MORGAN George A. and Griego V. ORLANDO. 1998. *Easy Use and Interpretation of SPSS for Windows: Answering Research Questions With Statistics*. Mahwah: Lawrence Erlbaum Associates.

CHAPTER 1

Research Problems, Approaches, and Questions

Research Problems

The research process begins with a problem. *What is a research problem?* Kerlinger (1986) formally describes a problem as "...an interrogative sentence or statement that asks: *What relation exists between two or more variables?*" (p. 16). Note that almost all research studies have more than two variables. Kerlinger suggests that prior to the problem statement "...the scientist will usually experience an obstacle to understanding, a vague unrest about observed and unobserved phenomena, a curiosity as to *why something is as it is*" (p. 11). Appendix A provides templates to help you phrase your research problem, and provides examples from the high school and beyond (HSB) data set.

Variables

A *variable* has one defining quality. *It must be able to vary or have different values.* For example, *gender* is a variable because it has two values, female or male. *Age* is a variable that has a large number of values. *Type of treatment/intervention* (or *type of curriculum*) is a variable if there is more than one treatment or a treatment and a control group. *Number of days to learn something or to recover from an ailment*, common measures of the effect of a treatment, are also variables. Similarly, *amount of mathematics knowledge* is a variable because it can vary from none to a lot. If a concept has one value in a particular study it is not a variable, e.g., ethnic group is not a variable if all participants are Caucasian.

Definition of a variable. We can define the term "variable" as a characteristic of the participants or situation of a given study that has different values in that study. In quantitative research, variables are defined operationally and are commonly divided into independent variables (active or attribute), dependent variables, and extraneous variables. Each of these topics will be dealt with in the following sections.

Operational definitions of variables. An operational definition describes or defines a variable in terms of the operations or techniques used to elicit or measure it. When quantitative researchers describe the variables in their study, they specify what they mean by demonstrating how they measured the variable. Demographic variables like age, gender, or ethnic group are usually measured simply by asking the participant to choose the appropriate category from a list. Types of treatment (or curriculum) are usually described/defined much more extensively so the reader can understand what the researcher meant by, for example, a cognitively enriching curriculum or sheltered work. Likewise, abstract concepts like mathematics knowledge, self-concept, or mathematics anxiety need to be defined operationally by spelling out in some detail how they were measured in a particular study. To do this, the investigator may provide sample questions, append the actual instrument, or provide a reference where more information can be found.

Independent Variables

Active independent variables. This first type of variable is often called a manipulated independent variable. A frequent goal of research is to investigate the effect of a particular intervention. An example might be the effect of a new kind of therapy compared to the traditional treatment. A second example might be the effect of a new teaching method, such as cooperative learning, on student performance. In the two examples provided above, the variable of interest was something that was *given to* the participants. Therefore, an *active independent variable* is a variable, such as a workshop, new curriculum, or other intervention, one level of which *can be given to a group of participants*, usually within a specified period of time *during the study*.

In traditional experimental research, independent variables are those that the *investigator can manipulate*; they presumably cause a change in some resulting behavior, attitude, or physiological measure of interest. An independent variable is considered to be manipulated or active when the investigator has the option to give one value to one group (experimental condition), and another value to another group (control condition).

However, there are many circumstances, especially in applied research, when we have an active independent variable but this variable *is not directly manipulated by the investigator*. Consider the situation where the investigator is interested in a new type of treatment. In order to carry out the study, it turns out that rehabilitation center *A* will be using that treatment. Rehabilitation center *B* will be using the traditional treatment. The investigator will compare the two centers to determine if one treatment works better than the other. Notice that the independent variable is active but has *not* been manipulated *by the investigator*.

Thus, active independent variables are *given* to the participants in the study but are not necessarily manipulated by the experimenter. They may be given by a clinic, school, or someone, other than the investigator. From the participants' point of view the situation was manipulated.

Attribute independent variables. Unlike some authors of research methods books, we do not restrict the term "independent variable" to those variables that are manipulated or active. We define an independent variable more broadly to include any predictors, antecedents, or *presumed* causes or influences under investigation in the study. Attributes of the participants as well as active independent variables fit within this definition. For the social sciences, education, and disciplines dealing with special needs populations, attribute independent variables are especially important. Type of disability or level of disability is often the major focus of a study. Disability certainly qualifies as a variable since it can take on different values even though they are not "given" in the study. For example, cerebral palsy is different from Down syndrome which is different from spina bifida, yet all are disabilities. Also, there are different levels of the same disability. People already have defining characteristics or *attributes* which place them into one of two or more categories. The different disabilities are already present when we begin our study. Thus, we are also interested in studying a class of variables that cannot be given during the study, even by other persons, schools, or clinics.

MORGAN George A. and Griego V. ORLANDO. 1998. *Easy Use and Interpretation of SPSS for Windows: Answering Research Questions With Statistics.* Mahwah: Lawrence Erlbaum Associates.

A variable which cannot be given, yet is a major focus of the study, is called an attribute independent variable (Kerlinger, 1986). In other words, the values of the independent variable are attributes of the persons or the environment that are not manipulated during the study. For example, *gender*, *age*, *ethnic group*, or *disability* are attributes of a person.

Other labels for the independent variable. SPSS uses a variety of terms such as **factor** (chapters 5, 15, 16, 17 and 18), **covariate** (chapter 13), and **grouping variable** (chapters 14, 15). In other cases (chapters 5, 9) SPSS does not make a distinction between the independent and dependent variable, just labeling them variables. Another common label for an attribute independent variable is a measured variable. However, we prefer attribute so it is not easily confused with the dependent variable, which is also measured. Sometimes variables such as gender or ethnic group are called moderator or mediating variables because they serve these functions; however, SPSS does not use these terms so we will not either in this book.

Type of independent variable and inferences about cause and effect. When we analyze data from a research study, the statistical analysis does not differentiate whether the independent variable is an active independent variable or an attribute independent variable. However, even though SPSS and most statistics books use the label independent variable for both active and attribute variables, there is a crucial difference in interpretation. A significant change or difference following manipulation of the active independent variable may reasonably lead the investigator to infer that the independent variable *caused* the change in the dependent variable.

However, a significant change or difference between or among values of an attribute independent variable should *not* lead one to the interpretation that the attribute independent variable caused the dependent variable to change. A major goal of scientific research is to be able to identify a causal relationship between two variables. For those in applied disciplines, the need to demonstrate that a given intervention or treatment causes change in behavior or performance is extremely important. Only the approaches that have an active independent variable (the randomized experimental and to a lesser extent the quasi-experimental) can be successful in providing data that allow one to infer that the independent variable caused the dependent variable.

Although studies with attribute independent variables are limited in what can be said about causation, they can lead to solid conclusions about the differences between groups and about associations between variables. Furthermore, they are the *only* available approach if the focus of your research is on attribute independent variables. The descriptive approach, as we define it, does not attempt to identify relationships. It focuses on describing variables.

As implied above, this distinction between active and attribute independent variables is important because terms such as *main effect* and *effect size* used by SPSS and most statistics books might lead one to believe that if you find a significant difference the independent variable *caused* the difference. These terms are misleading when the independent variable is an attribute.

Values of the independent variable. In defining a variable, we said that it must have more than one value. When describing the different categories of an independent variable, SPSS uses the word *values*. This does *not* necessarily imply that the values are ordered.[1] Suppose that an investigator is performing a study to investigate the effect of a treatment. One group of participants is assigned to the treatment group. A second group does not receive the treatment. The study could be conceptualized as having one independent variable (*treatment type)*, with two values or levels (treatment and no treatment). The independent variable in this example would be classified as an active independent variable. Instead, suppose the investigator was interested primarily in comparing two different treatments but decided to include a third no-treatment group as a control group in the study. The study still would be conceptualized as having one active independent variable (treatment type), but with three values (the two treatment conditions and the control condition). This variable could be diagrammed as follows:

Variable Label	Values	Value Labels
	1	= Treatment 1
Treatment type	2	= Treatment 2
	3	= No treatment (control)

As an additional example, consider gender, which is an attribute independent variable with two values, male and female. It could be diagrammed as follows:

	1	= Male
Gender		
	2	= Female

Note that in SPSS each variable is given a label; the values, which are numbers, may also have labels. It is especially important to know the value labels when the variable is nominal; i.e., when the values of the variable are just names and, thus, are not ordered.

Dependent Variables

The dependent variable is the presumed outcome or criterion. It is assumed to measure or assess the effect of the independent variable. Dependent variables are often test scores, ratings on questionnaires, readings from instruments (electrocardiogram, galvanic skin response, etc.), or measures of physical performance. When we discuss measurement in chapter 3, we are usually referring to the dependent variable. SPSS also uses a number of other terms for the dependent variable. The most common is **dependent list**, used in cases where you can do the same statistic several times, for a list of dependent variables. In discriminant analysis (chapter 13), the dependent variable is called the **grouping variable.** The term **test variable** is used in several of the chapters on *t* tests and analysis of variance.

---

[1] The terms categories, levels, groups, or samples are sometimes used interchangeably with the term values, especially in statistics books. Likewise the term factor is often used instead of independent variable.

Basic comparative approach. The comparative research approach differs from the experimental and quasi-experimental approaches because the investigator *cannot randomly assign participants* to groups and because there is *not an active independent variable*. Table 1.1 shows that, like experiments and quasi-experiments, comparative designs usually have a few levels or categories for the independent variable and make comparisons between groups. Studies that use the comparative approach examine the presumed effect of an *attribute independent variable*.

An example of the comparative approach is a study that compared two groups of children on a series of motor performance tests. The investigators attempted to determine whether the differences between the two groups were due to perceptual or motor processing problems. One group of children, who had motor handicaps, was compared to a second group of children who did not have motor problems. Notice that the independent variable in this study was an attribute independent variable with two levels, motor handicapped and not handicapped. Thus, it is not possible for the investigator to randomly assign participants to groups, or "give" the independent variable; the independent variable was not active. The independent variable had only two values

or categories so a statistical comparison between the groups would be performed. It is, of course, possible for comparisons to be made between three or more groups.[2]

Basic associational approach. Now, we would like to consider an approach to research where the independent variable is usually continuous or has several ordered categories, usually five or more. Suppose that the investigator is interested in the relationship between giftedness and self-perceived confidence in children. Assume that the dependent variable is a self-confidence scale for children. The independent variable is giftedness. If giftedness had been divided into high, average, and low groups (a few values or levels), we would have called the research approach comparative because the logical thing to do would be to compare the groups. However, in the typical associational approach, the independent variable is continuous or has at least five ordered levels or values.[3] All participants would be in a single group with two continuous variables--giftedness and self-concept. A correlation coefficient could be performed to determine the strength of the relationship between the two variables.

As implied above, it is somewhat arbitrary whether a study is considered to be comparative or associational. For example, a continuous variable such as age can always be divided into a small number of levels such as young and old. However, we make this distinction for two reasons. First, we think it is usually unwise to divide a variable with many ordered levels into a few because information is lost. For example, if the cut point for "old age" was 65, persons 66 and 96 would be lumped together as would persons 21 and 64. Second, different types of statistics are usually used with the two approaches (see Fig. 1.1). We think this distinction and the similar one made in the section on research questions will help you decide on an appropriate statistic, which we have found is one of the hardest parts of the research process for students.

Basic descriptive approach. This approach is different from the other four in that only one variable is considered at a time so that no relationships are made. Table 1.1 shows that this lack of comparisons or associations is what distinguishes this approach from the other four. Of course, the descriptive approach does not meet any of the other criteria such as random assignment of participants to groups.

Most research studies include some descriptive questions (at least to describe the sample), but do not stop there. It is rare these days for published quantitative research to be purely descriptive; we almost always study several variables and their relationships. However, political polls and consumer surveys are sometimes only interested in describing how voters *as a whole* react to issues or what products a group of consumers will buy. Exploratory studies of a new topic may just describe what people say or feel about that topic.

Most research books use a considerably broader definition for descriptive research. Some use the phrase "descriptive research" to include all research that is not randomized experimental or

---

[2] It is also possible to compare relatively large numbers of groups (e.g., 5 or 10) if one has enough participants that the group sizes are adequate, but this is atypical.

[3] It is possible, as we will see in chapters 7 and 8, to use the associational approach and statistics when one has fewer than five ordered values of the variables and even with unordered nominal variables, but this is not typical.

quasi-experimental. Others do not seem to have a clear definition, using descriptive almost as a synonym for exploratory or sometimes "correlational" research. We think it is clearer and less confusing to students to restrict the term descriptive research to questions and studies that use only *descriptive statistics*, such as averages, percentages, histograms, and frequency distributions, and do not test null hypotheses with inferential statistics.

Complex Research Approaches

It is important to note that most studies are more complex than implied by the above examples. In fact, almost all studies have more than one hypothesis or research question and may utilize more than one of the above approaches. It is common to find a study with one active independent variable (e.g., type of treatment) and one or more attribute independent variables (e.g., gender). This type of study combines the randomized experimental approach (if the participants were randomly assigned to groups) and the comparative approach. Most "survey" studies include both the associational and comparative approaches. As mentioned above, most studies also have some descriptive questions so it is common for published studies to use three or even more of the approaches.

Research Questions/Hypotheses

Next, we divide research questions into three broad types: *difference, associational,* and *descriptive*. For the difference type of question, we compare groups or values of the independent variable on their scores on the dependent variable. This type of question typically is used with the randomized experimental, quasi-experimental, and comparative approaches. For an associational question, we associate or relate the independent and dependent variables. Descriptive questions are not answered with inferential statistics; they merely describe or summarize data.

Basic Difference Versus Associational Research Questions or Hypotheses

Hypotheses are defined as *predictive statements about the relationship between variables*. Fig. 1.1 shows that both difference and associational questions/hypotheses have as a *general purpose* the exploration of relationships between variables: This similarity is in agreement with the statement by statisticians that all parametric inferential statistics are relational, and it is consistent with the notion that the distinction between the comparative and associational approach is somewhat arbitrary.[4] However, we believe that the distinction is educationally useful. Note that difference and associational questions differ in specific purpose and the kinds of statistics they use to answer the question.

[4] We use the term associational for this type of research question, approach, and statistics rather than relational or correlational to distinguish them from the *general purpose* of both difference and associational questions/hypotheses described above. Also we wanted to distinguish between correlation, as a specific statistical technique, and the broader types of approach, questions, and group of statistics.

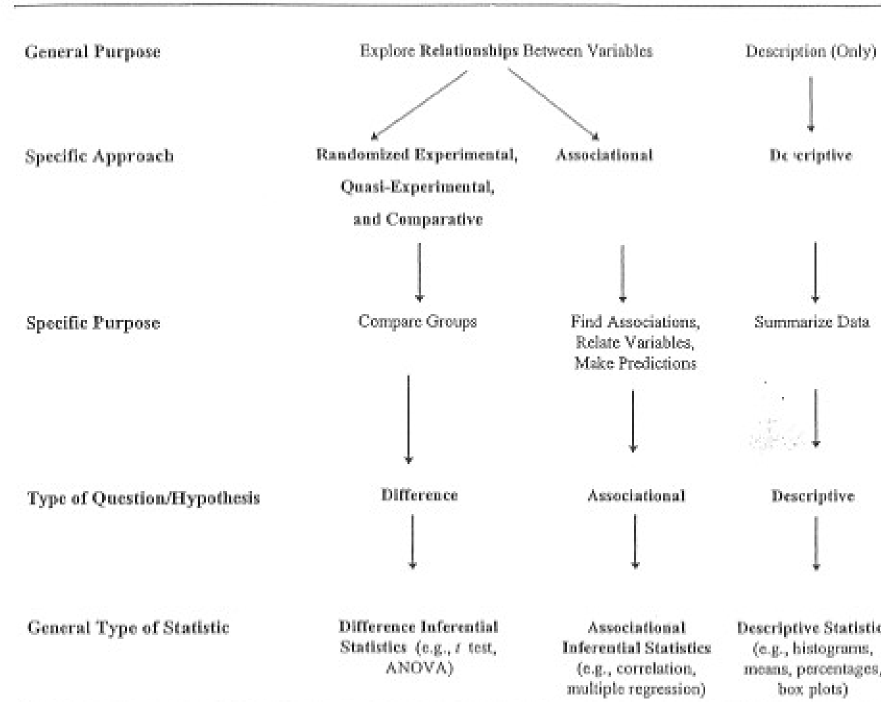

Fig. 1.1. Schematic diagram showing how the purpose, approach and type of research question correspond to the general type of statistic used in a study.

Table 1.2 provides the general format and one example of a basic *difference hypothesis* and of a basic *associational hypothesis*. Research questions are similar to hypotheses, but they are stated in question format. We think it is advisable to use the question format when one does not have a clear directional prediction and for the descriptive approach. More details and examples are given in Appendix A.

Table 1.2. *Examples of Basic Difference and Associational Hypotheses*

1. *Difference (group comparison) Hypothesis*

- For this type of hypothesis, the levels or values of the independent variable (e.g., gender) are used to divide the participants into groups (male and female) which are then compared to see if they differ in respect to the average scores on the dependent variable (e.g., empathy).
- An example of a directional research hypothesis is: Women will score higher than men on empathy scores. In other words, the average empathy scores of the women will be significantly higher than the average empathy scores for men.

2. *Associational (relational) Hypothesis*

- For this type of hypothesis, the scores on the independent variable (e.g., self-esteem) are associated with or related to the dependent variable (e.g., empathy). *It is often arbitrary which variable is considered the independent variable* but most researchers have an idea about what they think is the predictor (independent) and what is the outcome (dependent) variable.
- An example of a directional research hypothesis is: There will be a positive association (relation) between self-esteem scores and empathy scores. In other words, those persons who are high on self-esteem will tend to have high empathy, those with low self-esteem will tend also to have low empathy, and those in the middle on the independent variable will tend to be in the middle on the dependent variable.

Six Types of Research Questions

Table 1.3 expands our overview of research questions to include both basic and complex questions of each of the three types: *descriptive*, *difference*, and *associational*. The table also includes references to the tables in chapters 3 and 7, designed to help you select an appropriate statistic and examples of the types of statistics that we include under each of the six types of questions. Appendix A and the last section in this chapter provide examples of research questions for each of the six types. We use the terms basic and complex because the more common names, univariate and multivariate, are not used consistently in the literature.

Note that some complex descriptive statistics (e.g., a cross-tabulation table) could be tested for significance with inferential statistics; if they were so tested they would no longer be considered descriptive. We think that most qualitative/constructivist researchers ask complex descriptive questions because they consider more than one variable/concept at a time but do not use inferential/hypothesis testing statistics. Furthermore, complex descriptive statistics are used to check reliability (e.g., Cronbach's alpha) and to reduce the number of variables (e.g., factor analysis).

Table 1.3. *Summary of Types of Research Questions*

Type of Research Questions (Number of Variables)	Statistics (Example)
1) **Basic Descriptive Questions** – 1 variable	See Table 3.2 (mean, standard deviation, frequency distribution)
2) **Complex Descriptive Questions** – 2 or more variables, but no use of inferential statistics	(box plots, cross-tabulation tables, factor analysis, measures of reliability)
3) **Basic Difference Questions** – 1 independent and 1 dependent variable. Independent variable usually has a few values (ordered or not).	Table 7.1 (*t* test, one-way ANOVA)
4) **Complex Difference Question** – 3 or more variables. Usually 2 or a few independent variables and 1 or more dependent variables considered together.	Table 7.3 (factorial ANOVA, MANOVA)
5) **Basic Associational Questions** – 1 independent variable and 1 dependent variable. Usually at least 5 ordered values for both variables. Often they are continuous.	Table 7.2 (correlation tested for significance)
6) **Complex Associational Questions** – 2 or more independent variables and 1 or more dependent variables. Usually 5+ ordered values for all variables but some or all can be dichotomous variables.	Table 7.4 (multiple regression)

Difference versus associational inferential statistics. We think it is educationally useful, although not common in statistics books, to divide inferential statistics into two types corresponding to difference and associational hypotheses/questions. Difference inferential statistics are used for the experimental, quasi-experimental, and comparative approaches, which test for *differences between groups* (e.g., using analysis of variance). Associational inferential statistics test for *associations or relationships between variables* and use correlation or multiple regression analysis.[5] We will utilize this contrast between difference and associational inferential statistics in chapter 7 and later in this book.

[5] We realize that all parametric inferential statistics are relational so this dichotomy of using one type of data analysis procedure to test for differences (when there are a few values or levels of the independent variables) and another type of data analysis procedure to test for associations (when there are continuous independent variables) is somewhat artificial. Both continuous and categorical independent variables can be used in a general linear model