3. The frequency (across respondents) with which *J* attributes are assumed to be associated with *I* brands
4. The frequency (across respondents) with which respondents in each of *I* brand-favorite groups pick each of *J* attributes as important to their brand choice
5. The frequency (across respondents) with which each of *J* use occasions is perceived to be appropriate for each of *I* brands
6. The frequency (across respondents) with which each of *J* problems is perceived to be associated with using each of *I* brands.

These geometric spaces are often called *joint spaces* in that two different sets of points (e.g., brands and attributes) are represented in the MDS map. In Figure 14.6 we observe that brands are positioned in 3-dimensional space as points and the attributes are vectors extending a distance of 1 unit away from the origin. The brands project onto each attribute vector to define their degree of association with that attribute. The further out on the vector the brand projects, the stronger the association. In some cases three or more sets of entities may be scaled.

Figure 14.6 MDPREF Joint Space MDS Map



Type of Geometric Model

In applications of single-mode, two-way data the entities being scaled are almost always represented as points (as opposed to vectors). However, in the case of two-mode, two-way data, the two sets of entities might each be represented as points or, alternatively, one set may be represented as points while the other set is represented as vector directions. In the latter case the termini of the vectors are often normalized to lie on a common circumference around the origin of the configuration.

The point-point type of two-mode, two-way data representation is often referred to as an *unfolding* model (Coombs, 1964). If the original matrix consists of *I* respondents' preference evaluations of *J* brands, then the resulting joint-space map has *I* respondents' ideal points and *J* brand points. Brand points that are near a respondent's ideal point are assumed to be *highly* *preferred* by that respondent. Although the original input data may be based on between-set relationships, if the simple unfolding model holds, one can also infer respondent to- respondent similarities in terms of the closeness of their ideal points to each other. Brand to- brand similarities may be analogously inferred, based on the relative closeness of pairs of brand points.

The point-vector model of two-mode, two-way data is a *projection* model in which one obtains respondent *i*'s preference scale by projecting the *J* brand points onto respondent *i*'s vector (Figure 14.5). Point-vector models also show ideal points or points of "ideal preference". This ideal point is located at the terminus or end of the vector. Projections are made by drawing a line so that it intersects the vector at a 90-degree angle. The farther out (toward vector *i*'s terminus) the *projection* is, the more preferred the brand is for that respondent.

Collecting Data for MDS

The content side of MDS—dimension interpretation, relating physical changes in products to psychological changes in perceptual maps—poses the most difficult problems for researchers. However, methodologists are developing MDS models that provide more flexibility than a straight dimensional application. For example, recent models have coupled the ideas of cluster analysis and MDS into hybrid models of categorical-dimensional structure. Furthermore, conjoint analysis, to be discussed next, offers high promise for relating changes in the physical (or otherwise controlled) aspects of products to changes in their psychological imagery and evaluation. Typically, conjoint analysis deals with preference (and other dominance-type) judgments rather than similarities. However, more recent research has extended the methodology to similarities judgments.

On the input side, there are issues that arise concerning data collection methods. In MDS studies, there are four most commonly used methods of data collection:

- *Sorting task.* Subjects are asked to sort the stimuli into a number of groups, according to similarity. The number of groups is determined by the subject during the judgment task.
- *Paired comparison task.* Stimuli are presented to subjects in all possible pairs of two stimuli. Each subject has to rate each pair on an ordinal scale (the number of points can vary) where the extreme values of the scale represent maximum dissimilarity and maximum similarity.
- *Conditional ranking task.* Subjects order stimuli on the basis of their similarity with an anchor stimulus. Each of the stimuli is in turn presented as the anchor.
- *Triadic combinations task.* Subjects indicate which two stimuli of combinations of three stimuli form the most similar pair, and which two stimuli form the least similar pair.

When subjects perform a similarity (or dissimilarity) judgment they may experience increases in fatigue and boredom (Bijmolt and Wedel,1995, p. 364).

Bijmolt and Wedel examined the effect of the alternative data collection methods on fatigue, boredom and other mental conditions. They showed that when collecting data, conditional rankings and triadic combinations should be used only if the stimulus set is relatively small, and in situations where the maximum amount of information is to be extracted from the respondents. If the stimulus set is relatively large, sorting and paired comparisons are better suited for collecting similarity data. Which of these two to use will depend on characteristics of the application, such as number of stimuli and whether or not individual-level perceptual maps are desired.

Marketing Applications of MDS

MDS studies have been used in a variety of situations to help marketing managers see how their brand is positioned in the minds of consumers, vis-à-vis competing brands. Illustrations include (a) choosing a slogan for advertising a soft drink, (b) the relationship between physical characteristics of computers and perceptions of users and potential users, (c) effectiveness of a new advertising campaign for a high-nutrition brand of cereal, (d) positioning in physicians' minds of medical magazines and journals, and (e) positioning of new products and product concepts. There is no shortage of applications in real-world marketing situations.

Current research activity in MDS methods, including the increasing use of correspondence analyses for representing nominal data (Hoffman & Franke, 1986; Carroll, Green, & Schaffer, 1986; Whitlark & Smith, 2003), shows few signs of slowing down. In contrast, industry applications for the methods still seem to be emphasizing the graphical display and diagnostic roles that characterized the motivation for developing these techniques in the first place. The gap between theory and practice appears to be widening. A comprehensive overview of the developments in MDS is provided by Carroll and Arabie (1998).

FUNDAMENTALS OF CONJOINT ANALYSIS

Conjoint analysis is one of the most widely used advanced techniques in marketing research. It is a powerful tool that allows the researcher to predict choice share for evaluated stimuli such as competitive brands. When using this technique the researcher is concerned with the identification of *utilities*—values used by people making tradeoffs and choosing among objects having many attributes and/or characteristics.

There are many methodologies for conducting conjoint analysis, including two-factor-at a- time tradeoff, full profile, Adaptive Conjoint Analysis (ACA), choice-based conjoint, self explicated conjoint, hybrid conjoint, and Hierarchical Bayes (HB). In this chapter, two of the most popular methodologies are discussed: the full-profile and self-explicated models.

Conjoint analysis, like MDS, concerns the measurement of psychological judgments, such as consumer preferences. The stimuli to be presented to the respondent are often designed beforehand according to some type of factorial structure. In full-profile conjoint analysis, the objective is to decompose a set of overall responses to a set of stimuli (product or service attribute descriptions) so that the utility of each attribute describing the stimulus can be inferred from the respondent's *overall evaluations* of the stimuli. As an example, a respondent might be presented with a set of alternative product descriptions (automobiles). The automobiles are described by their stimulus attributes (level of gas mileage, size of engine, type of transmission, etc.). When choice alternatives are presented, choice or preference evaluations are made. From this information, the researcher is able to determine the respondent's utility for each stimulus attribute (i.e., what is the relative value of an automatic versus a five-speed manual transmission). Once the utilities are determined for all respondents, simulations are run to determine the relative choice share of a competing set of new or existing products.

Conjoint analysis models are constrained by the amount of data required in the data collection task. Managers demand models that define products with increasingly more stimulus attributes and levels within each attribute. Because more detail increases the size, complexity, and time of the evaluation task, new data collection methodologies and analysis models are continually being developed.

One early conjoint data collection method presented a series of attribute-by-attribute (two attributes at a time) tradeoff tables where respondents ranked their preferences of the different combinations of the attribute levels. For example, if each attribute had three levels, the table would have nine cells and the respondents would rank their tradeoff preferences from 1 to 9. The two-factor-at-a-time approach makes few cognitive demands of the respondent and is simple to follow . . . but it is both time-consuming and tedious. Moreover, respondents often lose their place in the table or develop some stylized pattern just to get the job done. Most importantly, however, the task is unrealistic in that real alternatives do not present themselves for evaluation two attributes at a time.

For the last 30 years, full-profile conjoint analysis has been a popular approach to measure attribute utilities. In the full-profile conjoint task, different product descriptions (or even different actual products) are developed and presented to the respondent for acceptability or preference evaluations. Each product profile is designed as part of a *fractional factorial experimental design* that evenly matches the occurrence of each attribute with all other attributes. By controlling the attribute pairings, the researcher can estimate the respondent's utility for each level of each attribute tested.

A third approach, Adaptive Conjoint Analysis, was developed to handle larger problems that required more descriptive attributes and levels. ACA uses computer-based interviews to adapt each respondent's interview to the evaluations provided by each respondent. Early in the interview, the respondent is asked to eliminate attributes and levels that would not be considered in an acceptable product under any conditions. ACA next presents attributes for evaluation and finally full profiles, two at a time, for evaluation. The choice pairs are presented in an order that increasingly focuses on determining the utility associated with each attribute.

A fourth methodology, choice-based conjoint, requires the respondent to choose a preferred full-profile concept from repeated sets of 3–5 concepts. This choice activity simulates an actual buying situation, thereby giving the respondents a familiar task that mimics actual shopping behavior.

The self-explicated approach to conjoint analysis offers a simple but robust approach that does not require the development or testing of full-profile concepts. Rather, the conjoint factors and levels are presented to respondents for elimination if not acceptable in products under any condition. The levels of the attributes are then evaluated for desirability. Finally, the relative importance of attributes is derived by dividing 100 points between the most desirable levels of each attribute. The respondent's reported attribute level desirabilities are weighted by the attribute importances to provide utility values for each attribute level. This is done without the regression analysis or aggregated solution required in many other conjoint approaches. This approach has been shown to provide results equal or superior to full-profile approaches, and requires less rigorous evaluations from respondents.

Most recently, academic researchers have focused on an approach called Hierarchical Bayes (HB) to estimate attribute level utilities from choice data. HB uses information about the distribution of utilities from all respondents as part of the procedure to estimate attribute level utilities for each individual. This approach again allows more attributes and levels to be estimated with smaller amounts of data collected from each individual respondent.

An Example of Full-Profile Conjoint Analysis

In *metric conjoint analysis*, the solution algorithm involves a dummy variable regression analysis in which the respondent's preference ratings of the product profile (service or other item) being evaluated serve as the dependent (criterion) variable, and the independent (predictor) variables are represented by the various factorial levels making up each stimulus. In the *nonmetric* version of conjoint analysis, the dependent (criterion) variable represents a ranking of the alternative profiles and is only ordinal-scaled. The full-profile methods for collecting conjoint analysis data will be illustrated to show how conjoint data are obtained.

The *multiple-factor approach* illustrated in Figure 14.7 consists of sixteen cards, each made up according to a special type of factorial design. The details of each card are shown on the left side of Figure 14.6.

Figure 14.7 Multiple-Factor Evaluations (Sample Profiles)

Advertising Appeal Study for Allergy Medication 4^4 Fractional Factorial Design (4 Levels$^{4\ \text{Factors}}$)	
0 , 0 , 0 , 0	Column 1: "Efficacy", 4 Levels
1 , 0 , 1 , 2	0="No med more effective"
2 , 0 , 2 , 3	1="No med works faster"
3 , 0 , 3 , 1	2="Relief all day"
0 , 1 , 1 , 1	3="Right Formula"
1 , 1 , 0 , 3	
2 , 1 , 3 , 2	Column 2: "Endorsements",4 Levels
3 , 1 , 2 , 0	0="Most recommended by allergists"
0 , 2 , 2 , 2	1="Most recommended by pharmacist"
1 , 2 , 3 , 0	2="National Gardening Association."
2 , 2 , 0 , 1	3="Professional Gardeners (Horticulture.)"
3 , 2 , 1 , 3	
0 , 3 , 3 , 3	Column 3: "Superiority",4 Levels
1 , 3 , 2 , 1	0="Less sedating than Benadryl"
2 , 3 , 1 , 0	1="Rec. 2:1 over Benadryl"
3 , 3 , 0 , 2	2="Relief 2x longer than Benadryl"
	3="Leading long acting Over The Counter "
Legend:	
Profile 1 has levels 0,0,0,0:	Column 4: "Gardening",4 Levels
No Medication is More Effective	0="Won't quit on you"
Most Recommended by Allergists	1="Enjoy relief while garden"
Less Sedating than Benadryl	2="Brand used by millions"
Won't Quit on You	3="Relieves allergy symptoms"

All card descriptions differ in one or more attribute level(s).1 The respondent is then asked to group the 16 cards (Figure 14.8) into three piles (with no need to place an equal number in each pile) described in one of three ways:

- Definitely like
- Neither definitely like nor dislike
- Definitely dislike

The criterion variable is usually some kind of preference or purchase likelihood rating. Following this, the respondent takes the first pile and ranks the cards in it from most to least liked, and similarly so for the second and third piles. By means of this two-step procedure, the full set of 16 cards is eventually ranked from most liked to least liked.

While it would be easier for the respondent to rate each of the 16 profiles on, a 1-10 rating scale, the inability or unwillingness of the respondent to conscientiously differentiate between all of the profiles considered typically results in end-piling of ratings where many profiles incorrectly receive the same score values.

Again, the analytical objective is to find a set of part-worths or utility values for the separate attribute (factor) levels so that, when these are appropriately added, one can find a total utility for each combination or profile. The part-worths are chosen so as to produce the highest possible correspondence between the derived ranking and the original ranking of the 16 cards. While the two-factor-at-a-time and the multiple-factor approaches, as just described, assume only ranking-type data, one could just as readily ask the respondent to state his or her preferences on (say), an 11-point equal-interval ratings scale, ranging from like most to like least. Moreover, in the multiple-factor approach, a 0-to-100 rating scale, representing likelihood of purchase, also could be used.

Figure 14.8 Product Descriptions for Conjoint Analysis (Allergy Medication)

Card 1 Efficacy No med more effective Endorsements Most recom. by allergists Superiority Less sedating than Benadryl Gardening Won't quit on you	Card 2 Efficacy No med works faster Endorsements Most recom. by allergists Superiority Rec. 2:1 over Benadryl Gardening Brand used by millions	Card 3 Efficacy Relief all day Endorsements Most recom. by allergists Superiority Relief 2x longer than Benadryl Gardening Relieves allergy symptoms	Card 4 Efficacy Right Formula Endorsements Most recom. by allergists Superiority Leading long acting OTC Gardening Enjoy relief while garden
Card 5 Efficacy No med more effective Endorsements Most recom. by pharmacist Superiority Rec. 2:1 over Benadryl Gardening Enjoy relief while garden	Card 6 Efficacy No med works faster Endorsements Most recom. by pharmacist Superiority Less sedating than Benadryl Gardening Relieves allergy symptoms	Card 7 Efficacy Relief all day Endorsements Most recom. by pharmacist Superiority Leading long acting OTC Gardening Brand used by millions	Card 8 Efficacy Right Formula Endorsements Most recom. by pharmacist Superiority Relief 2x longer than Ben Gardening Won't quit on you
Card 9 Efficacy No med more effective Endorsements Nat. Gardening Assoc. Superiority Relief 2x longer than Ben Gardening Brand used by millions	Card 10 Efficacy No med works faster Endorsements Nat. Gardening Assoc. Superiority Leading long acting OTC Gardening Won't quit on you	Card 11 Efficacy Relief all day Endorsements Nat. Gardening Assoc. Superiority Less sedating than Benadryl Gardening Enjoy relief while garden	Card 12 Efficacy Right Formula Endorsements Nat. Gardening Assoc. Superiority Rec. 2:1 over Benadryl Gardening Relieves allergy symptoms
Card 13 Efficacy No med more effective Endorsements Prof. Gardeners (Horticul.) Superiority Leading long acting OTC Gardening Relieves allergy symptoms	Card 14 Efficacy No med works faster Endorsements Prof. Gardeners (Horticul.) Superiority Relief 2x longer than Ben Gardening Enjoy relief while garden	Card 15 Efficacy Relief all day Endorsements Prof. Gardeners (Horticult.) Superiority Rec. 2:1 over Benadryl Gardening Won't quit on you	Card 16 Efficacy Right Formula Endorsements Prof. Gardeners (Horticul.) Superiority Less sedating than Benadryl Gardening Brand used by millions

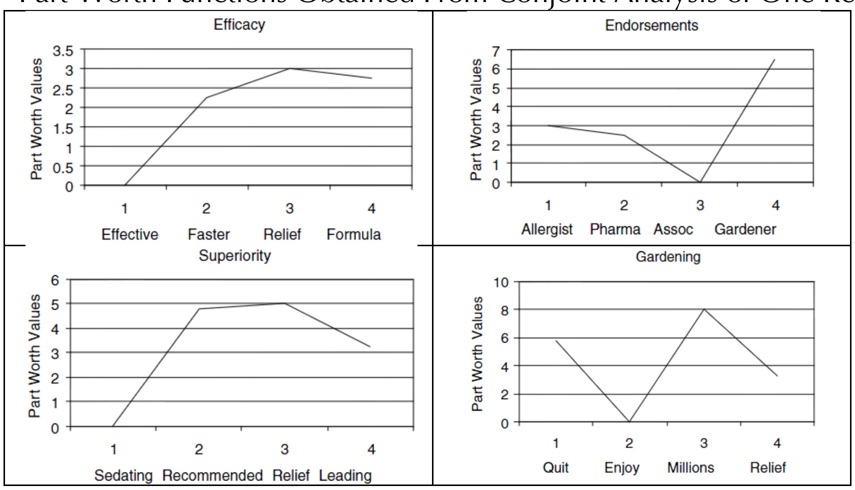
As may be surmised, the multiple-factor evaluative approach makes greater cognitive demands on the respondent, since the full set of factors appears each time. In practice, if more than six or seven factors are involved, this approach is often modified to handle specific *subsets* of interlinked factors across two or more evaluation tasks.

Consider the situation in which a manufacturer of over-the-counter allergy medication is interested in measuring consumers' tradeoffs among the four attributes identified in Figure 14.6.

Figure 14.9 shows a table of the resulting utility values for each of the attribute levels derived for one respondent. These values can be obtained from an ordinary multiple regression program using dummy-variable coding. All one needs to do to estimate the respondent's utility score for a given concept profile is to add each separate value (the regression coefficient) for each component of the described combination. (The regression's intercept term may be added in later if there is interest in estimating the absolute level of purchase interest.) For example, to obtain the respondent's estimated evaluation of card 1, one sums the part-worths:

Value for: Efficacy "No med more effective" = 0.00
Value for: Endorsements "Most recom. by allergists" = 3.00
Value for: Superiority "Less sedating than Benadryl" = 0.00
Value for: Gardening "Won't quit on you" = 6.00
Total = 9.00

Figure 14.9 Part-Worth Functions Obtained From Conjoint Analysis of One Respondent



In this instance we obtain an almost perfect prediction of a person's overall response to card one. Similarly, we can find the estimated total evaluations for the other 15 options and compare them with the respondent's original evaluations. The regression technique guarantees that the (squared) prediction error between estimated and actual response will be minimized. The information in Figure 14.8 also permits the researcher to find estimated evaluations for *all* combinations, including the 256 – 16 = 240 options never shown to the respondent. Moreover, all respondents' separate part-worth functions (as illustrated for the average of all respondents in Table 14.6) can be compared in order to see if various types of respondents (e.g., high-income versus low-income respondents) differ in their separate attribute evaluations. In short, while the respondent evaluates complete bundles of attributes, the technique solves for a set of part-worths—one for each attribute level—that are imputed from the *overall* tradeoffs.

Table 14.6 Average Utilities of All Respondents

LEVEL	Importance %	1	2	3	4
Efficacy		Effective	Faster	Relief	Formula
Importance	14.63	2.48	2.37	2.10	2.18
Endorsements		Allergist	Pharmacist	Nat. Garden	Prof. Garden
Importance	54.24	3.83	3.18	2.57	2.41

Superiority		Less	Recommend	Relief	Leading
Importance	17.76	2.78	2.63	2.71	2.31
Gardening		Quit	Enjoy	Millions	Relief
Importance	13.37	2.39	2.74	2.58	2.54

These part-worths can then be combined in various ways to estimate the evaluation that a respondent would give to *any* combination of interest. It is this high leverage between options that are actually evaluated and those that can be evaluated (after the analysis) that makes conjoint analysis a useful tool. It is clear that the full-profile approach requires much sophistication in developing the profiles and performing the regression analyses to determine the utilities. We will now consider the self-explicated model as an approach that provides results of equal quality, but does so with a much easier design and data collection task.

Self-Explicated Conjoint Analysis

The development of fractional factorial designs and required dummy regression for each respondent places a burden on the researcher and respondent alike, especially when the number of factors and levels require that a large number of profiles be presented to the respondent.

The self-explicated model provides a simple alternative producing utility score estimates equal to or superior to that of the ACA or full-profile regression models. The self-explicated model is based theoretically on the multi-attribute attitude models that combine attribute importance with attribute desirability to estimate overall preference. This model is expressed as

$$E_o = \sum_{j=1}^{m} \sum_{k=1}^{n} I_j D_{jk}$$

where *Ij* is the importance of attribute *j* and *Djk* is the desirability of level *k* of attribute *j*. In this model, *Eo*, the evaluation of profile for product or service *o*, is formed by summing the importance weighted desirabilities of the attributes and attribute levels that make up the profile.

The Self-Explicated Data Collection Task

Initially, all attribute levels are presented to respondents for evaluation to eliminate any levels that would not be acceptable in a product under any conditions. Next, the list of attribute levels is presented and each level is evaluated for desirability (0 –10 scale). Finally, based on these evaluations, the most desirable levels of all attributes are presented in a constant sum question where the relative importances of the attributes are evaluated. Using this information, the attribute importance scores are used to weight the standardized attribute level scores, thereby producing self-explicated utility values for each attribute level. This is done for each respondent and does not require a fractional factorial designs or regression analysis.

As with the full-profile model, these scores can be summed and simulations run to obtain a score for any profile of interest. This simple self-reporting approach is easier for the respondent to complete and straightforward in terms of determining the importance or desirability of attributes and attribute levels (Srinivasan, 1997). An easy to use online implementation of the self-explicated model is found at www.qualtrics.com. For this implementation, the conjoint analysis is automatically developed after the attribute level descriptors are entered into the question builder.

Conjoint Reliability and Validity Checks

Irrespective of the method used to carry out a conjoint analysis, it is useful to include the following ancillary analyses: (a) test-retest reliability; (b) a comparison of actual utilities with those of random respondents; and (c) an internal validity check on model-based utilities.

The test-retest reliability can be conducted by including a few replicate judgments (drawn from the original set of 16) at a later stage in the interview. The purpose here is to see if the judgments are highly enough correlated, on a test-retest basis, to justify the analysis of the respondent's data.

An internal validity check could, in the case of the allergy medication example, be carried out by collecting a few new evaluations (drawn randomly from the 240 stimulus combinations not utilized in Figure 14.8). These constitute a hold-out sample. Their rank order is to be predicted by the part-worths developed from the calibration sample of 16 combinations. Internal validity checks could be conducted in similar manner for other conjoint methods.

Other Models

So far our discussion has centered on the most widely applied conjoint models—a main effects model using rankings or ratings. Other models are available that permit some or all two-factor interactions to be measured (as well as main effects). Interactions occur when attribute levels combine to provide a differential effect. For example, this often happens in food products where combinations of attribute levels seemingly produce more acceptable combinations than would be predicted by the individual attribute levels when considered alone (oatmeal-raisin cookies vs. oatmeal cookies or raisin cookies). These models again make use of various types of fractional factorial designs or combine attributes. Specialized computer programs have been designed to implement them. In short, the users of conjoint analysis currently have a highly flexible set of models and data collection procedures to choose from.

OTHER ASPECTS OF CONJOINT ANALYSIS

The typical sequence that one goes through to implement a conjoint study involves four steps:

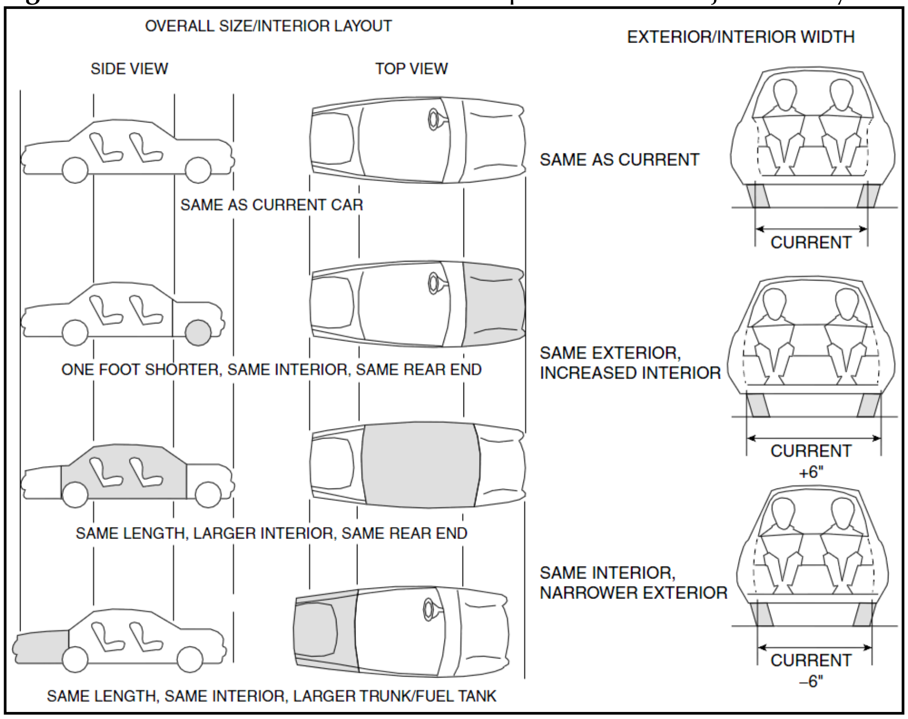
1. Using one of a variety of data collection procedures just described, obtain sufficient data at the individual respondent level to estimate the part-worths of each person's utility function.
2. Relate the respondent's attribute-level part-worth data to other subject background data in an effort to identify possible market segments based on similarities in part-worth functions.
3. Compose a set of product configurations that represent feasible competitive offerings. These product profiles are entered into a consumer choice simulator, along with the earlier computed individual utility functions.
4. Use the respondent's individual part-worth function to compute the utility for each of the competing profiles in the choice simulator. The respondent is then assumed to choose that profile with the highest utility (i.e., the choice process is deterministic).

Use of Visual Aids in Conjoint analysis

Another problem in the application of conjoint measurement is the pragmatic one of getting fairly complex concepts across to the respondent. Verbal descriptions of the type covered in Figure 14.8 are not only difficult for the respondent to assimilate, but also introduce unwanted perceptual differences. As an example, if conjoint analysis was used to test designs for automobiles; two respondents may have quite different perceptions of the car length and car-roominess if verbalizations were used.

Wherever possible, *visual props* can help to transmit complex information more easily and uniformly than verbal description. As an illustration of the value of visual props, mention can be made of a study involving styling designs for future compact cars. In the course of preparing the questionnaire, rather complex experimental factors such as overall size and interior layout, trunk size and fuel-tank capacity, exterior and interior width, and interior spaciousness and visibility had to be considered. To provide quick and uniform treatment of these style factors, visual props were prepared, as illustrated for two of the attributes in Figure 14.10. (These can be projected on screens in full view of the respondents during the interview or made part of the questionnaire itself.)

Figure 14.10 Illustrations of Visual Props Used on Conjoint Analysis



Visual props work particularly well for the multiple-factor approach, since a relatively large amount of information can be communicated realistically and quickly.

Strategic Aspects of Conjoint Analysis

The output of conjoint analysis is frequently employed in additional analyses. Since most conjoint analysis studies collect full sets of data at the individual respondent level, *individual utility functions and importance weights* can be computed. This fosters two additional types of analyses: (1) market segmentation and (2) strategic simulation of new factor-level combinations (most often used to test the viability of new product design configurations). Frequently, both kinds of analyses are carried out in the same study.

In segmentation studies, the respondents are usually clustered in terms of either their commonality of utility functions or their commonality of importance weights. Having formed the segments in one of these ways, the analyst can then determine how the segments differ with regard to other background data—product-class usage, brand-selection behavior, demographics, and so on.

Strategic simulations are also relatively easy to construct from conjoint analysis data by simply including each individual respondent's utility function in a computerized-choice model. Various combinations of factor levels can then be tried out to see what their share of choices would be under different assumptions regarding competitive offerings and total market demand.

The simulators can employ a variety of consumer-choice procedures, ranging from having each consumer simply select the alternative with the highest utility to more elaborate probability of choice rules, where probability is related to utility differences in the set of alternatives under evaluation.

Applications of Conjoint Analysis

Conjoint analysis is most conducive to predicting choice behavior when product or service involves a relatively high resource commitment and tends to be "analyzable" by the purchaser (e.g., banking or insurance services, industrial products). Conjoint analysis has already been applied to a wide variety of problems in product design, price elasticity of demand, transportation service design, and the like. Table 14.6 shows a representative list of applications. As can be noted, areas of application cover the gamut—products and services, as well as consumer, industrial, and institutional markets.

Table 14.6 Sample List of Conjoint Applications

Consumer Durables	***Other Products***	***Industrial Goods***
1. Automotive styling	1. Bar soaps	1. Copying machines
2. Automobile and truck tires	2. Hair shampoos	2. Printing equipment
3. Car batteries	3. Carpet cleaners	3. Facsimile transmission
5. Toaster ovens	4. Clothing: sweatshirts, bras	4. Data transmission
6. Cameras	5. Gasoline pricing	5. Personal computer design
Financial Services	***Other Services***	***Transportation***
1. Branch bank services	1. Car rental agencies	1. Domestic airlines
2. Auto insurance policies	2. Telephone services and pricing	2. Transcontinental airlines
3. Health insurance policies	3. Medical laboratories	3. Passenger train operations
4. Credit card features	4. Hotel design	4. Freight train operations
5. Consumer discount cards	5. Toll collection on toll ways	5. Electric car design

Recent Developments in Conjoint Analysis

Conjoint analysis has become a highly popular technique in a relatively short time. Researchers estimate that business firms' use of conjoint analysis entails several thousand studies each year. With statistical software and conjoint data collection algorithms built into online survey tools (Qualtrics.com), conjoint methodology is easily accessed by any interested user.

Software developments in data collection and analysis likewise make it easy to find orthogonal main effects plans. Conjoint methodology has also been extended to encompass use occasion and situation dependence in a series of dual-conjoint designs, called componential segmentation.

Perhaps the most interesting extension of the methodology, however, is the recent application of conjoint to the design of "optimal" products and product lines. Thus, it is feasible to extend conjoint beyond the simulation stage (where one finds the best of a limited set of options) to encompass the identification of the best product (or line) over the full set of possibilities. These may number in the hundreds of thousands or even the millions. In sum, conjoint methodology, like MDS, appears to be moving into the product-design-optimization arena, a most useful approach from a pragmatic managerial viewpoint.

Still, conjoint analysis, like MDS, has a number of limitations. For example, the approach assumes that the important attributes of a product or service can all be identified and that consumers behave rationally as though all tradeoffs are being considered. In some products where imagery is quite important, consumers may not evaluate a product analytically, or, even if they do, the tradeoff model may be only a gross approximation to the actual decision rules that are employed.

In short, MDS and conjoint are still maturing—both as techniques that provide intellectual stimulation and as practical tools for product positioning, segmentation, and strategic planning.

SUMMARY

Chapter 14 has focused on four multivariate techniques: factor analysis, cluster analysis, multidimensional scaling, and conjoint analysis.

The factor-analytic method stressed in this chapter was principal-components analysis. This procedure has the property of selecting sets of weights to form linear combinations of the original variables such that the variance of the obtained component scores is (sequentially) maximal, subject to each linear combination's being orthogonal to previously obtained ones.

The principal-components model was illustrated on a set of data from a study conducted by a grocery chain.

Cluster analysis was described in terms of three general questions: (a) selecting a proximity measure; (b) algorithms for grouping objects; and (c) describing the clusters. In addition, an application of clustering was briefly described.

MDS methods are designed to portray subjective similarities or preferences as points (or vectors) in some multidimensional space. Psychological distance is given a physical distance representation. We discussed metric and nonmetric MDS methods, and ideal-point and vector preference models. A variety of applications were described to give the reader some idea of the scope of the methodology.

Conjoint analysis was described along similar lines. We first discussed the primary ways of collecting tradeoff data and then showed how such data are analyzed via multiple regression with dummy predictor variables. The importance of fractional factorial designs was discussed, as well as other practical problems in the implementation of conjoint analysis. We next turned to some illustrative applications of conjoint analysis, including the design of new products and services. We then presented a brief description of future developments that could serve to increase the flexibility of conjoint methodology.

This chapter, together with Chapter 13, covers the major multivariate analysis techniques and has included brief discussions of lesser-used techniques. We have not discussed such extensions as canonical correlation of three or more sets of variables or tests for the equality of sums of squares and cross-products matrices. Other advanced, but related procedures such as moderated regression, multiple-partial correlation, discriminant analysis with covariate adjustment, factorial discriminant analysis, to name a few, have been omitted from discussion.

We have discussed the principal assumption structure of each technique, appropriate problems for applying it, and sufficient numerical applications to give the reader a feel for the kind of output generated by each program.

Our coverage of this vast and complex a set of methods is limited in depth as well as breadth. The fact remains, however, that marketing researchers of the future will have to seek grounding in multivariate methodology, if current research trends are any indication. This grounding will probably embrace three facets: (a) theoretical understanding of the techniques; (b) knowledge of the details of appropriate computer algorithms for implementing the techniques; and (c) a grasp of the characteristics of substantive problems in marketing that are relevant for each of the methods.

REFERENCES

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster analysis. Beverly Hills, CA: Sage.

Ball, G. H., & Hall, D. J. (1964, July). Background information on clustering techniques. (Working paper.) Menlo Park, CA: Stanford Research Institute.

Bijmolt, T. H. A., & Wedel, M. (1995). The effects of alternative methods of collecting similarity data for multidimensional scaling. International Journal of Research in Marketing, 12, 363–371.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In M. R. Rosenzweig & L. W. Porter (Eds.), Annual review of psychology. Palo Alto, CA: Annual Reviews.

Carroll, J. D., & Arabie, P. (1998). Multidimensional scaling. In M. H. Birnbaum (Ed.), Handbook of perception and cognition. Volume 3: Measurement, judgment and decision-making

Carroll, J. D., Green, P. E., & Schaffer, C. M. (1986). Interpoint distance comparisons in correspondence analysis. Journal of Marketing Research, 23, 271–280.

Clausen, S. E. (1998). Applied correspondence analysis: An introduction. Thousand Oaks, CA: Sage.

Coombs, C. H. (1964). A theory of data. New York: Wiley.

Green, R. T., & Larsen, T. L. (1985, May). Export markets and economic change. (Working paper 84/85-5-2). Austin: Department of Marketing Administration, University of Texas at Austin.

Green, P. E., Carmone, F. J. and Smith , S. M. (1989). Multidimensional Scaling: Concepts and Applications, Boston: Allyn and Bacon.

Greenacre, M. J. (1993). Correspondence analysis in practice. London: Academic Press.

Hoffman, D. L., & Franke, G. R. (1986, August). Correspondence analysis: Graphical representation of categorical data in marketing research. Journal of Marketing Research, 23, 213–217.

Johnson, S. C. (1967, September). Hierarchical clustering schemes. Psychometrika, 32, 241–154.

Kim, J.-O., & Mueller, C. W. (1978b). Factor analysis: Statistical methods and practical issues. Beverly Hills, CA: Sage.

Kim, J.-O., & Mueller, C. W. (1978b). Factor analysis: Statistical methods and practical issues. Beverly Hills, CA: Sage.

Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling. Beverly Hills, CA: Sage.

PC-MDS Statistical Software, Scott M. Smith, Brigham Young University.

Sneath, P. H. A., & Sokal, R. R. (1973). Numerical taxonomy. San Francisco: W.H. Freeman.

Sokal, R. R., & Sneath, P. H. A. (1963). Principles of numerical taxonomy. San Francisco: W.H. Freeman.

Srinivasan, V. (1997, May). Surprising robustness of the self-explicated approach to customer preference structure measurement. Journal of Marketing Research, 34, 286–291.

Whitlark, D., & Smith, S. (2001, Summer). Using correspondence analysis to map relationships: It's time to think strategically about positioning and image data. Marketing Research, 2, 23–27.

Whitlark, D., & Smith, S. M. (2003). How many attributes does it take to describe a brand? What is the right question when collecting associate data? Unpublished Paper, Graduate School of Management, Brigham Young University, Provo, UT.

Whitlark, D., & Smith, S. (2004). Measuring brand performance using online questionnaires: Advantages and issues with "pick any" data. (Research Paper.) Provo, UT: Graduate School of Management, Brigham Young University.

Chapter 15
PREPARING THE RESEARCH REPORT

Chapter 2 identified the research report as the culmination of the research process. Whatever is to be included in the report—the study's purpose, methodology, results, conclusions, or recommendations for management—should be presented clearly, accurately, and honestly. The key attributes of a report are completeness and conciseness. We can go further and state that, when managing a research project by *objectives*, the researcher should focus on the last step—*the research report* and the information needs of the user or client (Semon, 1998). The report should be designed and outlined in detail at the outset. Since the report needs to focus on information needs, this outline can be an invaluable aid in planning the other stages, such as the plans for analysis, measurement, and so forth.

This chapter provides information and guidelines that may be invaluable in preparing and presenting reports, both written and oral. All too often the research report is the only aspect of the project that others will ever see. Consequently, if the project is not effectively presented, everything done up to that point will be wasted. A research report must be read (or heard), and its results, conclusions, and recommendations evaluated, if the resources devoted to the project are to have been well spent.

COMMUNICATION

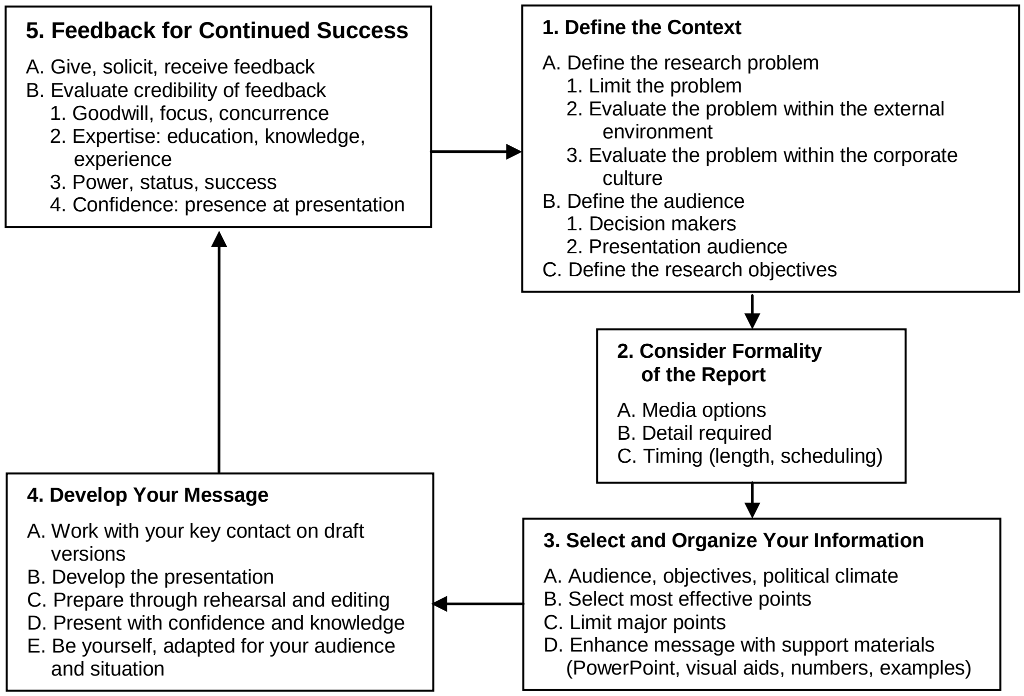
The importance of the research report derives from its purpose of communicating to others the objectives, processes and results of the research. It is important to realize that the research report and presentation may be the only measure of the quality of the research efforts. A poor quality presentation of excellent research leaves the impression of poor quality research. Effective communication has certain characteristics (Rockey, 1984; Vardaman, 1981).

First, the communication is has an *objective and a purpose*. Second, it is an *interchange of symbols*—verbal and graphic representations of a sender's ideas—between people. Third, it seeks workable *understanding and agreement* between the parties. This means that the sender of a report (the researcher) and the receiver (the client and stakeholders) have enough common understanding about, and willingness to accept or at least evaluate, the ideas presented in the report to use them in whatever manner is best. Finally, there is implicit recognition of the *need for feedback*. In order to realize satisfaction, there needs to be an opportunity to respond to a draft report. In short, communication is like playing catch with a baseball: it's not how many times the ball is thrown, but how many times it is caught. To be effective, communication must ultimately be two-way.

One view of the communication process involved in presenting a research report to management or an outside client is illustrated in Figure 15.1, where five basic steps of the communication process are identified to include context, formality, organization, message, and feedback.

Following this process will help you to communicate effectively and avoid a minefield of obstacles like multiple objectives, political agendas, and special interests. Further, reports are presented to groups with unique dynamics and individuals with personalities that guide, direct, and even dominate the project, process, or presentation. Each of these factors must be considered when determining how to best communicate the results of your research. Without considering these factors, your research report's effectiveness will be limited.

Figure 15.1 Preparing to Communicate the Research Process



SOURCE: From Bienvenu, S., & Timm, P. R., *Business Communications: Discovering Strategy, Developing Skills, 1st edition,* copyright © 2002. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.

THE RESEARCH REPORT

Written and oral reports can come in all sizes, shapes, and lengths. The nature of the problem studied, how it was studied, and the client's identity will influence the type of report to be prepared. The length of a report on qualitative research, for example, which can run from as short as eight pages to more than 100 pages, will be influenced by learning garnered, client requirements, writing style, and formatting (Greenberg, 1999). Also, it is important to remember that for many projects two reports are prepared. There is always need for a report to management. This so-called *popular* report minimizes technical details and emphasizes simplicity. Frequently, the simpler reports written in the marketer's language are most understood, read, and, thus, acted upon. However, there also may be need for a more detailed *technical* report. Such a report emphasizes the methods used and underlying assumptions, and presents findings in a detailed manner. This type of report can be a valuable resource for researchers, as well as managers, for future studies.

There are two key dimensions to the report-writing process: (a) analysis of the data; and (b) the actual writing. Analysis is the interpretation of the data obtained to provide useful and meaningful insights into the problem at hand. Writing takes the learning and expresses it in a coherent, logical, and succinct way (Greenberg, 1999). For both analysis and writing there are many ways in which data can be presented visually. According to Jacoby (1997, pp. 2–4), graphical statistical methods can be used to:

- Explore the contents of a data set
- Find structure in data
- Check assumptions in statistical models
- Communicate the results of an analysis

The requirements to both effectively analyze data and write a report may be beyond the abilities of a one person. Thus, someone other than the analyst often does the report writing.

Criteria for a Good Report

Before turning to what should be included in the report and what the format for presentation should be, we turn to a brief discussion of the criteria for a good report. The most basic criterion is how well the report communicates with the reader. Was the communication effective, in the way defined previously? Practitioners believe that reports written with clarity, brevity, and concreteness get action and commitment. The major concerns are that a research report be *complete*, *accurate*, *concise*, and *clear* in what is being said. All these are a reflection of writing style. These topics are discussed in more detail in Rehart (1994), Fink (2003), and Bailey (1990).

Completeness

A report is complete when it gives all the needed information in a language understood by the intended audience. The writing must be done on the level of the reader's understanding; this is fundamental! A report may be incomplete because it is too long *or* too short. Not all information obtained is significant or relevant enough to be reported. In the end, completeness is always defined by the person(s) who will be reading the report and asked to act upon it.

Accuracy

Obviously, accuracy is related to completeness. But it goes further. A report may not be completely accurate because the data (or the information in general) upon which it is based are not accurate. There may have been flaws in research design, measurement and scaling, sampling, and analysis that led to inaccurate results being presented. In addition, carelessness in handling data, interpretation of analyses, or writing style may also lead to inaccuracy. The preparation of a research report calls for attention to detail, including attention to the meaning of every word used, punctuation, and so forth.

Conciseness

Being concise means being selective. Thus, as we said above, not all information obtained needs to be reported. When writing a report, one should write not only so the reader can understand, but can also do so as quickly and easily as possible. Conciseness refers to what is included and how it is included. Sentences should be kept as short as possible. A concise report is not the same as a brief report. A brief report usually contains only the highlights or base essentials of what the researcher has to report. In contrast, a concise report may contain any amount of detail and be very long. A concise report may be very complete. But it is efficient; that is, it conveys all the researcher wants to present in the shortest, most direct way.

Clarity

Clarity may be the most misunderstood criterion for evaluating a research report. Clarity derives from clear and logical thinking and organization. Clarity involves connections between words, sentences, paragraphs, topics, or ideas. There is always some logical connection between, say, a topic or idea and the one that preceded it. For example, there are different structural ways to organize material in a logical framework, as illustrated in the list below:

1. **Time order.** Ideas are presented in chronological time, either forward or backward.
2. **Space order.** The relationship among places and locations determines organization.
3. **Cause and effect.** Cause may precede effect or the opposite may hold. A report may stress one or the other.
4. **Increasing difficulty.** Structure involves going from the simple to the complex or from the familiar to the unfamiliar, useful when the audience lacks expertise in the topic.
5. **Established category.** When a basic framework is understood, content can be organized in recognized categories. In a market segmentation report, using accepted categories for any of the demographic bases is useful.
6. **Comparison or contrast.** Readers grasp differences or similarities more easily when this structural organization is used.
7. **Pro and con order.** Presenting arguments for and against something, usually without favoring one or the other is another common structure.

A most important rule is that the report be well organized. At the same time, it must be written clearly. This means that researchers who prepare reports should develop a writing style conducive to clarity in presentation. Many different sets of rules for clear writing are available. Most say the same things, but perhaps in different ways. One such set is shown in Exhibit 15.1.

EXHIBIT 15.1 Some Rules for Developing a Writing Style

1. **Use concrete words.** Such words are clear and specific. Be willing to specify, itemize, give details or examples, define, and illustrate.
2. **Keep sentences short.** Shorter sentences usually are more readable, all other things being equal. Lengthy sentences are more likely to cause confusion or complexity.
3. **Vary sentence types and structures.** Use declarative (assertion), interrogative (question), imperative (command), and exclamatory sentences. Do not limit a report, but use all types where appropriate. Also, vary structure among simple, compound, complex, and compound-complex sentences. Varying sentence types and structure increases interest and reduces monotony.
4. **Maintain unity.** Try to build each paragraph around one idea or topic.
5. **Use active verbs.** Active sentences tend to be more forceful and more direct. This does not mean that the passive voice should never be used. Use it sparingly, however.
6. **Avoid wordiness.** Say what is to be said in as few words as possible.
7. **Use varying means of emphasis.** There are different ways to emphasize certain points made in the report: space, repetition, position, and mechanical means (e.g., arrows, color, or underscoring).
8. **Write and speak naturally.** Often it is difficult to identify what is natural. A helpful guide is to use conversational language, but this is only a guide.
9. **Write on the level of reader's understanding.** Do not overestimate the reader's understanding and confuse him or her with highly specialized technical jargon. Also, do not underestimate the reader by using overly simplistic and childish terms.
10. **Watch the pace.** Avoid trying to say too much in too few words. In the same light, it is bad to stretch out an idea by using too many words.

11. Keep the tone appropriate. The way in which words are put together says a great deal about the writer. Thus, tone of writing implies something about the personality of the writer and of the writer's organization or company. While it is rare for a person to have complete control of the tone of his or her writing, it is important to acknowledge that it has many shades: positive or negative, helpful or indifferent, courteous or impertinent, humble or arrogant, and so forth.

SOURCE: Rockey, 1984, pp. 101–106.

Practitioners' Views

Many of the ideas presented in the last few pages are widely shared by research practitioners. Four perspectives of report preparation are presented in Exhibit 15.2.

EXHIBIT 15.2 What the Research Practitioner Has to Say

Different researchers have different philosophies regarding what is important in report preparation. All agree that poorly communicated research is useless. All too often, the written report ends up as a second-class citizen to the demonstration of a new or repositioned statistical technique.

Perspective #1: Strengthen Your Communications

In order to strengthen written communication, Teresa Wrobel (1990), Vice-President of Research 100 in Princeton, New Jersey, suggests six things:

1. **Do the readers' work for them.** Presume that a reader feels uncomfortable manipulating or interpreting numbers. The text relates the major issues, while the numerical data serve as supporting information. Numbers should not interrupt the flow of the text, and the reader should not be forced to perform mental calculations in mid-sentence. Calculations should already be completed and presented in an easily to understand form.
2. **You don't have to analyze everything.** Not every attitude or behavior that has been quantified is worthy of discussion. Even statistically significant results do not always merit discussion. Select only those findings that illustrate the main issues.
3. **Write in plain English (or whatever language is being used).** Avoid industry jargon. Jargon tends to conceal the main ideas being conveyed and may be incomprehensible to the intended readers, who are often not researchers. The use of jargon can alienate or confuse the reader.
4. **Convey the main ideas quickly.** Most decision-makers don't have time to plow through pages of text to unearth the main findings. The front section of a research document should reduce the findings to their skeletal elements, and include the conclusions and recommendations supported by key findings.
5. **Did you answer the question at hand?** Conclusions should address the initial objectives with clarity and precision.
6. **Keep the report interesting.** Take advantage of the richness and breadth of language available. Keep searching for new ways of expressing points that convey ideas most effectively. While it is easiest to rely on language that is familiar and tested, expanding one's use of language can encourage idea generation and provide a more enjoyable document for the reader.
7. **Some format limitations must be followed.** In addition, using esoteric language for its own sake may lose the reader's attention. However, such limitations should not stifle creativity. Diversity in communication styles should be accepted and encouraged.

Perspective #2 Make It Readable

A slightly different view is that of Howard Gordon (1998), a principal of George R. Frerichs & Associates of Chicago. Mr. Gordon feels the following seven points can contribute to preparing more readable, usable, and action-oriented reports:

1. **Present tense works better**. Use present tense. Results and observations expressed in "now" terms sound better and are easier to read. Don't say: "The test panel liked the taste of the juice." Say: "People like the taste of the juice." We do studies for clients to help them make decisions today and tomorrow—not yesterday. Clients want to know what people think now.
2. **Use active voice**. Use active voice where possible, which is in most cases. There is nothing wrong with saying: "We believe . . ." rather than: "It is believed that . . . ." Passive voice is stilted. Use first person plural, not third person singular. Present tense and active voice make reports sound action oriented and businesslike. And active is easy to read. Passive voice makes the reader work harder.
3. **Findings—don't use the word findings**. This word makes your report sound as if some archaeologist just came across some old bones from the Paleolithic age. Marketers should use results, conclusions, or observations—not findings. Conclusions is the take-away section. What is it you want the reader to take away from your report? If somebody could spend only five minutes with your report, what should that person know?
4. **Use informative headlines, not label headlines**. Your reader will welcome this headline: "Convenience is the packaging's major benefit." Your reader will not generally welcome this headline: "Analysis of packaging." Nor this: "How people feel about packaging."
5. **Let your tables and charts work**. Help them with words when needed. Don't take a paragraph or a page to describe what a table already says. If there is nothing to say about a table or a chart, don't say it. The purpose of a table or a chart is to simplify. Use your words to point out significant items in the table—something that is not readily clear. Or use your words to offer interpretive comments. Use verbatims from sample respondents to support a point.
6. **Use the double-sided presentation whenever possible**. This format will reduce the verbiage in your report. It simply presents the table on the left side of the open report. Your informative headline and interpretive comments are on the right-hand page. The double-sided presentation is one of the most readable report formats. And it is much easier to write a report that uses it. The double-sided system also allows you to use white space effectively. White space helps readers get through your report—especially those readers in executive suites.
7. **Make liberal use of verbatims**. Great nuggets of marketing wisdom have come from people's comments. Use verbatims if you have them. Sprinkle the verbatims throughout your report in appropriate places as they make research reports interesting and readable

Perspective #3 Make Your Results Talk

Yet another approach views things differently. James Nelems (1997), President of The Marketing Workshop in Norcross, Georgia, cautioned us more than 20 years ago on what should be avoided. He believed that research reports have more faults beyond improper analysis!

1. **The longer the report, the better.** Research isn't bought or sold by the pound. Any research report—regardless of sample size, multivariate techniques, or number of questions— should be reducible to a one-page executive overview. The report itself also should be clearly written and easy to read, including marketing conclusions and implications—the indications for future action—drawn from the data. Management has both the time and desire to read such reports—and the confidence to accept.
2. **Indiscriminate use of fancy tools.** Some, particularly the multivariate ones, may confuse rather than enlighten.
3. **Letting questions do all the work.** Research is more than asking questions. How simple it is in a package test to ask if people like the package. Or if someone doesn't use a product, just ask him why he doesn't use it. Even though these are the wrong questions, they are often asked, and, of course, when questions are asked, answers are received. But wrong answers to wrong questions hardly constitute analysis.

4. **Reporting what happened, rather than what it means.** Too much research text simply repeats numbers in the accompanying table. What is the significance of the numbers? How are we better off by knowing the answers?
5. **The fallacy of "single-number research."** This refers to any kind of research, the outcome of which is a "single number" that purports to provide an easy decision, such as an on-air recall score, a positive buying intention, or the overall preference in a product test.
6. **Failure to relate findings to objectives or reality.** Low brand awareness should be improved, of course, but what is unaided brand awareness per $1,000 of advertising spent vs. competition? In a product test, preferences can vary between current users and nonusers. Yet the brand objectives—gain new users or hold current ones—can produce opposite results with the exact same preference data.
7. **Spurious accuracy.** This includes summary statistics to one decimal place. More than likely, they're only accurate within 10 percentage points. Yet we've actually seen in print a comment on "20% of the sample of five respondents . . . ."

Perspective #4 Write With Style

A final example of the thinking of practitioners is provided by Lynn Greenberg (1999, pp. 39–41), a principal of Lynn Greenberg Associates, a Scarsdale, New York, research consulting company. She suggests the following 11 tips for qualitative research reports, particularly focus groups:

1. **Capture your initial thoughts** in a research journal as you work. Some of your best insights may come at this time.
2. **Have a flexible plan**, one that initially has been developed before starting the writing.
3. **Break the task into small pieces and start with the easiest ones.** Once the basic sections have been written, it is easier to deal with the more challenging areas.
4. **Be succinct.** The reader should be able to clearly and quickly come away with two or three key points after reading the report.
5. **Provide overviews** to promote understanding of the learning. The first statement of a section should be an interpretation of what was learned, followed by more details.
6. **Turns negatives into constructive learning** and be clear about their meaning.
7. **Integrate information** from other sources where appropriate to provide more relevant meaning.
8. **Feature action-oriented conclusions** and implications that address all of the study's objectives, plus any important ancillary issues.
9. **Use a visually appealing format** that allows the readers to skim relevant issues. Use italics, boldface, underlining, bullet points, and so forth, wherever possible.
10. **Edit, edit, and re-edit** with *independent proofing* for clarity, grammar, and typos.
11. **Do it your way!** Write with your natural style.

REPORT FORMAT AND ORGANIZATION

Every report prepared for a research project is unique and custom-made. Although reports can be customized, there also is a degree of standardization to report format and organization.

Many techniques people use to organize research reports are based on *placement* of information to gain emphasis. A piece of information at the beginning—of a report itself, of a section in a report, or even in the first paragraph of a section—tends to get a disproportionate share of the reader's attention. Thus, something to be emphasized should be presented at the beginning of the message. In a similar manner, the end of a message also can be used as a point of strong emphasis, while the middle of a message is a point of minimum emphasis. The executive (or management) summary is placed at the front of the report and the conclusions/recommendations at the end of the report.

Parts of a Formal Report

A formal report consists of a number of components that can be organized into three major parts. Exhibit 15.3 shows all the components that would be included in a complete formal report. Research reports, however, vary in formality; not all components will always be included. Figure 15.2 illustrates how the components are affected by a decreased need for formality. Typically, marketing research project reports would follow formats II or III. There are many instances where another version of format IV may be used, in which an executive summary will be substituted for the table of contents.

Figure 15.2 Components of a Report and Formality

I	II	III	IV	V	VI
Title Fly Page Title Page Letter of Transmittal Letter of Authorization Table of Contents Summary Report Body Appendix	Title Page Letter of Transmittal Table of Contents Summary Report Body Appendix	Title Page Table of Contents Summary Report Body Appendix	Title Page Executive Summary Report Body Appendix	Title Page Report Body	Report Body

DIMINISHING NEED FOR FORMALITY →

SOURCE: Robinson, 1969, p. 295.

Prefatory Pages

Every research report must have a title page, and may have a title fly page preceding it (only the title appears on a fly page). Any report longer than a couple of pages should have a table of contents. The table of contents can be as thin as having only the major sections of the report listed, or as detailed as including all headings and subheadings. Lists of tables, charts, figures, and so forth, will be included.

Some research reports may include letters of transmittal and authorization. A letter of transmittal is the means by which a report is released to a particular person or group. Often, such a letter contains material usually found in a preface or a foreword. If there is no formal summary included, then a synopsis of the findings should be included in the letter of transmittal. Other times, the letter of transmittal is relatively short. A letter of authorization is a letter to the researcher approving the project and specifying some of the details. Except in the most formal of marketing research projects, a letter of authorization need not be included in the report. An illustration of letters of authorization and transmittal used for a project awarded to an outside firm is given in Figure 15.3.

From a manager's (or client's) perspective, the most important component of the research report is the executive (management) summary. This summary reduces the essence of the study—the why, what, how, conclusions, and recommendations—to one or two pages. Often, this is the only part of a project report that the manager reads. Exhibit 15.4 shows examples of management summaries.

EXHIBIT 15.3 Components of a Research Report

A formal report consists of the prefatory section, the report body, and the appended materials. A number of components can be organized and presented within these sections. This rather complete outline provides a general framework for organizing your research report.

PREFATORY PAGES

I. Title Page
- (a) Title
- (b) Author's name
- (c) Documentation numbers and project identification numbers
- (d) Classification
- (e) Circulation test
- (f) Issue date and destroy date

II. Table of Contents
- (a) Section subtitles and pages
- (b) Illustration titles and pages
- (c) Graph titles and pages
- (d) Figure titles and pages

III. Letters of Transmittal and Authorization

IV. Summary of Report
- (a) Problem definition and date
- (b) What was researched
- (c) When
- (d) Where
- (e) How and with what techniques
- (f) Major findings
- (g) Recommendations

REPORT BODY

V. Introduction
- (a) What prompted the undertaking of the project
- (b) Who prompted the project
- (c) How the problem was defined

VI. Statement of Objectives
- (a) Problem resolution
- (b) The research objective

VII. Research Methods
- (a) Research design
- (b) Data instruments (questionnaire, records, etc.)
- (c) Data collection methods
- (d) Sampling technique
- (e) Field work

VIII. Methodological Limitations
- (a) Weaknesses in research design
- (b) External events that may have influenced findings
- (c) Errors in research methods
- (d) Alternative causes for findings

IX. Analysis of Findings
- (a) Items of significance
- (b) Items of insignificance
- (c) Interpretations of findings

X. Conclusions and Recommendations
- (a) What research findings show
- (b) What actions should be taken (or not taken)

APPENDED PARTS

XI. Technical Appendix
- (a) Informational tables, graphs, illustrations
- (b) Technical discussion of research methodology and sample
- (c) Sample validation (if relevant)

XII. General Appendix
- (a) Selected portions of preliminary interviews
- (b) Project diary: dates, places, names, events
- (c) Copies of forms, questionnaires, records, and data instruments

XIII. Acknowledgments
- (a) Names, titles, and affiliations of contributors
- (b) Contribution of each contributor

XIV. References, Bibliography
- (a) Names, titles, and source-related research

Figure 15.3 Illustrations of Letters of Authorization and Transmittal

A. Letter of Authorization

Lane County / Private Industry Council

May 13, 2010

JVJ Research Associates
554 Pinto Way
Eugene, OR 97401

Dear Sirs:
Thank you for your response to our recent solicitation for proposals to conduct public opinion research for the Lane County Private Industry Council and Department of Employment & Training. I am pleased to inform you that your firm has been chosen as the contractor for this study, pending contract negotiation and discussion of certain aspects of your proposal.
Per recent discussions with Dan Cudaback of my staff, we will be meeting with XXX on Monday, May 18, at 4:00 p.m. in our offices to discuss your proposal further. Please call Mr. Cudaback with any questions between now and then.

I am looking forward to working with you.

Sincerely,
Steven J. Ickes, Director
EMPLOYMENT & TRAINING DEPARTMENT

B. Letter of Transmittal

JVJ Research Associates
554 Pinto Way
Eugene, Oregon 97401

August 21, 2010

Mr. Steven J. Ickes, Director
Lane County Department of Employment and Training
Eugene, Oregon 97401

Dear Mr. Ickes:

This letter accompanies our final report for the Lane County Public Opinion Study.

The findings contained in this report provide an insightful view of the opinions and desires of the residents of Lane County. We trust that this report will provide meaningful direction to the community as a whole and policy makers within Lane County.

Very truly yours,
JVJ Research Associates
John Researcher
Senior Research Associate

Body of the Report

This is the main component of the marketing research project report from the point of view of size or bulk. Included in the body is the introduction, which deals with why the project was undertaken and just what the problem was. Also included are research methods used, their strengths and limitations, and all definitions, results, and analyses. The main body of the report usually ends with the conclusions and any recommendations to be made.

Appended Materials

Several types of materials may be appended to the marketing research report. These may include detailed tables that provide more depth than the data presented in the body of the report, and all forms and measurement instruments used in data collection. If appropriate to the project, a bibliography and list of relevant references should be added to the end of the report.

EXHIBIT 15.4 Project Summaries Come in All Sizes and Shapes

The executive summary is an abstract that should be the last section of the report prepared. Typically, it will be a one- or two-page exposition. Example A is a one-page report summary of a study done for a pipe manufacturer that examined market potential for epoxy-coated pipe. Example B is drawn from a study of market needs and potential for a local motel. Example C is a two-page summary of a study done for a supplier of janitorial supplies.

A. Epoxy-Coated Pipe Study

The focus of this report is the market potential for epoxy-coated pipe. A survey of three separate samples was conducted to explore the market penetration potential of the product. These samples were: irrigators, municipalities, and industries. On the basis of data collected from the irrigator sample, it was concluded that the demand for pipe in 2010 would not be greater than that purchased in 2009. Respondents would not be willing to pay more for ECP than for PVC. There is some indication that a higher price would be paid for ECP relative to CTP. The irrigation market appears to be tightly controlled by PVC, decreasing the opportunity for market penetration by ECP.

On the basis of data collected from the municipality sample, it was concluded that the demand for pipe in 2010 would decrease from levels demanded in 2009. This expectation is moderated by the fact that demand in this market fluctuates yearly. The respondents would not be willing to pay more for ECP than for other types of pipe. The municipality market also appears to be tightly controlled by PVC, decreasing the opportunity for market penetration by ECP.

On the basis of data collected from the industry sample, it was concluded that demand for pipe in 2010 would increase relative to demand in 2009. Respondents would be willing to pay more for ECP than for other types of pipe. The industrial market is not tightly controlled by PVC. While other types of pipe dominate, however, ECP does not appear to be perceived as a viable substitute for types of pipe already in use.

Research was conducted on the future price levels of steel and oil to enable a comparison between plastic pipe (PVC) and steel pipe (ECP). Steel prices are extremely difficult to forecast. In general, they are expected to increase. Forecasts for oil prices (in order to predict prices) are also expected to show increase over time.

In general, it is recommended that ECP not be introduced into any of the markets at this time.

B. Motel

The focus of this report is the potential for increase in the occupancy at the Duck Inn Motel. To determine this, three different potential target markets were defined and examined. These markets include bus tourism, university departments, and general tourism. Three separate methodological tools were used and were assigned to the target markets as follows:

TARGET MARKET	METHODOLOGICAL TOOL
BUS TOURISM	FOCUS GROUP/MAIL SURVEY
UNIVERSITY DEPARTMENTS	PERSONAL INTERVIEWS
GENERAL TOURISM.	MAIL SURVEY

On the basis of data collected for the bus tourism market group, the most important thing to this segment is the attitude of the motel staff towards those traveling in a large group.

The results gathered for the university departments group indicate that the Duck Inn has many qualities attractive to campus departments, but the motel needs to increase its exposure to increase potential customer awareness.

And finally, for the general tourism target market, the target market is generally happy with the Duck Inn on most of the important choice criteria, but improvements can be made in advertising, appearance, and cohesion with the on-site restaurant to increase occupancy rates further.

C. Janitorial Supplies

Objectives

This study attempts to achieve several objectives.
There are two primary concerns the research design intends to determine:

1. The need for a distribution center in Eugene
2. The need for a product showroom, if a distribution outlet is established

Four support questions are also included:

1. Current market share of leading competitors
2. Degree of satisfaction with current suppliers
3. Degree of loyalty to current suppliers
4. Customer base and trends, including demand for Saturday access to supplies and the potential for telemarketing

Findings

The study shows that there is a substantial lack of confidence in the current suppliers operating in the Eugene/Springfield metropolitan area. Common complaints include poor credibility, insufficient product knowledge, and high-pressure sales tactics. Paulsen & Roles Laboratories (P&R) exhibited none of these traits. Quite to the contrary, genuine compliments for P&R's business tactics flowed freely from respondents associated with P&R.

Conclusions

The Eugene/Springfield metropolitan area is the second-largest statistical metropolitan area in Oregon. For this reason alone, it is critically important that Paulsen & Roles actively participate in this market. The foundation is in place, as exhibited by both current market share and current consumer awareness of Paulsen & Roles in the Eugene/Springfield area.

In order to further penetrate this market, it is necessary to increase local representation. P&R's current lack of sufficient local representation was the only significant complaint lodged against Paulsen & Roles during the entire interviewing process.

Recommendations

Survey respondents offered recommendations to Paulsen & Roles Laboratories:

- Aggressively pursue service as your primary competitive advantage. Develop a "full service" image.
- Continue with competitive pricing. Fully pursue the advantages of P&R's current level of vertical integration (i.e., both the manufacturer and distributor).
- Establish a distribution center in Eugene with a showroom, thus enabling a more complete utilization of P&R's competitive advantage of complete service to the consumer.
- Maintain Saturday business hours.
- Develop a seminar program for end-users in the Eugene/Springfield metropolitan area.
- Incorporate telemarketing techniques enabling (a) a reduction in the cost per sales call, and (b) expansion of coverage per sales representative.
- Establish a direct-mail campaign aimed primarily toward the janitorial contractor segment. The intent will be to minimize the cost associated with informing this market segment of Paulsen & Roles's new services in the Eugene area.
- Attempt to acquire the Johnson Wax franchise from Scot Supply.

USE OF GRAPHICS

Graphical tools are used to show the structure of data. They are relevant for both data analysis and the communication of data of results. When used properly, graphic material can enhance a research report by emphasizing the important points and more clearly showing complex relationships. However, when graphical material is used improperly or is of poor quality, it can be distracting, misleading, incomplete, confusing and erroneous (Wainer, 1984). Many types of graphic aids can be used in a marketing research report, including tables, charts, figures, maps, diagrams (e.g., a flow diagram), and other devices. Our brief discussion will concentrate mainly on tables and charts. Statistical packages and online survey research tools such as Qualtrics.com can prepare report-ready tables and charts directly from the data file. More complete discussion is given by Jacoby (1997, 1998) and Witzling and Greenstreet (1989).

Tables

There are many different ways to show numerical information in tabular form. Creating a good table of numerical data is an art as well as a science, and it goes far beyond simply putting data in columns, rows, or both. Using tables allows the researcher to minimize the amount of narrative material in the report. Only relatively short, perhaps even only summary, tables should be included in the report's body. More detailed and comprehensive tables should be placed in an appendix.

Each table should include a number of items:

1. **Table number**
2. **Title.** This should indicate by itself the contents of the table.
3. **Bannerhead and stubhead.** These are the headings for the columns and rows.
4. **Footnotes.** Explanations for particular items are placed here.
5. **Source note.** This acknowledges the source(s) of material used.

Figure 15.4 is taken from a study done for a chamber of commerce to examine involvement with business firms and is useful to illustrate the various parts of a table.

Not only are there different ways to show numerical data in a table, there also are different ways to present these data. Table 15.1 shows a typical way to present survey response data. These data were generated in a study for a manufacturer of business forms. Multiple responses from a respondent were possible. It is not always necessary to show the specific question asked in a table. If the title cannot be short and descriptive as well, then it may be necessary to show the question asked.

Table 15.1 Presenting Survey Response Data

"What role do you play in the purchase of business forms?"

	Role played				
Market segments	*Specify*	*Recommend*	*Purchase*	*None*	*Total sample size*
Hospital	25%	44%	59%	17%	259
Finance	29%	32%	62%	11%	164
Manufacturing	18%	27%	81%	3%	205
Retail	10%	0%	52%	37%	52
Government	39%	48%	4%	8%	213
Transportation	28%	28%	74%	6%	109

Figure 15.4 Table Example

Table Number → Table 5

Title → Extent of Agreement with Attitude Statement about the Chamber of Commerce[a]

Attitude Statement	Member			Non-member			General Public		
	Percent Agree	Percent Disagree	Mean Value[c]	Percent Agree	Percent Disagree	Mean Value[c]	Percent Agree	Percent Disagree	Mean Value[c]
Offers business contacts	58.3	10.8	2.5	23.8	57.1	3.5	(b)	(b)	(b)
Enhance sales	41.2	23.6	2.8	10.0	50.0	3.6	(b)	(b)	(b)
Achievement is noticeable	42.5	23.0	2.9	34.8	30.4	3.0	62.9	14.3	2.5
Costs too much to be a member	32.6	39.3	3.1	83.3	5.6	1.9	(b)	(b)	(b)
Unwilling to aid small businessman	17.3	55.6	3.5	22.7	27.2	3.0	(b)	(b)	(b)
Benefits community	75.6	8.9	2.2	60.8	13.0	2.5	65.1	9.3	2.3
Is socially attractive to be a member	29.5	31.8	3.0	9.1	45.4	3.4	(b)	(b)	(b)
Membership is prestigious	12.6	35.6	3.3	12.5	50.0	3.4	40.5	8.1	2.6
Enhance image of member firm	24.4	26.8	3.1	13.0	52.1	3.4	48.8	12.8	2.6

Stubhead → (Attitude Statement column); Boxhead ← (column headings)

Footnotes → [a]Responses were either "strongly agree," "agree," "neutral," "disagree," or "strongly disagree."
[b]Not available.
[c]Scoring ranges from 1 for "strongly agree" to 5 for "strongly disagree." Thus, the lower the mean value the greater the tendency to agree with the statement.

When using statistical analysis software, a researcher may receive report-ready tables. This is quite common for the many specialized cross-tabulation packages, and even in online packages like Qualtrics where, for example, one dependent variable could allow for simultaneous cross-tabulations and drill downs of the data for multiple demographic variables. Such a table is known as a *stub-and-banner* table.

Exhibit 15.5 Principles of Graph Construction

Clear Vision

- Make the data stand out. Avoid superfluity.
- Use visually prominent graphical elements to show the data.
- Use a pair of scale lines for each variable. Make the data region the interior of the rectangle formed by the scale lines. Put tick marks outside the data region.
- Do not overdo the number of tick marks.
- Use a reference line when an important value must be seen across the entire graph, but do not let the line interfere with the data.
- Do not allow data labels in the data region to interfere with the quantitative data or to clutter the graph.
- Avoid putting notes, keys, and markers in the data region. Put keys and markers just outside the data region and put notes in the legend or in the text.
- Overlapping plotting symbols must be visually distinguishable.
- Superposed data sets must be readily visually discriminated.
- Visual clarity must be preserved under reduction and reproduction.

Clear Understanding

- Put major conclusions into graphical form. Make legends comprehensive and informative.
- Error bars should be clearly explained.
- When logarithms of a variable are graphed, the scale label should correspond to the tick mark labels.
- Proofread graphs.
- Strive for clarity.

Scales

- Choose the range of the tick marks to include or nearly include the range of data.
- Subject to the constraints that scales have, choose the scales so that the data fill up as much of the data region as possible.

- It is sometimes helpful to use the pair of scale lines for a variable to show two different scales.
- Choose appropriate scales when graphs are compared.
- Do not insist that zero always be included on a scale showing magnitude.
- Use a logarithmic scale when it is important to understand the percent change or multiplicative factors.
- Showing data on a logarithmic scale can improve resolution.
- Use a scale break only when necessary. If a break cannot be avoided, use a full scale break. Do not connect numerical value on two sides of a break.

General Strategy

- A large amount of quantitative information can be packed into a small region.
- Graphing data should be an iterative, experimental process.
- Graph data two or more times when it is needed.
- Many useful graphs require careful, detailed study.

SOURCE: Cleveland, 1985, pp. 100–101. Reprinted with kind permission of Kleuwér Publications.

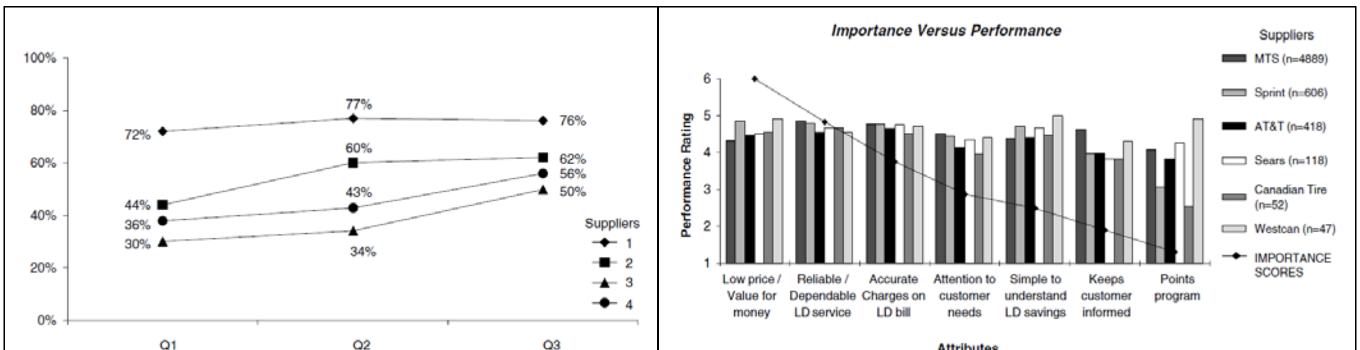
Charts and Graphs

Charts and graphs are a popular way to present numerical data. In one way or another, the information contained in a table can be presented with the help of a chart. The overall advantages are that relationships and structure can be more easily seen and the narrative can be minimized. Indeed, charts and graphs are often constructed with the premise "a picture is worth a thousand words," and as such make results more easily understandable (Gutsche, 2001). There are many different kinds of charts. These include line charts, bar charts, pie charts, and dot charts. At times, some combinations may be used in the form of dashboards that provide a real-time overview of a market as the data is continually updated.

Line Chart

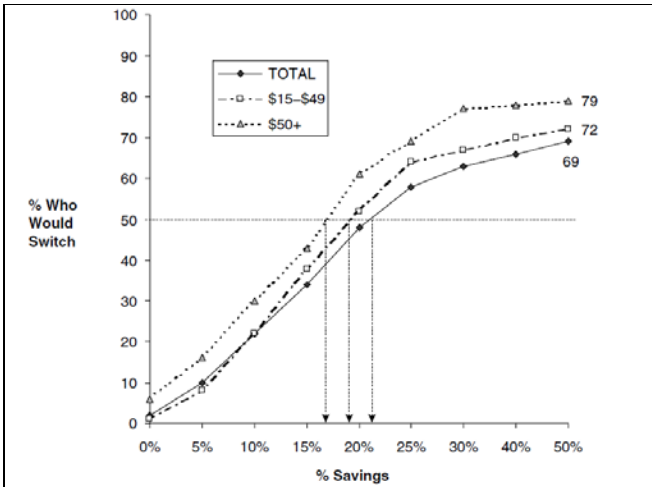
Also known as a line graph, the *line chart* is useful to show the relationship between variables. A *simple line chart* relates two variables, a *multiple line chart* shows the relationship between the independent variable and more than one dependent variable, and a *stratum line chart* (or *stacked line chart*) consists of a set of line charts whose quantities are grouped together. Figures 15.5a–d illustrates horizontal and vertical line charts using categorical and ratio scales, as well as in combination with vertical and horizontal bar charts. It is not necessary that the independent variable be continuous or ordered (The Faneuil Group, Inc., 1999).

Figure 15.5a-d Example Horizontal Line Charts

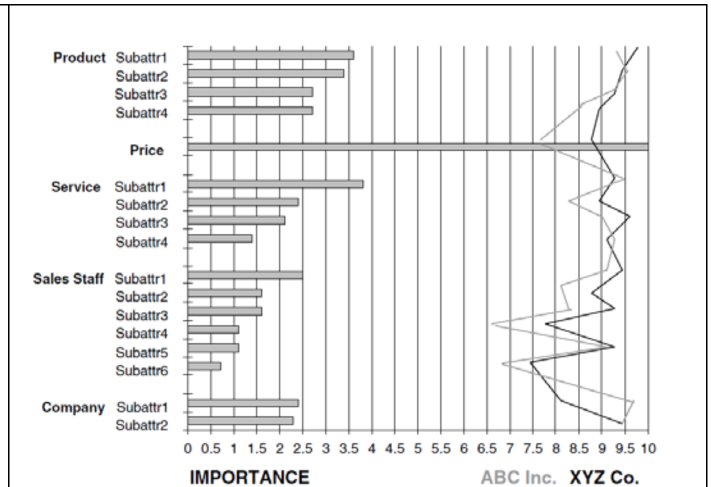


(a) Multiple Horizontal Line Chart: Percentage Change Over Time

(c) Single Horizontal Line Chart: Attributes x Performance Rating



(b) Multiple Horizontal Line Chart: Switching Percent x Percent Savings

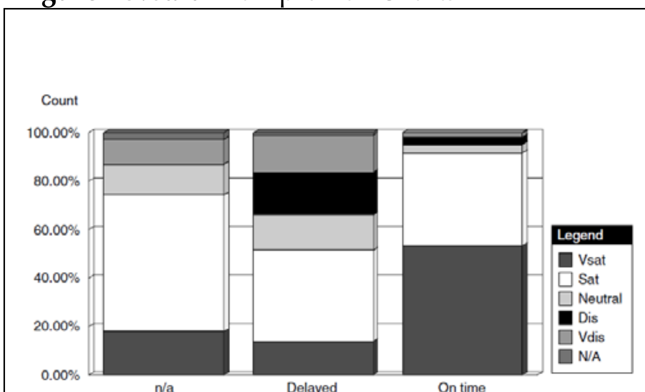


(d) Multiple Horizontal Line Chart with Horizontal Bars

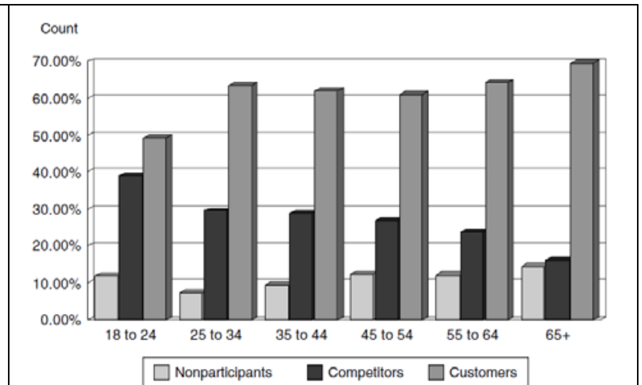
Bar Chart

A *bar chart* is another useful type of graph. Such a chart shows the relationship between a dependent variable and some independent variable at discrete levels. The stacked vertical bar chart shown in Figure 15.6a is used to show the relative satisfaction given on-time performance (*x*-axis). Figure 15.6b shows percentages associated with frequency counts of three different market segments by age categories.

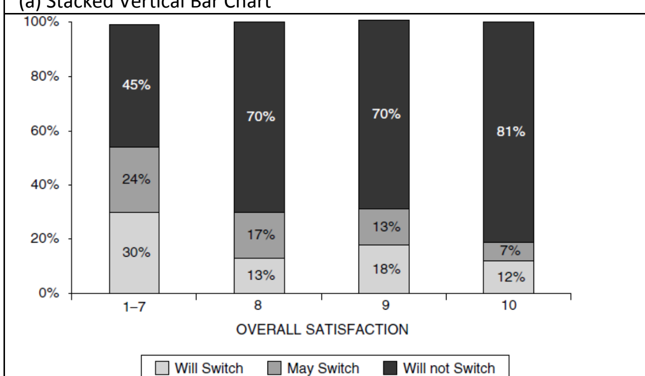
Figure 15.6a-d Example Bar Charts



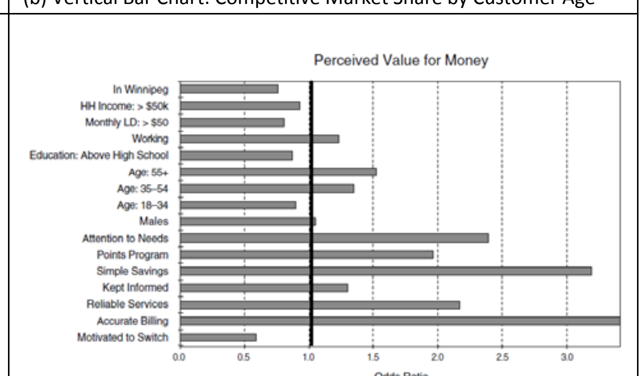
(a) Stacked Vertical Bar Chart



(b) Vertical Bar Chart: Competitive Market Share by Customer Age



(c) Stacked Vertical Bar Chart: Loyalty x Satisfaction



(d) Multivariable Horizontal Bar Chart with Threshold Line

There are times when a researcher may feel the need to draw special attention to a particular result and so may add multiple variables with demarcation lines or special cables. When using this type of notation, the researcher should make sure that he or she has not added too much information that confuses the reader. See Figures 15.6c and 15.6d for more detail (The Faneuil Group, Inc., 1999).

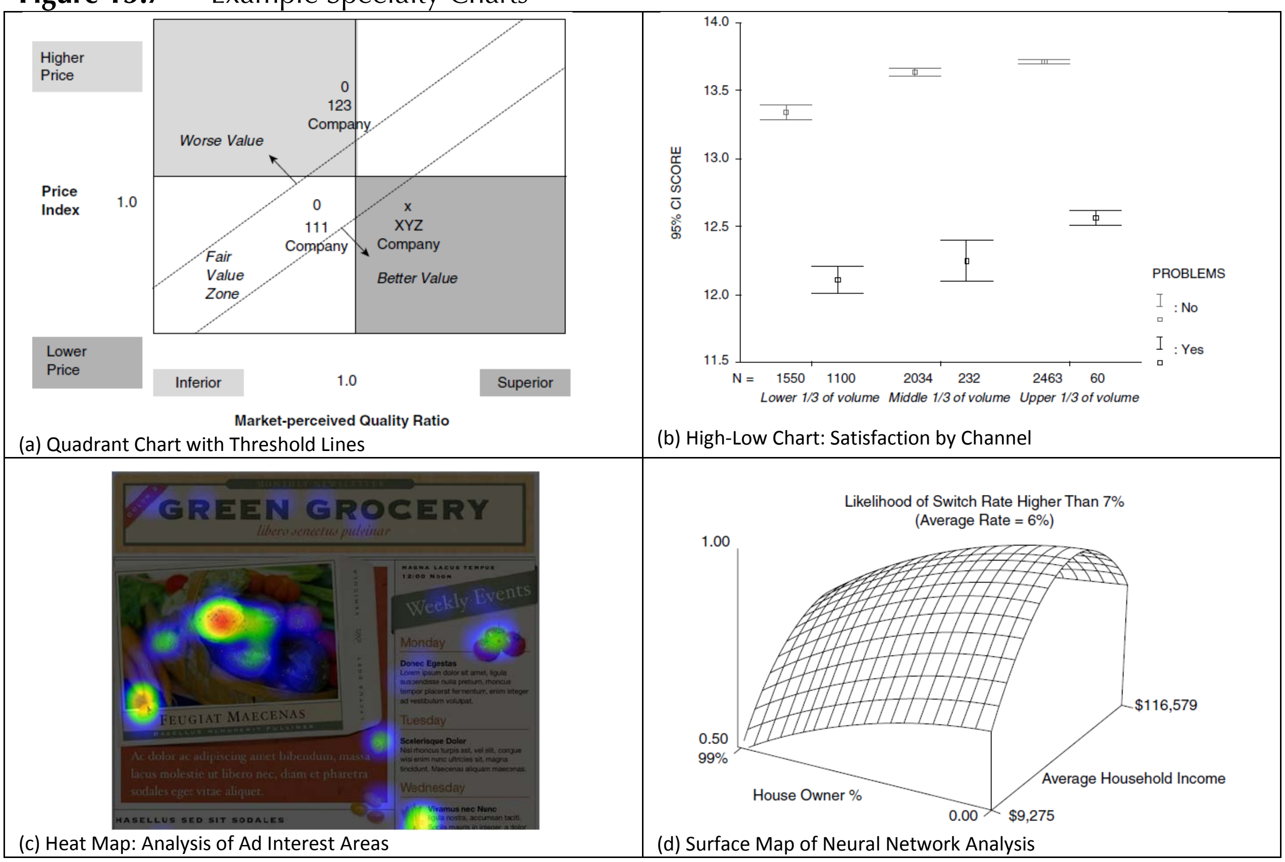
Specialty Charts

Other widely used chart formats include the quadrant, high-low, hierarchical tree, and surface charts (see Figures 15.7a-d). These charts are particularly useful to display a quantity that is subdivided into parts. When more categories are analyzed, tree diagrams may be used. Care should be exercised when too many subdivisions are shown and some portions are so small and narrow it is difficult to see them.

Dot Charts

A simple, yet powerful, type of chart is the *dot chart*. This is a way to display measurements of one or more quantitative variables in which each value may have a label associated with it. We have already shown such charts by the scatter plots (diagrams) in Chapter 13.

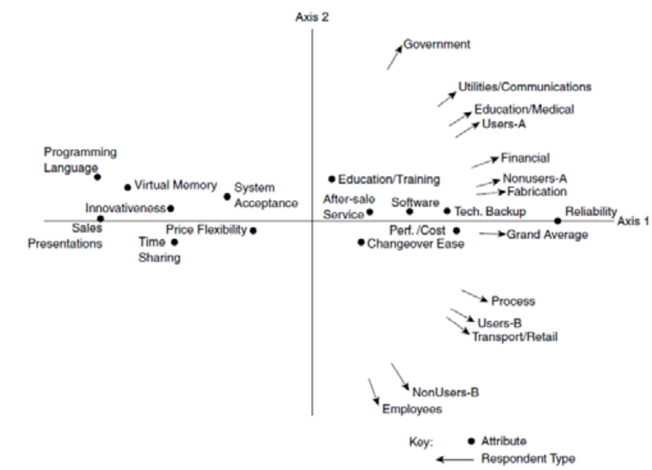
Figure 15.7 Example Specialty Charts



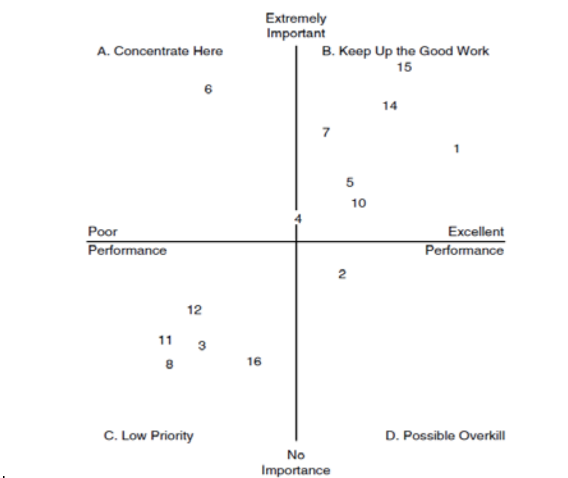
(a) Quadrant Chart with Threshold Lines
(b) High-Low Chart: Satisfaction by Channel
(c) Heat Map: Analysis of Ad Interest Areas
(d) Surface Map of Neural Network Analysis

Dot charts are used as output from multidimensional scaling (MDS) analyses (see Chapter 14). Figure 15.8a shows a MDS result for a computer manufacturer where attributes of computer manufacturers are plotted as dots. A dot chart may be used in a similar manner; Figure 15.8c presents utility functions obtained using conjoint analysis of a 3^3 fractional factorial design examining product attributes (see Chapter 14). When they appear, the solid lines (they could be dotted as well) should be made light so that they do not detract from the dots themselves. Although a literal use of this type of chart would show only dots in the body of graph, this need not be the case. Either letters or numbers also can be used, as illustrated by the Importance/Performance analysis of Figure 15.8b. The numbers are a result of plotting both mean ratings of importance and performance.

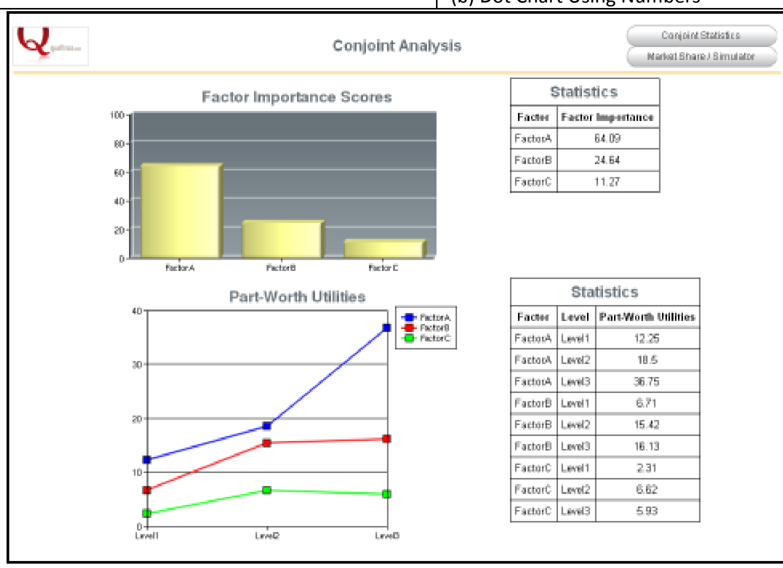
Figure 15.8a-c Example Dot Charts



(a) Dot Chart of MDS Analysis



(b) Dot Chart Using Numbers



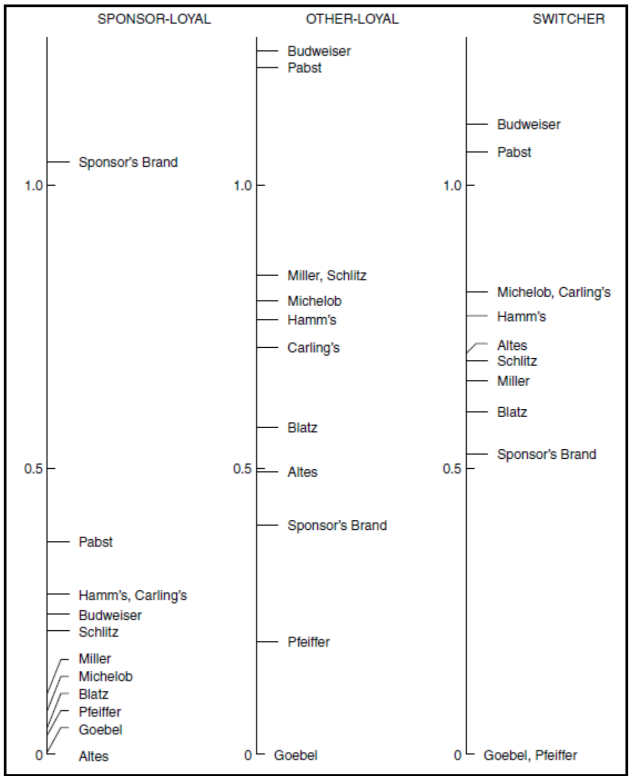
(c) Dot Chart for Conjoint-Analysis Derived Utility Functions

Another useful modification of the dot chart is one that essentially uses tick marks, each of which is labeled. A beer company was interested in knowing something about preferences for its beer and a number of competitive brands. Preference rankings for 12 brands were obtained and analyzed by Thurstonian Case V scaling (see Chapter 10). Prior to applying this procedure, each respondent was classified in one of three ways:

1. *Sponsor brand-loyal*—drinking the sponsor's brand regularly for at least a year
2. *Other brand-loyal*—drinking a particular (competitive) brand regularly for at least one year
3. *Switcher*—others not meeting either of the above conditions

A Thurstonian scale of preferences was then developed for each of the preceding segments and appears in Figure 15.9.

Figure 15.9 Example Thurstone Case V Scale



Other Graphic Aids

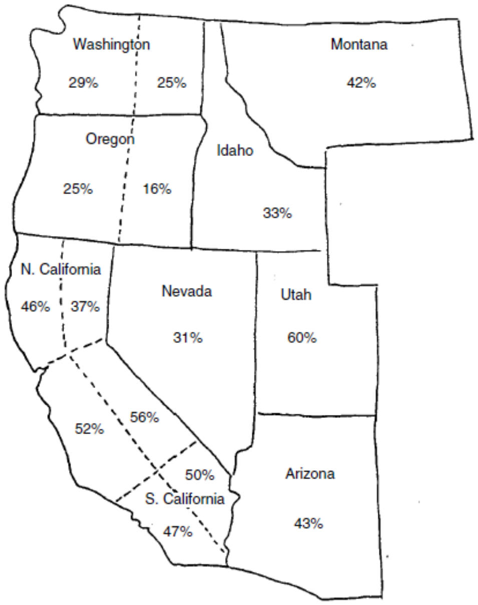
Sometimes key results from a study can be presented by use of maps. Often we see a map of the United States or another political entity used to display census data. Maps are used to enhance the meaning of data. For example, the situation facing the plastic irrigation pipe manufacturer regarding PIP in different selling areas of the country is shown in Figure 15.10a, which shows the geographic boundaries of selling areas. Figure 15.10b shows a map of the eight western states and reports the percent of respondent pipe sellers who sell PIP in each area. This type of map gives some life to the data beyond listing market area and showing percentages.

Another device that can enhance the presentation is a pictorial chart or pictogram. A *pictogram* depicts data with the help of symbols, which can be anything—stars, stacks of coins, trees, castles, facial expressions, money, caricatures of people, and so forth. Each symbol represents a specific and uniform amount or value. An example is shown in Figure 15.10c. There are situations where the size of the pictorial symbol is made proportional to the values that are to be shown, as illustrated in Figure 15.10d. One must be careful when doing this so as to not confuse the reader by distorting symbols. For example when doubling the size of a three-dimensional symbol, if each dimension is doubled the volume is increased eight times.

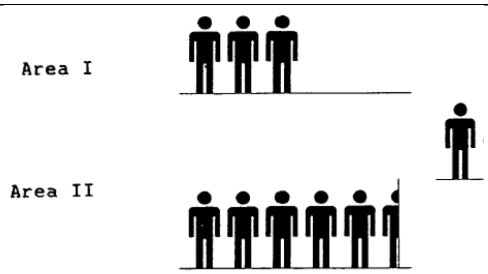
Figure 15.10a-e Use of Maps and Pictograms in Displaying Research Results

Selling Area	Percentage of All PIP Respondents Selling in the Area	Percentage of PIP Sellers With More Than 5% Sales in the Area
Area 1	78%	43%
Area 2	71%	71%
Area 3	57%	57%
Area 4	14%	0%
Area 5	57%	57%
Area 6	43%	29%

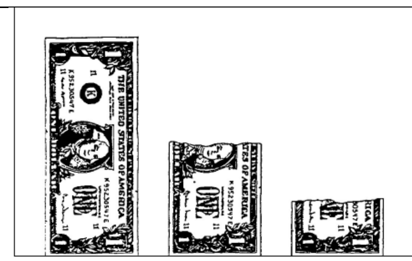
(a) Use of Maps in Plastic Irrigation Pipe Study: Selling Areas of PIP Manufacturers



(b) Use of Maps in Plastic Irrigation Pipe Study: PIP Respondent Selling Areas as a Percentage of All Respondents Reporting Sales in Area



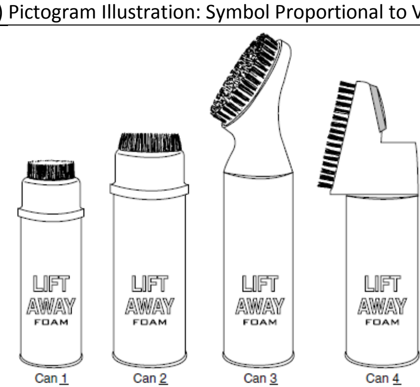
(c) Pictogram Illustration of Number of Customers by Sales Area



(d) Pictogram Illustration: Symbol Proportional to Value

					Tc
	Most Prefer	8	6	124	82
	2nd Most Prefer	12	16	75	116
Preference	3rd Most Prefer	47	152	9	12
Without Price	Least Prefer	152	46	12	10

(e) Frequencies of Preference for New Package Designs

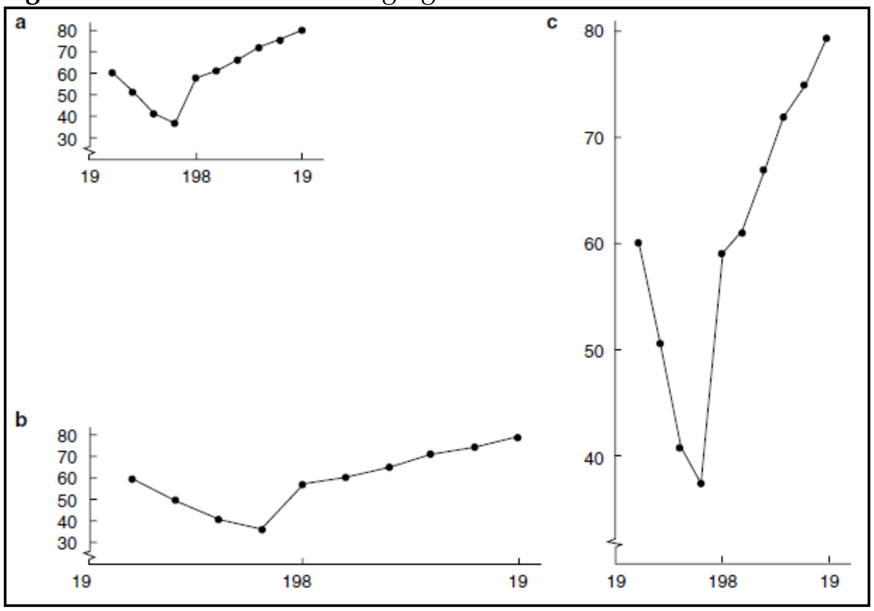


Pictures or drawings themselves may be useful to show what something looks like. A manufacturer of carpet cleaning agents developed a new type of foam cleaner called *Lift Away*, made especially for spot cleaning. A study was done to test four alternative aerosol packages, which used a sample of 220 housewives. Figure 15.10e shows sketches of the four package designs and frequencies of choice. It would have been difficult to describe the packages in words only. The sketches greatly aid understanding of the differences between the test items.

A Warning

When developing graphical material one must be very careful. It is not too difficult to get different impressions from graphs of identical data, each of which is technically correct. Figure 15.11 illustrates this point.

Figure 15.11 Effects of Changing an Axis



By stretching or compressing the horizontal or vertical axis, as the case may be, one can convey different meanings.

A researcher needs to follow sound principles when constructing graphical material. Such principles as those discussed in this chapter are focused on obtaining a clear vision, clear understanding, sound scales, and sound overall strategy for presentation.

THE ORAL REPORT AND PRESENTATION

In general, the written report will be supplemented by an oral/PowerPoint presentation. The objective behind this oral report is to identify and emphasize the major findings of the study and to allow the client or manager to ask questions about matters not clear to him or her. In some instances, however, the oral presentation may be more important than the written report. When this is the case, the written report consists of an extended management summary (3–4 pages) and copies of all overhead projector transparencies, photocopies, slides, or PowerPoint slides presented in the oral report.

When planning for the oral presentation, the researcher should consider the basic ingredients of such a presentation: (a) the *target* (purpose of the report); (b) the *receiver* of the report (the willingness and capacity of the listeners to understand and accept the report); (c) the *impact* needed; and (d) the *methods* that must be used to achieve the desired impact, given the receiver and target. The presenter must adapt to the audience, and this may require having to defend the results being presented. This is not the same as being in an adversarial position. It does mean, however, that a presenter should be prepared to deal with any and all questions in a professional, expert, and competent manner.

Each graphic needs to be of sufficient quality that it can be interpreted by the audience. When the researcher wants to use visual aids, it is important to keep in mind the difference between type of visual aid and medium of presentation. Any one of the types of visual aids (or graphics) available from graphics programs can be adapted and presented through many of the media available. In addition, actual actions and behaviors can be shown via videos, DVDs, and personal computers.

Programs such as PowerPoint exist to incorporate text, graphics, video and statistical data in the same report. Exhibit 15.6 presents some rules-of-thumb for using PowerPoint to give great presentations.

EXHIBIT 15.6 Rules-of-Thumb for Presentation Graphics

Presenting with PowerPoint: 10 dos and don'ts
Jeff Wuorio

Cherie Kerr knows how PowerPoint can be both provocative and persuasive in a business meeting. She's also aware that precisely the opposite can occur. "It can be the very best friend you have," says the Santa Ana, Calif., public relations consultant. "But you have to use it right." Many embrace the values of PowerPoint as a potent business tool, but there are others who contend that it's a drag on effective interaction — that it confuses, distorts and even strangles communication. PowerPoint must be used to best advantage. Here are 10 ways to use PowerPoint to help make your business look brilliant, not brainless.

1. Hold up your end with compelling material.
In a way, PowerPoint's ease of use may be its own worst enemy. However simple and engaging it can be to build eye catching slides and graphics, bear in mind that PowerPoint isn't autonomous. The audience has come to hear you, not merely to stare at images tossed onto a screen. Build a strong PowerPoint program, but make sure that your spoken remarks are no less compelling. "PowerPoint doesn't give presentations — PowerPoint makes slides," says Matt Thornhill, president of Audience First. "Remember that you are creating slides to support a spoken presentation."

2. Keep it simple.
The most effective PowerPoint presentations are simple — charts that are easy to understand, and graphics that reflect what the speaker is saying. Some authorities suggest no more than five words per line and three to five lines per individual slide. "Don't gum up the works with too many words and graphics," Kerr says. "Do you really need to have everything up on the screen?"

3. Minimize numbers in slides.
PowerPoint's lure is the capacity to convey ideas and support a speaker's remarks in a concise manner. That's hard to do through a haze of numbers and statistics. For the most part, most effective PowerPoint displays don't overwhelm viewers with too many figures and numbers. Instead, leave those for a later, more thorough digestion in handouts distributed at presentation's end. If you want to emphasize a statistic, consider using a graphic or image to convey the point. "For instance, when I once was talking about the prevalence of Alzheimer's patients, I used a photograph of an old woman rather than just throwing up a number on the screen," Kerr says.

4. Don't parrot PowerPoint.
One of the most prevalent and damaging habits of PowerPoint users is to simply read the visual presentation to the audience. Not only is that redundant — short of using the clicker, why are you even there? — but it makes even the most visually appealing presentation boring to the bone. PowerPoint works best with spoken remarks that augment and discuss, rather than mimic, what's on the screen. You've got to make eye contact with your audience, so never turn your back to your audience.

5. Time your remarks.
A well-orchestrated PowerPoint program brings up a new slide and gives the audience a chance to read and digest it, then follows up with remarks that broaden and amplify what's on the screen. "It's an issue of timing," Kerr says. "Never talk on top of your slides." Plan on no more than 1 slide per minute.

6. Give it a rest.
PowerPoint is most effective as a visual accompaniment to the spoken word. Don't be bashful about letting the screen go blank on occasion. Not only can that give your audience a visual break, it's also effective to focus attention on more verbally-focused give and take, such as a group discussion or question and answer session.

7. Use vibrant colors.
A striking contrast between words, graphics and the background can be very effective in conveying both a message and emotion. The standard theme colors are developed by graphic artists to be pleasing and not overbearing.

8. Import other images and graphics.
Don't limit your presentation to what PowerPoint offers. Use outside images and graphics for variety and visual appeal, including video. It can add humor, convey a message and get the audience on board.

9. Distribute handouts at the end — not during the presentation.
Some people may disagree with me here. But no speaker wants to be chatting to a crowd that's busy reading a summation of her remarks. Unless it is imperative that people follow a handout while you're presenting, wait until you're done to distribute them.

10. Edit ruthlessly before presenting.
Never lose the perspective of the audience. Once you're finished drafting your PowerPoint slides, assume you're just one of the folks listening to your remarks as you review them. If something is unappealing, distracting or confusing, edit ruthlessly. Chances are good your overall presentation will be the better for it. A great presentation often takes several weeks to prepare.

Source: Jeff Wuorio, Presenting with PowerPoint: 10 dos and don'ts http://www.microsoft.com/smallbusiness/resources/technology/business-software/presenting-with-powerpoint-10-dos-and-donts.aspx#Powerpointtips

SUMMARY

This chapter has covered preparing the research report. In order to plan properly for the report, the researcher must understand some basic concepts of communications.

We discussed the components of a formal research report in some depth. Reports need to be complete, accurate, concise, and clear in presentation. There is no one correct way to prepare a marketing research report, although we do discuss some general guidelines or rules for report preparation. The intended audience must be kept in mind at all times. There are many degrees of formality that may be used. Increasingly, marketing research reports seem to be written less formally, with a major emphasis being placed on the oral presentation.

We discussed in some depth the use of graphics to show the structure of data. Tables, charts, figures, maps, and other devices were presented. Numerous examples drawn from real-world marketing research projects were given to illustrate the various graphic aids.

Next we turned to the oral report. In many situations, the oral report is used to supplement a formal written report. This allows the potential user of the study results to ask questions and seek clarification of aspects that may be confusing. In other situations, the oral report is the major vehicle for the presentation of data and information. The written report then consists of an extended management summary and copies of all overhead transparencies used in the presentation.

REFERENCES

Bailey, E. P., Jr. (1990). *The plain English approach to business report writing*. New York: Oxford University Press.

Cleveland, W. S. (1985). *The elements of graphing data*. Monterey, CA: Wadsworth Advanced Books and Software.

Faneuil Group, Inc. (1999). *Training Manual*. Boston.

Fink, A. (2003). *How to report on surveys*. Thousand Oaks, CA: Sage.

Gordon, H. (1994, December), in S. J. Levy, G. R. Frerichs, & H. L. Gordon (Eds.) *The Dartnell marketing manager's handbook*. Chicago, IL: Dartnell Corporations.

Greenberg, L. (1999, December). The challenge of qualitative report writing. *Quirk's Marketing Research Review*, 36–41.

Gutsche, A. (2001, September 24). Visuals make the case. *Marketing News*, 21–23.

Jacoby, W. G. (1997). *Statistical graphics for univariate and bivariate data*. Thousand Oaks, CA: Sage.

Jacoby,W. G. (1998). *Statistical graphics for visualizing multivariate data*. Thousand Oaks, CA: Sage.

Nelems, J. H. (1979, January 12). Report results implications, not just pounds of data. *Marketing News*, *12*, 7.

Rehart, M. J. V. (1994). *Writing business research reports: A guide to scientific writing*. Los Angeles: Pyrczak Publishing.

Robinson, D. M. (1969). *Writing reports for management decisions*. Columbus, OH: Charles E. Merrill.

Rockey, E. H. (1984). *Communicating in organizations*. Lanham, MD: University Press of America, Inc.

Semon, T. T. (1998, November 23). To make research effective, start by drafting the report. *Marketing News*, *32*, 24, 17.

Timm, P. R., & Bienvenu, S. (2001). *Business communications: Discovering strategy, developing skills.* Upper Saddle, NJ: Prentice Hall.

Vardaman, G. T. (1981). *Making successful presentations*. New York: AMACOM.

Wainer, H. (1984, May). How to display data badly. *The American Statistician*, *38*, 137–147.

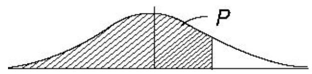
Witzling, L. P., & Greenstreet, R. C. (1989). *Presenting statistics: A manager's guide to the persuasive use of statistics*. New York: John Wiley & Sons.

Wrobel, T. K. (1990, September 3). Research is useless if it's poorly communicated. *Marketing News*, *24*, 18, 41ff.

Wuorio, J. Presenting with PowerPoint: 10 dos and don'ts. http://www.microsoft.com

APPENDIX A Statistical Tables

Table A.1 Cumulative Normal Distribution—Values of Probability



Z Values of P corresponding to Z for the normal curve. Z is the standard normal variable.
The value of P for −Z equals 1 minus the value of P for +Z, e.g., the P for –1.62 equals 1 – .9474 = .0526

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8360	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9960	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table A.2 Upper Percentiles of the *t* Distribution

1– α df	*.75*	*.90*	*.95*	*.975*	*.99*	*.995*	*.9995*
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	.816	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	1.328	1.729	2.093	2.339	2.861	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	3.551
60	.679	1.296	1.671	2.000	2.390	2.660	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	.674	1.282	1.645	1.960	2.326	2.576	3.291

df = degrees of freedom
SOURCE: From Fisher, R. A., & Yates, F., Statistical Tables for Biological, Agricultural, and Medical Research, 6/e, copyright © 1963. Reprinted with permission from Pearson Education, Ltd.

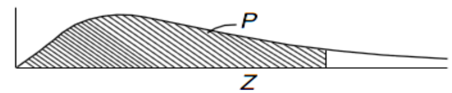
Table A.3 Percentiles of the χ^2 Distribution

df	χ2 .005	χ2 .01	χ2 .025	χ2 .05	χ2 .10	χ2 .90	χ2 .95	χ2 .975	χ2 .99	χ2 .995
1	.000039	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

For large degrees of freedom. $\chi^2 = \frac{1}{2}(Z + \sqrt{2v-1})^2$, (approximately) where *v* = degrees of freedom and *Z* is given in Table A.1.
SOURCE: Adapted from *Introduction to Statistical Analysis*, 2nd ed., by W. J. Dixon and F. J. Massey, Jr., Copyright 1957, McGraw-Hill Book Company.

Table A.4 Percentiles of the *F* Distribution

F Table for alpha=.10

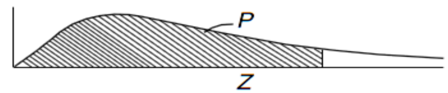


^{n}1= degrees of freedom for numerator) α= 0.10

^{n}2= degrees of freedom for denominator) α= 0.10

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.863	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.858	60.195	60.705	61.220	61.740	62.002	62.265	62.529	62.794	63.061	63.328
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392	9.408	9.425	9.441	9.450	9.458	9.466	9.475	9.483	9.491
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.216	5.200	5.184	5.176	5.168	5.160	5.151	5.143	5.134
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.896	3.870	3.844	3.831	3.817	3.804	3.790	3.775	3.761
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.268	3.238	3.207	3.191	3.174	3.157	3.140	3.123	3.105
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.905	2.871	2.836	2.818	2.800	2.781	2.762	2.742	2.722
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703	2.668	2.632	2.595	2.575	2.555	2.535	2.514	2.493	2.471
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.502	2.464	2.425	2.404	2.383	2.361	2.339	2.316	2.293
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.379	2.340	2.298	2.277	2.255	2.232	2.208	2.184	2.159
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.284	2.244	2.201	2.178	2.155	2.132	2.107	2.082	2.055
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248	2.209	2.167	2.123	2.100	2.076	2.052	2.026	2.000	1.972
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188	2.147	2.105	2.060	2.036	2.011	1.986	1.960	1.932	1.904
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138	2.097	2.053	2.007	1.983	1.958	1.931	1.904	1.876	1.846
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095	2.054	2.010	1.962	1.938	1.912	1.885	1.857	1.828	1.797
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059	2.017	1.972	1.924	1.899	1.873	1.845	1.817	1.787	1.755
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028	1.985	1.940	1.891	1.866	1.839	1.811	1.782	1.751	1.718
17	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001	1.958	1.912	1.862	1.836	1.809	1.781	1.751	1.719	1.686
18	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.977	1.933	1.887	1.837	1.810	1.783	1.754	1.723	1.691	1.657
19	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956	1.912	1.865	1.814	1.787	1.759	1.730	1.699	1.666	1.631
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937	1.892	1.845	1.794	1.767	1.738	1.708	1.677	1.643	1.607
21	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948	1.920	1.875	1.827	1.776	1.748	1.719	1.689	1.657	1.623	1.586
22	2.949	2.561	2.351	2.219	2.128	2.061	2.008	1.967	1.933	1.904	1.859	1.811	1.759	1.731	1.702	1.671	1.639	1.604	1.567
23	2.937	2.549	2.339	2.207	2.115	2.047	1.995	1.953	1.919	1.890	1.845	1.796	1.744	1.716	1.686	1.655	1.622	1.587	1.549
24	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877	1.832	1.783	1.730	1.702	1.672	1.641	1.607	1.571	1.533
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.820	1.771	1.718	1.689	1.659	1.627	1.593	1.557	1.518
26	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884	1.855	1.809	1.760	1.706	1.677	1.647	1.615	1.581	1.544	1.504
27	2.901	2.511	2.299	2.165	2.073	2.005	1.952	1.909	1.874	1.845	1.799	1.749	1.695	1.666	1.636	1.603	1.569	1.531	1.491
28	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865	1.836	1.790	1.740	1.685	1.656	1.625	1.593	1.558	1.520	1.478
29	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857	1.827	1.781	1.731	1.676	1.647	1.616	1.583	1.547	1.509	1.467
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.773	1.722	1.667	1.638	1.606	1.573	1.538	1.499	1.456
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763	1.715	1.662	1.605	1.574	1.541	1.506	1.467	1.425	1.377
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707	1.657	1.603	1.543	1.511	1.476	1.437	1.395	1.348	1.291
120	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684	1.652	1.601	1.545	1.482	1.447	1.409	1.368	1.320	1.265	1.193
∞	2.706	2.303	2.084	1.945	1.847	1.774	1.717	1.670	1.632	1.599	1.546	1.487	1.421	1.383	1.342	1.295	1.240	1.169	1.000

Table A.4 Percentiles of the *F* Distribution (continued)



F Table for alpha=.05

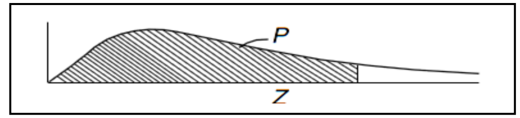
^{n}1= degrees of freedom for numerator) α= 0.05

^{n}2= degrees of freedom for denominator) α= 0.05

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	10.128	9.552	9.277	9.117	9.014	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.660	8.639	8.617	8.594	8.572	8.549	8.526
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.803	5.774	5.746	5.717	5.688	5.658	5.628
5	6.608	5.786	5.410	5.192	5.050	4.950	4.876	4.818	4.773	4.735	4.678	4.619	4.558	4.527	4.496	4.464	4.431	4.399	4.365
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.874	3.842	3.808	3.774	3.740	3.705	3.669
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.445	3.411	3.376	3.340	3.304	3.267	3.230
8	5.318	4.459	4.066	3.838	3.688	3.581	3.501	3.438	3.388	3.347	3.284	3.218	3.150	3.115	3.079	3.043	3.005	2.967	2.928
9	5.117	4.257	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.937	2.901	2.864	2.826	2.787	2.748	2.707
10	4.965	4.103	3.708	3.478	3.326	3.217	3.136	3.072	3.020	2.978	2.913	2.845	2.774	2.737	2.700	2.661	2.621	2.580	2.538
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.646	2.609	2.571	2.531	2.490	2.448	2.405
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.544	2.506	2.466	2.426	2.384	2.341	2.296
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.459	2.420	2.380	2.339	2.297	2.252	2.206
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.388	2.349	2.308	2.266	2.223	2.178	2.131
15	4.543	3.682	3.287	3.056	2.901	2.791	2.707	2.641	2.588	2.544	2.475	2.403	2.328	2.288	2.247	2.204	2.160	2.114	2.066
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.276	2.235	2.194	2.151	2.106	2.059	2.010
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.308	2.230	2.190	2.148	2.104	2.058	2.011	1.960
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.191	2.150	2.107	2.063	2.017	1.968	1.917
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.234	2.156	2.114	2.071	2.026	1.980	1.930	1.878
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.124	2.083	2.039	1.994	1.946	1.896	1.843
21	4.325	3.467	3.073	2.840	2.685	2.573	2.488	2.421	2.366	2.321	2.250	2.176	2.096	2.054	2.010	1.965	1.917	1.866	1.812
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.226	2.151	2.071	2.028	1.984	1.938	1.889	1.838	1.783
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.204	2.128	2.048	2.005	1.961	1.914	1.865	1.813	1.757
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.183	2.108	2.027	1.984	1.939	1.892	1.842	1.790	1.733
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.237	2.165	2.089	2.008	1.964	1.919	1.872	1.822	1.768	1.711
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.266	2.220	2.148	2.072	1.990	1.946	1.901	1.853	1.803	1.749	1.691
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.132	2.056	1.974	1.930	1.884	1.836	1.785	1.731	1.672
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.118	2.041	1.959	1.915	1.869	1.820	1.769	1.714	1.654
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.105	2.028	1.945	1.901	1.854	1.806	1.754	1.698	1.638
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.092	2.015	1.932	1.887	1.841	1.792	1.740	1.684	1.622
40	4.085	3.232	2.839	2.606	2.450	2.336	2.249	2.180	2.124	2.077	2.004	1.925	1.839	1.793	1.744	1.693	1.637	1.577	1.509
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.917	1.836	1.748	1.700	1.649	1.594	1.534	1.467	1.389
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.911	1.834	1.751	1.659	1.608	1.554	1.495	1.429	1.352	1.254
∞	3.842	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831	1.752	1.666	1.571	1.517	1.459	1.394	1.318	1.221	1.000

Table A.4 Percentiles of the *F* Distribution (continued)

F Table for alpha=.01



n_1= degrees of freedom for numerator) α= 0.01

n_2= degrees of freedom for denominator) α= .01

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.1	6022.5	6055.8	6106.3	6157.3	6208.7	6234.6	6260.6	6286.8	6313.0	6339.4	6365.9
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.020
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.312	3.162	3.084	3.003	2.920	2.835	2.746	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.931	2.781	2.702	2.620	2.535	2.447	2.354	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.211
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.805
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.192	2.035	1.950	1.860	1.763	1.656	1.533	1.381
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.039	1.878	1.791	1.696	1.592	1.473	1.325	1.000

Table A.5 Short Table of Random Digits

9221	3452	9754	8813	6679	3081	3945	9982	1510	827
8127	9422	6665	2154	5450	7042	1481	8014	0978	865
4927	7883	9205	7590	3627	1888	1641	8781	4144	508
6213	5446	3717	0552	9623	0088	9843	0634	6831	832
5334	7757	4781	5395	3258	0296	7318	2323	2485	471
3716	5373	2907	9488	1166	1247	4451	6553	0183	394
5145	3765	9427	6725	4461	1222	8457	8517	5350	868
9851	0178	9628	7183	4597	7034	6444	4269	3567	546
3074	1224	2657	6255	2151	2313	6834	6522	7554	639
1588	1272	6511	3349	2966	8131	6845	8587	1967	646
6687	3025	2377	8365	8044	1782	0422	6664	3636	834
0637	9907	1325	1033	6821	3481	5749	6173	1494	546
7231	7647	3772	5057	0121	6369	2426	2618	2569	811
7484	4844	5451	1561	4101	5482	5315	5877	6473	786
7642	6822	8268	5671	6734	5773	6103	5138	2626	053
3371	6578	6175	6397	8748	0442	4084	6057	6865	695
2147	2406	3484	4852	3824	3382	4844	6785	9836	173
8599	7373	4168	6777	5902	3235	0784	0686	2108	841
0165	0783	5414	6605	8781	4367	4422	9073	1186	852
6838	7236	7321	7445	2493	4579	8267	1116	5165	726
6239	7121	3293	3523	8586	1748	6837	2329	6150	101
8630	9465	1743	7576	8735	7719	8041	4547	4376	674
8136	9768	0114	3089	7523	1724	3166	6376	8269	636
6457	8345	2083	1333	6147	8113	5458	2718	9540	360
4911	0984	1290	0011	7488	7126	8475	6076	7131	782
8326	4220	5646	2386	6453	8472	3784	5781	6127	346
7831	5687	6449	3576	9733	1519	7275	2332	5845	677
9599	9382	3147	0763	4663	7017	0611	4003	7556	261
1774	1331	7482	1843	7638	4427	6117	2794	2153	624
0715	6621	5781	3165	8465	2177	1975	3224	0256	993
3208	8414	9263	0753	1886	8883	8593	6527	8207	762
4730	3362	4449	4956	6728	5518	1642	4717	6252	366
7827	3312	2765	4981	7261	3083	9756	2406	5770	146
4137	1877	1339	1047	1243	6217	6452	2915	9888	582
6514	1122	3156	6239	3702	7011	7979	4711	5875	648
8394	5148	6209	8063	2389	1893	9125	6024	6337	845
5531	8886	8814	2202	2345	2212	3722	3606	5267	112
6271	8587	2617	8619	5073	3825	8246	5528	2517	770
2846	0853	9442	5267	4900	6451	1633	1359	4895	233
1996	0414	6748	2185	4977	3271	9411	4189	4257	087
8093	6724	3822	4761	9475	1245	3694	3453	1188	637
8688	7831	9877	4522	1423	1922	3852	1543	8419	081
1756	7154	3675	7702	0852	3859	5641	0651	9345	058
8112	8434	8807	0362	7829	1944	4644	5104	5348	422
1421	2538	3284	9826	9634	7782	8541	2381	6327	316
7298	6803	2634	8535	0604	7546	1468	5351	7794	206
2782	3515	8334	5127	4282	7832	2779	5837	3176	467
3209	9124	7177	1739	6871	7124	8845	8928	8434	900
9572	2166	0252	0571	2982	2319	3913	8976	2482	579
1877	3400	4822	3143	6211	3138	0253	4542	7575	530
0635	8728	9471	6782	8265	1065	7239	6273	6455	201
2282	4461	1485	2323	1556	2248	8216	7776	3554	655
3638	1484	5648	2673	5642	6048	9777	6035	6918	235
6938	4589	4821	5443	5515	3476	5885	2462	6393	943
7337	3261	6451	8698	0813	0249	5328	4866	1410	617
8782	5228	8387	7138	8954	5868	5456	5640	6977	857
5840	0548	4630	1764	5476	9654	7243	1354	9241	681
3307	5215	7673	3446	2861	1863	6363	4484	4684	676
1001	6776	0247	5929	9125	7368	3679	8355	3816	267
4877	4617	2257	0256	0135	4269	2567	7410	7322	877
5032	9802	9831	6230	3636	8888	2354	7093	7216	768
5230	3531	0676	9427	0702	7546	5664	9757	0353	110
3342	5442	6198	6621	7387	9691	1833	1387	1754	511

Table A.6 Short Table of Random Normal Deviates

$(\mu = 0, \sigma = 1)$

−0.670	0.518	0.387	0.523	0.641	1.243	0.322	−2.607	−1.097	−0.012
−2.912	1.448	1.343	−0.122	0.726	−0.617	0.609	2.319	−0.450	−1.197
−0.028	−0.790	0.057	1.425	1.940	1.161	−0.878	−0.716	−0.244	−1.151
−1.257	0.774	0.003	0.388	1.060	1.028	−0.236	1.172	0.442	−0.157
2.372	−1.376	−1.318	1.236	0.738	0.337	−0.534	0.090	0.886	0.676
−0.970	0.438	−0.672	−0.180	0.667	1.370	−0.481	0.329	0.842	0.449
−1.228	0.129	−0.426	−0.165	0.028	2.696	1.201	−1.351	0.724	−1.017
−0.369	0.310	0.432	0.237	0.884	−1.224	0.539	0.852	0.497	−0.283
1.161	1.219	1.615	0.336	1.100	−0.528	0.161	0.278	0.675	−1.143
−0.284	2.609	0.792	1.825	−0.249	1.654	0.621	0.979	−1.472	−1.173
−0.578	−0.789	0.106	0.832	−0.597	0.496	−0.561	−1.033	−0.578	−0.378
0.074	0.261	−0.766	−1.046	0.361	−0.043	−1.927	1.527	0.605	1.475
0.230	0.046	0.978	−1.901	1.162	−0.545	0.697	1.151	2.033	0.080
2.162	−0.562	1.190	0.925	−1.057	0.015	−1.371	1.067	−1.080	−1.129
−1.020	−1.130	−0.315	0.628	−0.140	2.050	−0.030	−0.629	0.128	−1.221
1.323	−0.836	−0.284	−0.249	−0.768	1.242	−1.879	−0.417	0.013	−0.502
2.329	1.884	0.033	0.598	−0.217	0.260	0.431	−1.914	0.205	1.155
2.761	1.800	−0.562	0.714	−0.407	0.009	−0.724	−1.168	0.247	1.166
−0.742	0.210	−0.741	−1.099	0.158	2.112	−0.765	−0.319	−0.247	0.345
−1.410	0.413	0.705	1.444	1.057	−0.843	0.043	−0.571	−0.001	0.203
2.272	−0.719	0.679	2.007	−0.180	0.698	−1.137	0.688	−0.571	−0.100
2.832	0.925	−1.350	1.529	−0.260	−1.007	−2.350	−1.501	0.289	1.522
−1.086	−0.558	−0.973	−1.285	−0.021	0.077	0.915	−0.241	−0.249	−0.529
0.134	1.815	0.313	1.571	−0.216	2.261	0.696	−0.130	0.393	0.017
0.783	0.600	−0.745	1.127	−0.684	−0.519	0.125	−0.499	1.543	−0.082
0.174	−0.897	0.575	−0.751	0.694	−2.959	0.529	1.587	0.339	−0.813
−1.319	0.556	2.963	1.218	1.199	−1.746	1.611	0.467	−0.490	0.202
1.298	−0.940	−1.143	−1.136	−1.516	0.548	0.629	0.250	−1.087	0.322
−0.676	−1.107	−1.483	0.278	0.493	−0.442	1.078	−0.336	−0.177	−0.057
−1.287	0.775	−1.095	1.161	−1.877	1.874	1.703	−1.619	−0.725	−1.407
0.260	−0.028	−1.982	0.811	0.999	1.662	0.908	1.476	−1.137	−0.945
0.481	1.060	1.441	0.163	0.720	1.490	−0.026	−0.502	0.427	−0.351
0.794	0.725	1.971	0.384	−0.579	−1.079	−1.440	−0.859	−0.346	0.007
0.584	−0.554	1.460	0.791	−0.426	−0.682	0.430	1.922	−2.099	0.221
−0.114	0.379	−0.698	1.570	−0.511	−0.725	0.680	−0.591	−1.091	0.357
−1.128	−1.707	0.921	−0.859	−1.566	1.523	−0.900	−0.988	0.264	0.282
0.691	0.153	0.076	1.691	0.553	0.457	−1.107	0.322	0.633	0.007
1.115	0.777	−0.738	0.868	1.484	−0.792	0.950	−0.842	−0.192	0.620
−0.389	0.559	0.670	−0.315	1.234	0.475	1.117	1.286	−0.649	−1.880
0.330	0.750	−0.642	0.148	−0.608	0.866	−1.720	0.653	−0.210	−0.959
−0.333	−0.084	1.239	−0.049	−0.095	−0.197	−0.213	−1.420	−0.491	0.102
1.718	1.111	−0.548	−0.653	1.534	−0.456	−0.395	1.614	−0.531	−0.785
−0.182	0.620	1.178	−1.071	0.444	−0.072	−1.001	1.325	−0.302	−1.119
1.260	−1.192	0.182	−0.397	−0.705	−1.085	−1.492	1.642	0.673	−0.707
−1.204	−1.725	1.695	1.473	0.665	−0.489	0.020	0.267	1.230	0.865
−0.619	0.307	−0.226	−0.096	0.987	−1.195	−1.412	0.433	2.052	0.022
−0.272	−0.096	0.137	−0.361	0.653	−0.156	1.309	−0.480	−0.397	1.302
0.245	−0.690	0.493	−1.123	1.465	0.132	0.582	−0.429	0.225	0.125
0.101	−0.855	0.782	−1.040	2.113	−1.423	−1.010	0.158	0.106	−1.232

GLOSSARY OF TERMS

Accuracy of information Degree to which information reflects reality.

Accuracy in sampling Is affected by nonresponse bias, memory error, misunderstanding of questions, problem of definition of terms, and processing error,

A.C. Nielsen BASES Panel-based research conducted by the marketing research/consulting firm A.C. Nielsen using the Internet for data collection.

Adequate sample A sample that is of sufficient size to provide confidence in the stability of its characteristics.

After-only with control group design A true experiment requiring one treatment and an "after" measurement of both the experimental and control groups.

All-variable regression Multiple regression with all independent variables included simultaneously.

Alpha error See *Type I error*.

Alternate forms reliability The extent to which measurement scores from equivalent forms of the measure administered to the same sample are the same.

Alternative hypothesis A hypothesis making statement of expected difference.

Ambiguity of questions and responses Errors made in the transmission of information in interpreting written or spoken words or behavior

Analysis The categorizing, the aggregating into constituent parts, and the manipulation of data to obtain answers to the research question(s) underlying the research project.

Analysis of variance (ANOVA) A test of the differences in k group means for one or more variables (factors).

Applied research See *decisional research*.

Area sample A probability sample where the primary sampling is of geographical units.

Arithmetic mean See *mean*.

ASP model See *Online Application Service Provider*.

Associative variation A measure of the extent to which occurrences of two variables, or changes of two variables, are associated. Also known as "concomitant variation."

Attitude measurement Measurement of the person's mental state toward a certain stimuli according to specified sets of dimensions and instructions.

Audimeter An electromechanical devise used to record automatically the times a household's television set is turned on and off and the stations which it is tuned to.

Availability of information Availability of information when a decision is being made.

Balanced scale A rating scale with an equal number of response alternatives in opposite directions from some mid-point, e.g., positive and negative alternatives.

Balancing See *matching*.

Bases for inferring causation The means by which causation can be inferred: (1) associative variation, (2) sequence of events, and (3) absence of other possible causal factors.

Basic research See *fundamental research*.

Before-after with one control group design A true experimental design with one treatment and with measurements of the experimental and control groups made both "before" and "after" the treatment is administered.

Behaviorally-anchored rating scale (BARS) A rating scale using behavioral incidents to define each position on the rating scale rather than verbal, graphic, or numeric labels.

Beta error See *Type II error*.

Bias The difference between the true value of that which is being measured and the average value derived from a number of independent measures of it.

Bivariate analysis The analysis of relationships between two variables.

Bounded recall An approach for reducing telescoping by asking questions about events of concern in previous time periods as well as the time period of research interest.

Bulletin Board Qualitative research online technology at a website that allows users to register and then participate in unmoderated discussions of topics of interest to participants of the bulletin board.

Canonical correlation analysis A generalization of multiple correlation analysis to two or more criterion variables.

Canonical correlation coefficient In discriminant analysis it is a measure of the association that summarizes how related the discriminant function is to the groups, and in canonical correlation analysis it is a measure of the association between the independent variables and the set of dependent variables.

CATI A Computer Aided Telephone Interview in which the interviewer reads the questions from a computer and enters the responses directly.

Causal study Research design which attempts to determine the causes of what is being predicted—i.e., the "reasons why."

Census All members of the population are included for study.

Central tendency error Reluctance of respondents to give extreme scores or to use an extreme position on an individual scale item.

Cheater question A question included in a questionnaire that will disclose a respondent's giving fabricated answers.

Chi-square test For a simple tabulation it tests the goodness-of-fit between the observed distribution and the expected distribution of a variable, and for cross tabulation it tests whether the observed association or relationship between the variables is statistically significant.

Classification matrix A cross-tabulation-type table showing the results of discriminant analysis' ability to correctly classify observations into categories of a criterion variable

Cluster analysis A class of statistical techniques whose objective is to separate objects into groups such that the similarity of objects within each group is maximized while maximizing the difference between groups.

Cluster sampling A probability sample in which a simple random or stratified random sample is selected of all primary sampling units, and then all elements within the selected primary units are sampled.

Codebook/coding manual A manual which shows how the data have been coded for analysis.

Coding The process by which responses are assigned to data categories and symbols (usually numbers) are assigned to identify them with the categories.

Coefficient alpha A measure of internal consistency reliability for a multi-item measure which is a type of mean reliability coefficient for all possible ways to split the items into two groups.

Coefficient of determination (R^2) A measure of the strength of association between variables in regression, it specifies how much of the variation in the dependent variable can be explained by the variation in the independent variable(s).

Coefficient of multiple correlation Correlation coefficient when the number of independent variables is two or more.

Commercial data Data sold in the form of syndicated services. Collected by commercial marketing research firms or industry associations. See *syndicated services*.

Comparative rating scale Objects are rated in comparison with other objects.

Completely randomized design A statistical design where experimental treatments are assigned to test units on a random basis.

Concept An abstraction formed by generalization about particulars.

Conceptual definition A construct is defined in terms of other constructs. Also known as *constitutive definition*.

Concept testing How people, without prompting, interpret deliberately a sketchy idea for a new product or service.

Concomitant variation See *associative variation*.

Concurrent validity See *criterion validity*.

Confidence interval A range of values with a given probability of covering the true population value (parameter).

Confounding In an experiment, it is the tangling effects of two or more levels of a treatment variable or two or more treatment variables.

Confusion matrix See *classification matrix*.

Conjoint analysis A technique of research that measures psychological judgments by decomposing a set of overall responses to a set of factorially designed stimuli so that the utility of each stimulus attribute and attribute level can be inferred from the respondent's overall evaluations of the stimuli.

Constant sum question A question in which the respondent is asked to allocate a fixed sum of points to options to show the importance (or some other attribute or characteristic) of each option.

Constitutive definition See *conceptual definition*.

Construct A concept that is the conscious invention of researchers to be used for a special research purpose.

Construct validity A form of external validity that assesses the extent to which generalizations can be made about higher-order constructs from research operations; it is a measurement issue and is concerned with how and why a measurement works.

Content analysis Coding of free responses to open-end questions.

Content validity The extent to which a scale or measurement instrument represents the universe of the property or characteristic being measured.

Continuity correction A correction factor in chi-square analysis that adjusts for the use of a continuous distribution to estimate probability in a discrete distribution. Known as the Yates continuity correction.

Continuous panel A sample of individuals, households, or firms from whom information is obtained at successive time periods.

Controlled experiment A research design in which the investigator intervenes by manipulating at least one assumed causal variable and in which subjects (respondents) are assigned randomly to experimental and control groups.

Convergent validity A type of construct validity concerned with the correspondence in results between measuring the same construct by two or more independent methods.

Convenience sample A nonprobability sample chosen by a convenient process and because elements are easy to obtain.

Conversational interviewing In personal or telephone interviewing the interviewer answers queries from respondents about question meaning and understanding.

Correlation analysis The analysis of the extent to which changes in one variable are related to changes in one or more other variables.

Correlation coefficient A measure of the association between two or more variables.

Correspondence analysis A special case of canonical correlation analysis which examines the relations between the categories of two discrete variables.

Coverage error Occurs when the sample frame or group from which the sample is drawn does not represent the population as a whole. See also *frame error*.

Criterion validity The extent to which the measurement instrument works in predicting the future (predictive validity) or reflects the present (concurrent validity). Also known as *pragmatic validity*.

Criterion variable See *dependent variable*.

Critical path method (CPM) A network approach in which the component activities are diagrammed in sequence of performance and a time estimate for each activity is presented. See *Program Evaluation and Review Technique (PERT)*.

Cross-over design A statistical experimental design in which different treatments are applied to the same test units in different time periods.

Cross sectional design A research design where several groups are measured at the same time, with each group having been exposed to a different level of the treatment variable.

Cross tabulation The simultaneous counting of the number of observations that occur in each of the data categories of two or more variables.

Cross validation A procedure in regression analysis and discriminant analysis for examining whether the predictive equations derived hold up beyond the data on which parameters are based.

Cumulative scale A scale constructed of a set of items with which the respondent indicates agreement or disagreement, it is unidimensional, and there is a pattern of item responses that is related to the total score.

Data collection techniques The means by which data are collected by communication or by observation.

Data matrix A rectangular array of data entries where the rows are a respondent's responses and the columns are the variables or data fields for the responses. Also known as the *basic data matrix*.

Decision model See *problem-situation model*.

Decisional research Applied research which attempts to use existing knowledge to aid in the solution of some given problem (s).

Dependent variable The effect of interest or outcome in an experiment.

Depth interview An often unstructured interview that is used to explore the underlying predisposition, needs, desires, feelings, and emotions of consumers toward products and services. May consist of direct and/or indirect questions.

Descriptive study Provides information such as describing market characteristics or functions on groups and phenomena that already exist; there is prior formulation of specific research questions. Also known as an "observational study."

Determinants of a research project The problem, the researcher, the respondent/subject, and the client.

Deterministic cause. Any event that is necessary and sufficient for the subsequent occurrence of another event.

Dichotomous question A multiple-choice question with only two alternative responses.

Differential scale A rating scale, assumed to be interval, in which a respondent is asked to agree with only a subset of items (statements) about an object, with each item having a predetermined scale value (position on the scale), and the items agreed with correspond to the respondent's position on the dimension being measured.

Direct interview An interview in which the purposes of the questions are not purposely disguised.

Direct-judgment rating method A respondent is asked to give a numerical rating to each stimulus with respect to some designated attribute.

Directory-based sampling A directory or other physical listing is used as a sample frame to select sample elements to be called in a telephone survey.

Discriminant analysis A multivariate technique of analysis that examines the ability of predictor variables to discriminate (i.e., to separate) between categories of a criterion (dependent) variable, or groups.

Discriminant validity A form of construct validity which assesses the extent to which a measure is unique and not simply a reflection of other variables.

Disk-by-Mail A type of survey where the questionnaire is placed on a personal computer diskette and mail is used to send it to a respondent and by the respondent to return it to the research organization.

Disproportionate stratified sampling A stratified sample in which characteristics other than just relative size, such as relative size of stratum variances, are taken into account.

Distance A characteristic of the real number system where the differences between ordered numbers are ordered.

Double opt-in mailing list A list of people with e-mail addresses who have indicated agreement to receive surveys.

Drop-off/pick-up survey A survey where the questionnaire is left with a respondent and a representative of the research organization returns later to pick up the completed questionnaire, or the questionnaire can be returned by mail.

Editing The process of reviewing the data to ensure maximum accuracy and clarity.

Electronic surveying See *online research*.

E-mail submission form used when the researcher builds a HTML online survey, distributes actively to the respondent, and receives the respondent's answers as part of an e-mail message that is directed back to the researcher.

Emic vs. etic issue An issue in cross-national/cultural research in which a researcher must decide whether constructs and methods are culture-specific (emic) or culture-free (etic).

E-panel Doing panel research using the Internet for data collection.

Equal-appearing intervals See *differential scale*.

Ethical behavior Behavior conforming to professional standards of conduct; it is what most people in a given society view as being moral, good, or right.

Ethics Moral principles, quality, or practice.

Ethnography A qualitative approach to research that studies human behavior within a cultural context.

Executive summary That part of a formal research report which reduces the essentials of the study—the why, the what, the how, the conclusions, and the recommendations—to one or two pages.

Experimental error Noncorrespondence of the true (or actual) impact of, and the impact attributed to, the independent variable (s).

Experimentation A research method where there is researcher intervention and control over the factors affecting the response variable of interest, thus allowing for the establishment of causal relationships.

Explicit model A model described verbally, graphically or diagrammatically, mathematically (symbolically), or as a logical sequence of questions (logical flow).

***Ex post facto* design** A quasi-experiment in which the test and control groups are not known until after the treatment has been administered.

Exploratory study A study whose purposes include the identification of problems, more precise formulation of problems (including identification of relevant variables), and the formulation of new alternative courses of action.

External secondary information Secondary information that must be obtained from outside sources.

External validity The generalizability of a relationship beyond the circumstances under which it is observed.

Extraneous variable A variable other than the manipulated independent variable that could influence the dependent variable.

Factor A variable or construct that is not directly observable but is developed as a linear combination of observed variables.

Factor analysis A class of statistical techniques whose purpose often is data reduction and summarization which is accomplished by representing a set of observed variables, persons, or occasions in terms of a smaller number of hypothetical, underlying and unknown dimensions which are called factors.

Factorial design A statistical experimental design where there is an equal number of observations made of all combinations involving at least two levels of at least two variables.

False negative error A respondent reports not to have an attitude when he or she really does have one.

False positive error Statements by respondents that appear to be complimentary, but really are not, or when respondents appear to have an attitude and they do not.

Field experiment An experiment conducted in a natural environmental setting.

Fixed-size sampling The number of elements to be included in the sample is decided upon in advance.

Focus group A group of topic knowledgeable people who jointly participate in an interview that does not use a structured question-and-answer methodology. Usually consists of 8 to 12 people selected purposively.

Formal research report Consists of a number of components which an be organized in to three components: prefatory pages, report body, and appended parts.

Four-group, six-study design Combines an after-only with control group design and a before-after with control group design.

Fractionation A rating scale in which the respondent is given two stimuli at a time and is asked to give some numerical estimate of the ratio between them, with respect to some attribute.

Frame error Noncorrespondence of the sought sample to the required sample. Occurs when the sample frame is incomplete, has multiple entries for elements, or has elements included that are not in the relevant population.

Framing effects The difference in response to objectively equivalent information depending upon the manner in which the information is labeled or framed.

Free answer question A question that has no fixed alternatives to which the answer must conform. Also known as *open-ended text*.

Frequency distribution See *simple tabulation*.

Full profile conjoint analysis Conjoint analysis where different stimulus (e.g., a product) descriptions are developed and presented to the respondent for acceptability or preference evaluations.

Fundamental research Seeks to extend the boundaries of knowledge in a given area with no necessary immediate application to existing problems.

Funnel approach An approach to questionnaire design that specifies a sequence of questions where one proceeds from the general to the specific or from the easier questions to answer to those that are more difficult to answer.

Goodness-of-fit test An analysis of whether the data obtained in a research study fit or conform to a model or distribution.

Graphic positioning scale A semantic differential used for multiple object ratings where all objects are evaluated on each scale item.

Graphic rating scale A rating scale in which a respondent indicates his/her rating of a stimulus on a graphical response item.

Guided imagery A modified TAT where participants are asked to appraise a product or brand by concentrating on creating and experiencing an associated image.

Guttman scalogram analysis See *cumulative scale*.

History Events outside an experimental design that affect the dependent variable.

Hypothesis An assertion about the "state of nature" or the relation between things that often, from a practical standpoint, implies a possible course of action with a prediction of the outcome if the course of action is followed.

Implicit model A model that guides a decision but has not been specified in an explicit or formal manner.

Inaccuracy in response Errors made in the formulation of information. May be *concurrent* or *predictive* (e.g., when reported intentions are not carried out).

Independent variable In an experiment, it is a variable whose effect upon some other variable the experiment is designed to measure; it is the variable that is manipulated and is also known as the *treatment* variable.

Indexes of agreement Measures of the strength of association between two variables in a cross tabulation, including the phi correlation coefficient, the contingency coefficient, the lambda-asymmetric coefficient, and the lambda-symmetric coefficient.

Indirect interview An interview that is neither fully structured nor unstructured, and in which the purposes of the questions asked are intentionally disguised.

Information Recorded experience that is useful for decision making; communicated knowledge which changes the state of knowledge of the person who receives it.

In-store interviewing A type of mall intercept in which the interviews take place in a single store, usually at the point of purchase.

Instrument effect Changes in the measuring instrument or process that may affect the measurement obtained in an experiment.

Intentions Presently planned actions to be taken in a specified future time period.

Interaction The situation in an experiment where the response to changes in the levels of one treatment variable is dependent on the level of some other treatment variable(s).

Interactive interviewing Interviews that are conducted by having a respondent respond on a personal computer. Some software may customize new questions based on responses to previously answered questions.

Intercoder reliability The reliability of coding done by multiple persons.

Internal consistency reliability Reliability within single testing occasions in which the variables are grouped.

Internal secondary information Secondary information that is available from within the company or the organization.

Internal validity Assesses whether the observed effect is due solely to the experimental treatments and not due to some extraneous variable(s).

Interpretation The process of taking the results of analysis, making inferences relevant to the research relationships studied, and drawing managerially useful conclusions about these relationships.

Interval scale A measurement scale that possesses the characteristics of order and distance, and the zero point of the scale is arbitrary.

Interview A form of person-to-person (dyadic) communication between two parties that involves the asking and answering of questions.

Interviewer A person who asks questions in an interview of a respondent.

Judgment sample A nonprobability sample where the elements to be included are selected on the basis of the researcher’s sound judgment or expertise and an appropriate strategy.

Kolmogorov-Smirnov one-sample test A goodness-of-fit test of the agreement between an observed distribution of a set of sample values and some specified theoretical distribution

Kolmogorov-Smirnov two-sample test A test of whether two independent samples come from the same population or from populations with the same distribution.

Kurtosis The shape of a data distribution in terms of height or flatness.

Laboratory experiment An experiment conducted in a controlled laboratory or laboratory-type setting.

Laddering See *means-end analysis*.

Leniency error occurs when respondents consistently use the extreme positions on a rating scale with relatively little use of intermediate scale positions.

Likert scale A balanced rating scale in which a respondent is asked to indicate extent of agreement with a series of statements, using a set of verbal categories from "strongly agree to "strongly disagree" for response. See also *summated scale*.

Limited-response category scale A rating scale in which a respondent is limited to choosing from a predetermined set of response categories.

Logit A type of multiple regression analysis where the categorical dependent variable is assumed to follow a logistic distribution.

Mail interview A type of survey where the questionnaire is sent to a respondent by mail and the respondent returns the completed questionnaire by mail.

Make or buy decision The decision by a research client whether the research is to be done in-house (make) or by an outside supplier (buy).

Mall intercept Interviews are stationed at selected places in a shopping mall or other centralized public place and they request interviews from people who pass by.

Management summary See *executive summary*.

Mann-Whitney U test A test of whether two independent groups providing data are from the same population and whether there is a relationship between two variables.

Marketing Decision Support System (MDSS) A coordinated collection of data, systems, tools, and techniques with supporting software and hardware by which an organization gathers and interprets relevant information from the business and the environment and turns it into a basis for marketing action.

Marketing information system A "formal" system within an organization for obtaining, processing, and disseminating decision information. Subsystems are marketing research, internal records, marketing intelligence, and information analysis.

Marketing intelligence A subsystem of a MIS in which a set of procedures and sources are used to provide information about relevant developments in the marketing environment.

Marketing research The systematic and objective search for, and analysis of, information relevant to the identification and solution of any problem in the field of marketing.

Matching A control technique where subjects are equated on the variable(s) to be controlled. Also known as *balancing*.

Maturation Changes that occur with the passage of time in the people involved in an experimental design.

Mean The point on a scale around which the values of a distribution balance; it is the sum of all the values divided by the number of respondents.

Means-end analysis An in-depth one-on-one interviewing technique that identifies the linkages people make between product attributes (means), the benefits derived from those attributes (the consequences), and the values that underlie why the consequences are important (the ends). Also known as "Laddering" and "Means-End Chain."

Measurement A way of assigning symbols to represent the properties of persons, objects, events, or states, which symbols have the same relevant relationships to each other as do the things represented.

Measurement error The difference between the information obtained and the information wanted by the researcher; it is generated by the measurement process itself.

Median The midpoint of the data in a distribution.

Memory error Inaccuracy in response that occurs when a respondent gives the wrong factual information because of not remembering an event asked about.

Method of Choices A procedure for indirectly arriving at paired comparison proportions of the form p(B>A) by asking respondents to choose the one of a set of stimuli that has the "most of," is the "best," or is "preferred," etc. on the basis of the attribute or characteristic being studied.

Method of inquiry The broad approach to conducting a research project and the philosophy underlying the approach. Methods include objectivist, subjectivist, Bayesian, and phenomenologist.

Metric measurement Direct numerical judgments made by a respondent which are assumed to be either interval- or ratio-scaled.

Metric multidimensional scaling Multidimensional scaling in which the input data are ratio-scaled.

Mind Track A brainwave-to-computer interface developed by Advanced Neurotechnologies, Inc. that measures direct emotional response to most any communication medium.

MIS See *marketing information system*.

MIS activities Discovery, collection, interpretation, analysis, and intra-company dissemination of information.

Misunderstanding error Inaccuracy in response often due to careless question design.

Mode The typical or most frequently occurring value in a distribution.

Model The linking of propositions together in a way that provides a meaningful explanation for a system or process.

Moderator A person conducting a focus group whose job is to direct the group's discussion to the topics of interest.

Monadic rating scale Each object is rated by itself independently of any other objects being rated.

Multicollinearity A condition in multiple regression analysis where the predictor variables show very high correlation among themselves.

Multidimensional scaling A set of techniques that portray psychological relations among stimuli—either empirically obtained similarities or preferences (or other kinds of orderings)—as geometric relationships among points in a multidimensional space.

Multi-item scale A scale consisting of a number of closely related individual rating scales whose responses are combined into a single index or composite score or value. See also *summated scale*.

Multiple choice question A question that has at least two fixed alternative response categories and the respondents can select **k** out of **n** choices.

Multiple correlation analysis Correlation analysis when the number of independent variables is two or more.

Multiple regression analysis Regression analysis with two or more independent variables.

Multiplicity sample See *snowball sample*.

Multistage sampling A multilevel probability sample in which a sample is selected of larger areas (or groups), and then a sample is selected from each of the areas (groups) selected at the first level, and so on.

Multitrait Multimethod Matrix A generalized approach for establishing the validity and reliability of a set of measurements (traits).

Multivariate analysis Statistical procedures that simultaneously analyze measurements of multiple variables on each individual or object under study.

Natural experiment An experiment in which the investigator intervenes only to the extent required for measurement, and there is no manipulation of an assumed causal variable. The variable of interest has occurred in a natural setting, and the investigator looks at what has happened.

Nominal scale A measurement scale that does not possess the characteristics of order, distance, and origin.

Nomogram A graphic instrument for specifying sample size relating allowable error, confidence level, mean or proportion, and standard deviation.

Nomological validity A form of construct validity which attempts to relate measurements to a theoretical model that leads to further deductions, interpretations, and tests.

Nonmetric multidimensional scaling Multidimensional scaling in which input data are rank order data (ordinally-scaled), but which output is interval-scaled.

Nonparametric statistical methods Distribution-free methods in which inferences are based on a test statistic whose sampling distribution does not depend upon the specific distribution of the population from which the sample is drawn.

Nonprobability sample A sample selected based on the judgment of the investigator, convenience, or by some other means not involving the use of probabilities.

Nonresponse error Noncorrespondence of the obtained sample to the original sample.

Nonsampling error All errors other than sampling error that are associated with a research project; typically is a systematic error but can have a random component.

Null hypothesis A hypothesis which states no difference.

Numerical comparative scale A semantic differential used for multiple object ratings where all objects are evaluated on each scale item using a verbally-anchored numerical scale.

Numerical rating scale A rating scale that uses a series of integers that may, or may not have verbal descriptions, to represent degrees of some property.

Observation technique Information on respondents' behavior is obtained by observing it rather than by asking about it.

"One more question" syndrome The tendency to add an additional question to a survey because the cost is very low to do so.

One-on-one interview See *depth interview*.

Online Application Service Provider (ASP) Accessed through the Internet, where surveys are built online, requiring no user software, server, or IT support.

Online research Using the Internet as a mode of data collection. Often used in conjunction with e-mail.

Operational definition Assigns meaning to a variable by specifying what is to be measured and how it is to be measured.

Order A characteristic of the real number series in which the numbers are ordered.

Ordered-category sorting A respondent assigns (sorts) a set of stimuli into different categories, which are ordered on the basis of some property.

Ordinal scale A measurement scale that possesses only the characteristic of order; it is a ranking scale.

Origin A characteristic of the real number series where there is a unique origin indicated by the number zero.

Paired comparisons The respondent is asked to choose one of a pair of stimuli on the basis of some property of interest.

Pantry audit A data collection technique whereby a field worker takes an inventory of brands, quantities, and package sizes that a consumer has on hand.

Parameter A summary property of a collectivity, such as a population, when that collectivity is not considered to be a sample.

Partially structured indirect interview An interview using a predevised set of words, statements, cartoons, pictures, or other representation to which a person is asked to respond, and the interviewer is allowed considerable freedom in questioning the respondent to ensure a full response.

Personal interview An interviewer asks questions of respondents in a face-to-face situation.

Pictogram A pictorial chart that depicts data with the help of symbols such as stars, stacks of coins, trees, facial expressions, caricatures of people, and so forth.

Pilot study A small-scale test of what a survey will be, including all activities that will go into the final survey.

Planned information Exists when a manager recognizes a need and he or she makes a request that information be obtained.

Politz-Simmons method A method of estimating both the direction and magnitude of nonresponse error.

Population The totality of all the units or elements (individuals, households, organizations, etc.) possessing one or more particular relevant features or characteristics in common, to which one desires to generalize study results.

Population specification error Noncorrespondence of the required population to the population selected by the researcher.

Popular report A research report that minimizes technical details and emphasizes simplicity.

Postcoding Coding done after the data are collected.

Power of a hypothesis test It is 1 minus the probability of a Type II error ($1-\beta$).

Practical significance See *substantive significance*.

Pragmatic validity See *criterion validity*.

Precision Refers to *sampling error* and the size of the confidence limits placed on an estimate.

Precoding Coding done before the data are collected.

Predictive validity See *criterion validity*.

Predictor variable See *independent variable*.

Pre-experimental design A research design with total absence of control.

Pretesting The testing of a questionnaire or measurement instrument before use in a survey or experiment.

Probabilistic cause Any event that is necessary, but not sufficient, for the subsequent occurrence of another event.

Probability sampling Every element in the population has a known nonzero probability (chance) of being selected for inclusion in a study.

Probit A type of multiple regression analysis where the categorical dependent variable is assumed to be normally distributed.

Problem formulation A stage in the research process in which a management problem is translated into a research problem.

Problem-situation model A conceptual scheme that specifies a measure of the outcome(s) to be achiever, the relevant variables, and their functional relationship to the outcomes(s).

Program Evaluation and Review technique (PERT) A probabilistic scheduling approach using three time estimates: optimistic, most likely, and pessimistic. See also *critical path method (CPM)*.

Projection A research technique whereby a respondent projects his/her personality characteristics, etc. to a non-personal, ambiguous situation that he/she is asked to describe, expand, or build a structure around it.

Proportionate stratified sampling A stratified sample in which the sample that is drawn from each stratum is proportionate in size to the relative size of the stratum in the population.

Proposition A statement of the relationship between variables, including the form of the relationship.

Protocol A record of a respondent's verbalized thought processes while performing a decision task or while problem solving (concurrent) or just after the task is completed (retrospective)

Psychogalvanometer A device for measuring the extent of a subject's response to a stimulus, such as an advertisement.

Purposive sampling See *judgment sample*.

Q-sort A scaling technique in which the respondent is asked to sort a number of statements or other stimuli into a predetermined number categories, formed on the basis of some criterion, with a specified number having to be placed in each category.

Quasi-experimental design A controlled experiment design where there is manipulation of at least one assumed causal variable but there is not random assignment of subjects to experiment and control groups.

Questionnaire An instrument for data collection that requests information from respondents by asking questions.

Quota sample A nonprobability sample in which population subgroups are classified on the basis of researcher judgment and the individual elements are selected by interviewer judgment.

Random-digit-dialing A probability sampling procedure used in telephone surveys where the telephone number to be called is generated by selecting random digits.

Randomized response technique A technique for obtaining information about sensitive information.

Random sampling error See *sampling error*.

Rank correlation The correlation between variables that are measured by ranking. Measures used are Spearman *rho* and Kendall *tau*.

Ranking Respondents are asked to order stimuli with respect to some designated property.

Rank order question A question where the answer format requires the respondent to assign a rank (order) position for the first, second,…, to the *n*th item to be ordered.

Rating A measurement method where a respondent paces that which is being rated along a continuum or in one of an ordered set of categories.

Ratio scale A measurement scale possessing all the characteristics of the real number series: order, distance, and origin.

Reactive effects of experimental situation Effects that may arise from subjects' reacting to the situation surrounding the conduct of an experiment rather than the treatment variable.

Reactive effects of testing The learning or conditioning of the persons involved in an experimental design as a result of knowing that their behavior is being observed and/or that the results are being measured.

Regression analysis The mathematical relationship between a dependent variable and one or more independent variables.

Regression coefficient Represented by *b*, it shows the amount of change that will occur in the dependent variable for a unit change in the independent variable it represents.

Relevancy of information Pertinence and applicability of information to the decision.

Reliability The consistency of test results over groups of individuals or over the same individual at different times.

Repeated measures design A research design where subjects are measured more than once on a dependent variable. See also *cross-over design*.

Repertory grid A partially structured measurement technique that requires the respondent to compare objects along dimensions that he or she selects.

Representative sample A relatively small piece of the population that mirrors the various patterns and subclasses of the population.

Research design The specification of methods and procedures for acquiring the information needed to structure or to solve problems. The operational design stipulates what information is to be collected, from which sources, and by what procedures.

Research method Experimental or non-experimental; the major difference between the two lies in the control of extraneous variables and the manipulation of at least one assumed causal variable by the investigator in an experiment.

Research plan A formal written document that serves as the overall master guide for conducting a research project.

Research process A series of interrelated steps that define what a research project is all about, starting with problem formulation and ending with the research report.

Research proposal A shorter and less technical version of a research plan that is used to elicit the project and gain a commitment of funding.

Research question States the purpose of the research, the variables of interest and the relationships to be examined.

Research report The major vehicle by which researchers communicate by a written statement and/or oral presentation research results, recommendations for strategic and tactical action, and other conclusions to management in the organization or to an outside organization.

Respondent A person who participates in a research project by responding and answering questions verbally, in writing, or by behavior.

Response bias See *response error*.

Response error The difference between a reported value and the true value of a variable.

Robust statistical technique A technique of analysis whereby if certain assumptions underlying the proper use of the technique are violated, the technique performs okay and can handle such a violation.

Sample A subset of the relevant population selected for inclusion in a research study.

Sample design A statement about a sample that specifies where the sample is to be selected, the process of selection, and the size of the sample; it is the theoretical basis and the practical means by which data are collected so that the characteristics of the population can be inferred with known estimates of error.

Sample frame A means of accounting for the elements in a population, usually a physical listing of the elements, but may be a procedure which produces a result equivalent to a physical listing, from which the sampled elements are selected.

Sampling distribution The probability distribution of a specified sample statistic (e.g., the mean) for all possible random samples of a given size *n* drawn from the specified population.

Sampling error Variable error resulting from the chance specification of population from elements according to the sampling plan. Often called random sampling error, it is the non-correspondence of the sample selected by probability means and the representative sample sought by the researcher.

Sampling unit A population element which is actually chosen by the sampling process.

Scaling Generation of a continuum on which measured objects are located.

Scanner data Data on products purchased in retail stores that are obtained by electronic scanning at checkout of the Universal Product Code (UPC); unit and price information are recorded.

Scree chart In factor analysis, it is a discrete line chart that relates the amount of variance accounted for by each factor to the factor number (1 … k).

Secondary information Information that has been collected by persons or agencies for purposes other than the solution of the problem at hand, and which is available for the project at hand.

Selection error The sampling error for a sample selected by a nonprobability method. It is also a term used for the effect of the selection procedure for the test (treatment) and control groups on the results of an experimental study.

Self-hosted server software Survey building software that requires housing on the researcher's server.

Semantic differential A rating procedure in which the respondent is asked to describe a concept or object by means of ratings on a set of bipolar adjectives or phrases, with the resulting measurements assumed to be interval-scaled.

Sentence completion test A respondent is given a sentence stem (the beginning phrase) and is asked to complete the sentence with the first thought that occurs to him or her.

Sequential sample An approach to selecting a sample size whereby a previously determined decision rule is used to indicate when sampling is to be stopped during the process of data collection.

Simple random sample A probability sample where each sample element has a known and equal probability of selection, and each possible sample of *n* elements has a known and equal probability of being the sample actually selected.

Simple tabulation A count of the number of responses that occur in each of the data categories that comprise a variable. Also known as *marginal tabulation*.

Simulation A set of techniques for manipulating a model of some real-world process for the purpose of finding numerical solutions that are useful in the real process that is being modeled.

Single-source data Obtaining all data from one research supplier on product purchases and causal factors such as media exposure, promotional influences, and consumer characteristics from the same household.

Skewness A measure of a given data distribution's asymmetry.

Snowball sampling A nonprobability sample in which initial respondents are selected randomly but additional respondents are obtained by referrals or by some other information provided by the initial respondents.

Socioeconomic characteristics The social and economic characteristics of respondents, including for example, income, occupation, education level, age, gender, marital status and size of family.

Split-half reliability A measure of internal consistency reliability where the items in a multi-item measure are divided into two equivalent groups and the item responses are correlated.

Standard deviation A measure of dispersion (variation) around the sample mean, it is the square root of the variance.

Standard error The standard deviation of the specified sampling distribution of a statistic.

Standard error of the difference The standard deviation of the sampling distribution of the difference between statistics such as means and proportions.

Standardized interviewing In a survey using personal or telephone interviewing the interpretation of questions asked is left up to the respondent as the interviewer is not allowed to answer any query.

Stapel scale An even-numbered balanced nonverbal rating scale that is used in conjunction with single adjectives or phrases.

State of nature An environmental condition.

Static-group comparison A quasi-experimental design in which a group exposed to a treatment is compared to a group that was not exposed.

Statistical conclusion validity Involves the specific question whether the presumed independent and dependent variables are indeed related.

Statistical experimental design After-only designs in which there are at least two treatment levels. Includes completely randomized, factorial, Latin-Square, randomized block, and covariance designs.

Statistical power Ability of a sample to protect against the type II error (beta risk).

Statistical regression The tendency with repeated measures for scores to regress to the population mean of the group.

Stepwise regression Multiple regression analysis in which the independent variable explaining the most variance is sequentially included one at a time.

Story completion A qualitative research technique where a respondent is presented with the beginning of a situational narrative and is asked to complete it.

Stratified sampling A probability sample where the population is broken into different strata or subgroups based on one or more characteristics and then a simple random sample is taken from each stratum of interest in the population.

Structured interview An interview in which a formal questionnaire has been developed and the questions asked in a prearranged order.

Stub-and-banner table A table that presents one dependent variable cross-tabulated by multiple independent variables.

Substantive significance An association that is statistically significant and of sufficient strength.

Sufficiency of information Degree of completeness and/or detail of information to allow a decision to be made.

Summated scale A rating scale constructed by adding scores from responses to a set of Likert scales with the purpose of placing respondents along an attitude continuum of interest. See also *Likert scale* and *multi-item scale*.

Surrogate information error Noncorrespondence of the information being sought and that required to solve the problem.

Survey A research method in which the information sought is obtained by asking questions of respondents.

Survey tracking and address books Online survey technology that uses imbedded codes to facilitate the identification and tracking of survey respondents and non-respondents.

Syndicated services Information collected and tabulated on a continuing basis by research organizations for purposes of sale to firms; data are made available to all who wish to subscribe. See *commercial data*.

Systematic error See *nonsampling error*.

Systematic sampling A probability sample where the population elements are ordered in some way and then after the first element is selected all others are chosen using a fixed interval.

Tabulation The process of sorting data into previously established categories, making initial counts of responses and using summarizing measures.

Technical report A research report that emphasizes the methods used and underlying assumptions, and presents the findings in a detailed manner.

Telephone interview Interviews that are conducted by telephone.

Telescoping A response error that occurs when a respondent reports an event happening at a time when it did not happen. It may be *forward* (report it happening more recently than it did) or *backward* (reporting it happening earlier than it did).

Testing effect The effect of a first measurement on the scores of a second measurement.

Test of independence A test of the significance of observed association involving twop or more variables.

Test-retest reliability The stability of response over time.

Thematic Apperception Test (TAT) A test consisting of one or more pictures or cartoons that depict an ambiguous situation relating to the subject being studied, and research subjects are asked to make up a story about what is happening, or the subject is asked to assume the role of a person in the situation and then describe what is happening and who the others in the scene are.

Third-person technique A projective qualitative research method in which a respondent is indirectly interviewed by asking for his or her view of what a neighbor or some other person would respond to the interview.

Thurstone Case V Scaling Based on the Thurstone's Law of Comparative Judgment, this method allows the construction of a unidimensional interval scale using responses from ordinal measurement methods, such as paired comparisons.

Time series design Data are obtained from the same sample (or population) at successive points in time.

Total study error Sampling error plus non-sampling error.

Treatment variable See *independent variable*.

Trend design Data are obtained from statistically matched samples drawn from the same population over time.

True experiment See *controlled experiment*.

***t* – test** A test of the difference in mans of two groups of respondents that focuses on sample means and variances.

Type I error The probability that one will incorrectly reject H_o, the null hypothesis of no difference, or any hypothesis.

Type II error The probability that one will incorrectly accept a null hypothesis, or any hypothesis.

Unlimited-response category scale A direct-judgment rating scale where the respondent is free to choose his/her own number or insert a tick mark along some line to represent his/her judgment about he magnitude of the stimulus relative to some reference points.

Unobtrusive measures Nonreactive measures of behavior, past and present.

Unsolicited information Information which may, in fact, exist within and be obtainable within the company, but which potential users do not know is available unless they happen to chance upon it.

Unstructured interview An interview in which there is no formal questionnaire and the questions may not be asked in a prearranged order.

Useful information Information which is accurate, current, sufficient, available, and relevant.

Validity of measurement The extent to which one measures what he or she believes is being measured.

VALS A syndicated segmentation scheme known as Values and Lifestyle segmentation which combines demographic, attitudinal, and psychographic data, according to pre-defined segments,

Variance A measure of dispersion, it is the mean of the squared deviation of individual measurements from the arithmetic mean of the distribution.

Variation in measurement Differences in individual scores within a set of measurements that may be due to the characteristic or property being measured (the true difference) and/or the measurement process itself.

Verbal measures Include spoken and written responses, including responses provided interactively with a personal computer.

Verbal rating scale A rating scale using a series of verbal options for rating an object.

Warranty form of interview A type of mail interview where the questions asked are included on the warranty card to be returned to the manufacturer.

Weighting data Procedures used to adjust the final sample so that the specific respondent subgroups of the sample are found in identical proportions to those found in the poulation.

Wilcoxon rank sum (T) test A test of the relationship between two sets of measurements from dependent samples in which the data are collected in matched pairs.

Wilks' lambda In discriminant analysis it is a multivariate measure of group differences over discriminating variables.

Word association test A series of stimulus words are presented to a respondent who is asked to answer with the first word that comes to mind after hearing each stimulus word.