Antonius Rachard. 2003. *Interpreting Quantitative Data with SPSS*. London: Sage.

1 THE BASIC LANGUAGE OF STATISTICS

This chapter is an introduction to statistics and to quantitative methods. It explains the basic language used in statistics, the notion of a data file, the distinction between descriptive and inferential statistics, and the basic concepts of statistics and quantitative methods.

After studying this chapter, the student should know:

- the basic vocabulary of statistics and of quantitative methods;
- what an electronic data file looks like, and how to identify cases and variables;
- the different uses of the term 'statistics';
- the basic definition of descriptive and inferential statistics;
- the type of variables and of measurement scales;
- how concepts are operationalized with the help of indicators.

Introduction: Social Sciences and Quantitative Methods

Social sciences aim at studying social and human phenomena as rigorously as possible. This involves describing some aspect of the social reality, analyzing it to see whether logical links can be established between its various parts, and, whenever possible, predicting future outcomes.

The general objective of such studies is to understand the patterns of individual or collective behavior, the constraints that affect it, the causes and explanations that can help us understand our societies and ourselves better and predict the consequences of certain situations. Such studies are never entirely objective, as they are inevitably based on certain assumptions and beliefs that cannot be demonstrated. Our perceptions of social phenomena are themselves subjective to a large extent, as they depend on the *meanings* we attribute to what we observe. Thus, we *interpret* social and human phenomena much more than we describe them, but we try to make that interpretation as objective as possible.

Some of the phenomena we observe can be *quantified*, which means that we can translate into numbers some aspects of our observations. For instance, we can quantify

population change: we can count how many babies are born every year in a given country, how many people die, and how many people migrate in or out of the country. Such figures allow us to estimate the present size of the population, and maybe even to predict how this size is going to change in the near future. We can quantify psychological phenomena such as the degree of stress or the rapidity of response to a stimulus; demographic phenomena such as population sizes or sex ratios (the ratio of men to women); geographic phenomena such as the average amount of rain over a year or over a month; economic phenomena such as the unemployment rate; we can also quantify social phenomena such as the changing patterns of marriage or of unions, and so on.

When a social or human phenomenon is quantified in an appropriate way, we can ground our analysis of it on figures, or statistics. This allows us to describe the phenomenon with some accuracy, to establish whether there are links between some of the variables, and even to predict the evolution of the phenomenon. If the observations have been conducted on a sample (that is, a group of people smaller than the whole population), we may even be able to generalize to the whole population what we have found on a sample.

When we observe a social or human phenomenon in a systematic, scientific way, the information we gather about it is referred to as *data*. In other words, **data** is information that is collected in a systematic way, and organized and recorded in such a way that it can be interpreted correctly. Data is not collected haphazardly, but in response to some questions that the researchers would like to answer. Sometimes, we collect information (that is, data) about a character or a quality, such as the mother tongue of a person. Sometimes, the data is something measurable with numbers, such as a person's age. In both cases, we can treat this data numerically: for instance we can count how many people speak a certain language, or we can find the average age of a group of people. The procedures and techniques used to analyze data numerically are called *quantitative methods*. In other words, **quantitative methods** are procedures and techniques used to analyze data numerically; they include a study of the valid methods used for collecting data in the first place, as well as a discussion of the limits of validity of any given procedure (that is, an understanding of the situations when a given procedure yields valid results), and of the ways the results are to be interpreted.

This book constitutes an introduction to quantitative methods for the social sciences. The first chapter covers the basic vocabulary of quantitative methods. This vocabulary should be mastered by the student if the remainder of the book is to be understood properly.

Data Files

The first object of analysis in quantitative methods is a **data file**, that is, a set of pieces of information written down in a codified way. Figure 1.1 illustrates what an **electronic data file** looks like when we open it with the SPSS program.

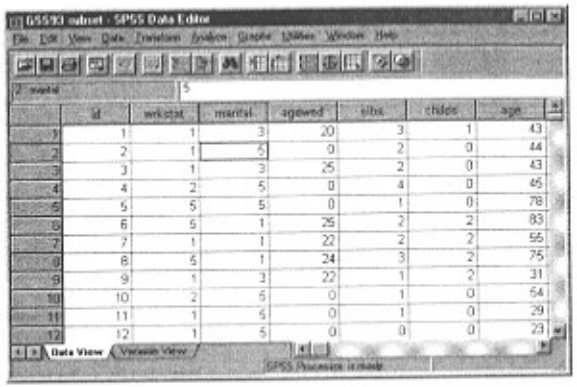

Figure 1.1 **The Data window in SPSS version 10.1. © SPSS. Reprinted with permission.**

This data file was created by the statistical software package SPSS Version 10.1, which will be used in this course. The first lab in the second part of this manual will introduce you to SPSS, which stands for *Statistical Package for the Social Sciences*. On the top of the window, you can read the name of the data file: **GSS93 subset**. This stands for **Subset of the General Social Survey**, a survey conducted in the USA in 1993.

When we open an SPSS data file, two views can be displayed: the Data View or the Variable View. Both views are part of the same file, and one can switch from one view to the other by clicking on the tab at the bottom left of the window.

The Data View: The information in this data view is organized in rows and columns. Each row refers to a **case**, that is, all the information pertaining to one individual. Each column refers to a **variable**, that is, a character or quality that was measured in this survey. For instance, the second column is a variable called **wrkstat**, and the third is a variable called **marital**.

But what are the meanings of all these numbers and words? A data file must be accompanied by information that allows a reader to interpret (that is, understand) the meanings of the various elements in it. This information constitutes the **codebook**. In SPSS, we can find the information of the codebook by clicking the word **Variables...** under the **Utilities** menu. We get a window listing all the variables contained in this data file. By clicking once on a variable, we see the information pertaining to this variable:

- the short name that stands on the top of the column;
- what the name stands for (the **label** of the variable);
- the numerical type of the variable (that is, how many digits are used, and whether it includes decimals);
- other technical information to be explained later;
- and the **Value Labels**, that is, what each number appearing in the data sheet stands for.

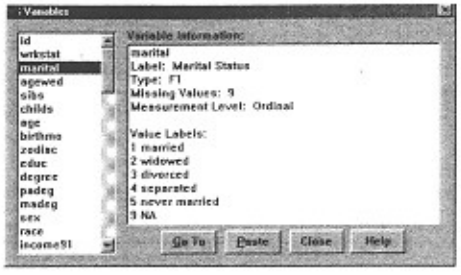

Figure 1.2 **The Variables window in SPSS. The codes and value labels of the variable Marital Status are shown**

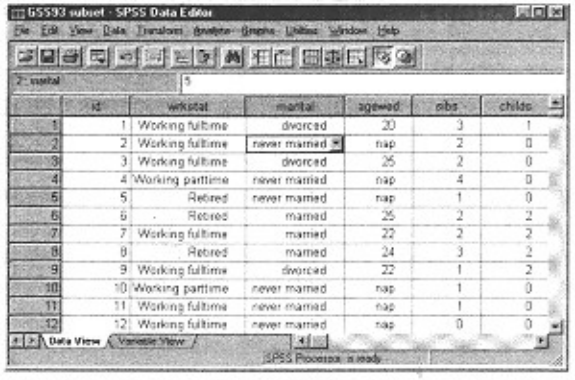

Figure 1.3 **The Data View window in SPSS when the Show Labels command is ticked in the View menu. The value labels are displayed rather than the codes**

Figure 1.2 shows the codes used for the variable Marital Status.

You may have noticed that:

1 stands for married
2 stands for widowed
3 stands for divorced
etc.

The numbers 1, 2, 3, etc. are the **codes,** and the terms married, widowed, divorced, etc. are the **value labels** that correspond to the various codes. The name *marital*, which appears at the top of the column, is the **variable name**. *Marital Status* is the **variable label**: it is a usually longer, detailed name for the variable. When we print tables or graphs, it is the variable labels and the value labels that are printed.

Figure 1.4 **The Variable View window in SPSS. The variables are listed in the rows, and their properties are displayed**

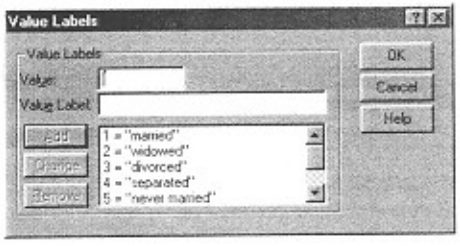

Figure 1.5 **The Value Labels window in SPSS. In this window it is possible to add new codes and their corresponding value labels, or to modify or delete existing ones**

There is a way of showing the value labels instead of the codes. This is done by clicking **Value Labels** under the **View** menu. The Data View window looks now as shown in Figure 1.3.

We can see that case number 4, for example, is a person who works part time, and who has never been married. To understand the precise meaning of the numbers written in the other cells, we should first read the variable information found in the codebook for each of the variables.

In version 10.0 and version 11 of SPSS, you can read the information pertaining to the variables in the Variable View. By clicking on the tab for Variable View, you get the window shown in Figure 1.4.

In the Variable View, no data is shown. You can see, however, all the information pertaining to the variables themselves, each variable being represented by a line. The various variable names are listed in the first column, and each is followed by information about the corresponding variable: the way it is measured and recorded, its full name, the values and their codes, etc. All these terms will be explained in detail later on. The label, that is, the long name of the variable **marital**, is **Marital Status**. By clicking on the **Values** cell for the variable **marital**, the window shown in Figure 1.5 pops up.

We can see again the meanings of the codes used to designate the various marital statuses. We can now raise a number of questions: How did we come up with this data? What are the rules for obtaining reliable data that can be interpreted easily? How can we analyze this data? Table 1.1 includes a systematic list of such questions. The answers to these questions will be found in the various chapters and sections of this manual.

Table 1.1 **Some questions that arise when we want to use quantitative methods**

Questions	Chapters
How did we come up with this data? What are the questions we are trying to answer? What is the place of quantitative analysis in social research, and how does it link up with the qualitative questions we may want to ask? What is the scientific way of defining concepts and operationalizing them?	1. The Basic Language of Statistics
How do we conduct social research in a scientific way? What procedures should we follow to ensure that results are scientific? What are the basic types of research designs? How do we go about collecting the data?	2. The Research Process
Once collected, the data must be organized and described. How do we do that? When we summarize the data what are the characteristics that we focus on? What kind of information is lost? What are the most common types of shapes and distributions we encounter?	3. Univariate Descriptive Statistics 5. Normal Distributions
What are the procedures for selecting a sample? Are some of them better than others?	6. Sampling Designs
Some institutions collect and publish a lot of social data. Where can we find it? How do we use it?	7. Statistical Databases
Sometimes we notice coincidences in the data: for instance, those who have a higher income tend to behave differently on some social variables than those who do not. Is there a way of describing such relationships between variables, and drawing their significance?	8. Statistical Association
Sometimes the data comes from a sample, that is, a part of the population, and not the whole population. Can we generalize our conclusions to the whole population on the basis of the data collected on a sample? How can this be done? Is it precise? What are the risks that our conclusions are wrong?	9. Statistical Inference: Estimation 10. Statistical Inference: Hypothesis Testing

The Discipline of Statistics

The term *statistics* is used in two different meanings: it can refer to the *discipline* of statistics, or it can refer to the *actual data* that has been collected.

As a scientific discipline, the object of *statistics* is the numerical treatment of data that pertain to a large quantity of individuals or a large quantity of objects. It includes a general, theoretical aspect which is very mathematical, but it can also include the study of the concrete problems that are raised when we apply the theoretical methods to specific disciplines. The term *quantitative methods* is used to refer to methods and techniques of statistics which are applied to concrete problems. Thus, the difference between statistics and quantitative methods is that the latter include practical concerns such as finding solutions to the problems arising from the collection of real data, and interpreting the numerical results as they relate to concrete situations. For instance, proving that the mean (or average) of a set of values has certain mathematical properties is part of statistics. Deciding that the mean is an appropriate measure to use in a given situation is part of quantitative methods. But the line between statistics and quantitative methods is fuzzy, and the two terms are sometimes used interchangeably. In practice, the term statistics is often used to mean quantitative methods, and we will use it in that way too.

The term *statistics* has also a different meaning, and it is used to refer to the actual data that has been obtained by statistical methods. Thus, we will say for instance that the latest statistics published by the Ministry of Labor indicate a decrease in unemployment. In that last sentence, the word *statistics* was used to refer to data published by the Ministry.

Populations, Samples, and Units

Three basic terms must be defined to explain the subject matter of the discipline of statistics:

- unit (or element, or case),
- population, and
- sample.

A **unit** (sometimes called **element**, or **case**) is the smallest object of study. If we are conducting a study on individuals, a unit is an individual. If our study were about the health system (we may want to know, for instance, whether certain hospitals are more efficient than others), a unit for such a study would be a hospital, not a person.

A **population** is the collection of all units that we wish to consider. If our study is about the hospitals in Quebec, the population will consist of all hospitals in Quebec. Sometimes, the term *universe* is used to refer to the set of all individuals under consideration, but we will not use it in this manual.

Most of the time, we cannot afford to study each and every unit in a population, due to the impossibility of doing so or to considerations of time and cost. In this case, we study a smaller group of units, called a **sample**. Thus, a sample is *any* subset (or subgroup) of our population.

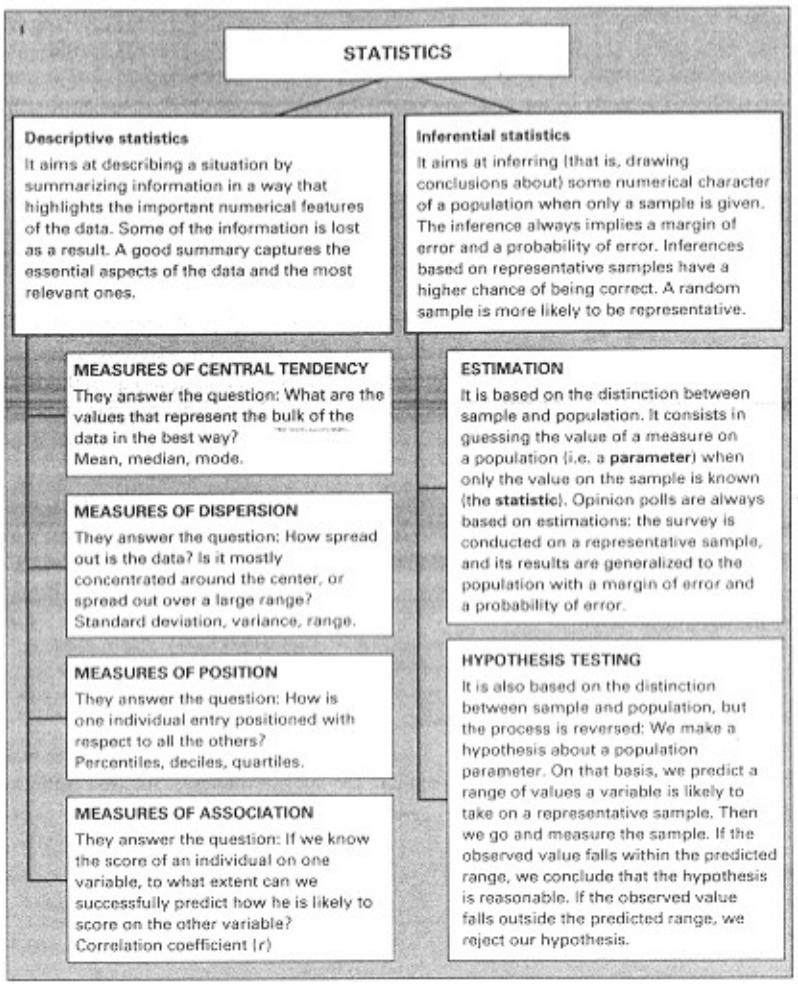

Figure 1.6 **The discipline of statistics and its two branches, descriptive statistics and inferential statistics**

The distinction between sample and population is absolutely fundamental. Whenever you are doing a computation, or making any statement, it must be clear in your mind whether you are talking about a sample (a group of units generally smaller than the population) or about the whole population.

The discipline of statistics includes two main branches:

- descriptive statistics, and
- inferential statistics.

The following paragraphs explain what each branch is about. Refer also to Figure 1.6. Some of the terms used in the diagram may not be clear for now, but they will be explained as we progress.

Descriptive Statistics

The methods and techniques of descriptive statistics aim at summarizing large quantities of data by a few numbers, in a way that highlights the most important numerical features of the data. For instance, if you say that your average GPA (grade point average) in secondary schooling is 3.62, you are giving only one number that gives a pretty good idea of your performance during all your secondary schooling. If you also say that the *standard deviation* (this term will be explained later on) of your grades is 0.02, you are saying that your marks are very consistent across the various courses. A standard deviation of 0.1 would indicate a variability that is 5 times bigger, as we will learn later on. You do not need to give the detailed list of your marks in every exam of every course: the average GPA is a sufficient measure in many circumstances. However, the average can sometimes be misleading. When is the average misleading? Can we complement it by other measures that would help us have a better idea of the features of the data we are summarizing? Such questions are part of descriptive statistics.

Descriptive statistics include measures of central tendency, measures of dispersion, measures of position, and measures of association. They also include a description of the general shape of the distribution of the data. These terms will be explained in the corresponding chapters.

Inferential Statistics

Inferential statistics aim at generalizing a measure taken on a small number of cases that have been observed, to a larger set of cases that have not been observed. Using the terms explained above, we could reformulate this aim, and say that inferential statistics aim at generalizing observations made on a sample to a whole population. For instance, when pre-election polls are conducted, only one or two thousand individuals are questioned, and on the basis of their answers, the polling agency draws conclusions about the voting intentions of the whole population. Such conclusions are not very precise, and there is always a risk that they are completely wrong. More importantly, the sample used to draw such conclusions must be a *representative sample*, that is, a sample in which all the relevant qualities of the population are adequately represented. How can we ensure that a sample is representative? Well, we can't. We can only increase our chances of selecting a representative sample if we select it randomly. We will devote a chapter to sampling methods.

Inferential statistics include estimation and hypothesis testing, two techniques that will be studied in Chapters 9 and 10.

A few more terms must be defined to be able to go further in our study. We need to talk a little about variables and their types.

Variables and Measurement

A variable is a characteristic or quality that is observed, measured, and recorded in a data file (generally, in a single column). If you need to keep track of the country of birth of the individuals in your population, you will include in your study a variable called *Country of birth*. You may also want to keep track of the nationality of the individuals: you will then have another variable called *Nationality*. The two variables are distinct, since some people may carry the nationality of a country other than the one they were born in. Here are some examples of variables used widely in social sciences:

Socio-demographic variables
Age
Sex
Religion
Level of education
Highest degree obtained
Marital status
Country of birth
Nationality
Mother tongue

Psychological variables
Level of anxiety
Stimulus response time
Score obtained in a personality test
Score obtained in an aptitude test

Economic variables
Working status
Income
Value of individual assets
Average number of hours of work per week

Variables that refer to units other than the individual
Number of hospitals in a country
Percentage of people who can read
Percentage of people who completed high school
Total population
Birth rate
Fertility rate
Number of teachers per 1000 people
Number of doctors per 10,000 people
Population growth
Predominant religion

You may have noticed that some of these variables refer to qualities (such as mother tongue) and others refer to quantities, such as the total population of a country. In fact, we can distinguish two basic **types of variables**:

1. **quantitative variables;**
2. **qualitative variables.**

Quantitative variables are characteristics or features that are best expressed by numerical values, such as the age of a person, the number of people in a household, the size of a building, or the annual sales of a product. Qualitative variables are characteristics or qualities that are not numerical, such as mother tongue, or country of origin. The scores of the individuals of a population on the various variables are called the **values** of that variable.

Example

Suppose you have the information shown in Table 1.2 about five students in your college.

Table 1.2 **Examples of qualitative and quantitative variables**

Name	Age	Program of Study	Grade Point Average
John	19	Social Science	3.78
Mary	17	Pure and Applied Science	3.89
Peter	18	Commerce	3.67
Colette	19	Office Systems Technology	3.90
Suzie	20	Graphic Design	3.82

There are three variables: **Age** (quantitative), **Program of Study** (qualitative), and **Grade Point Average** (quantitative).

The values, or scores, taken by the individuals for the variable **Age** are 17, 18, 19 (twice), and 20. The values taken for the variable **Program of Study** are Social Science, Pure and Applied Science, Commerce, Office Systems Technology, and Graphic Design. Qualitative variables are sometimes referred to as **categorical** variables because they consist of categories in which the population can be classified. For instance, we can classify all students in a college into categories according to the program of study they are in.

Careful attention must be given to the way observations pertaining to a variable are *recorded*. We must find a system for recording the data that is very clear, and that can be interpreted without any ambiguity. Consider, for instance, the following characteristics: age, rank in the family, and mother tongue. The first characteristic is a quantity; the second is a rank, and the third is a quality. The systems used to record our observations about these characteristics will be organized into three **levels of measurement**:

- measurement at the **nominal** level;
- measurement at the **ordinal** level; and
- measurement at the **numerical scale** level.

Each level of measurement allows us to perform certain statistical operations, and not others.

The **nominal level of measurement** is used to measure qualitative variables. It is the simplest system for writing down our observations: when we want to measure a characteristic at the nominal level, we establish a number of categories in such a way that each observation falls into one and only one of these categories. For example, if you want to write down your observations about mother tongue in the Canadian context, you may have the following categories:

- English,
- French,
- Native, and
- Other.

Depending on the subject of your research, you may have more categories to include other languages, or you may want to make a provision for those who have two mother tongues.

It is important to note that when a variable is measured at the nominal level, the categories must be

- exhaustive, and
- mutually exclusive.

The categories are said to be **exhaustive** when they include the whole range of possible observations, that is, they exhaust all the possibilities. That means that every one of the observations can fit in one of the available categories. The categories are said to be **mutually exclusive** if they are not overlapping: every observation fits in only one category. These two properties ensure that the system used to write down the observations is clear and complete, and that there are no ambiguities when recording the observations or when reading the data file. Table 1.3 displays examples of measurements made at the nominal level.

Qualitative variables must be measured at the nominal level.

The **ordinal level of measurement** is used when the observations are organized in categories that are *ranked*, or *ordered*. We can say that one category precedes another, but we cannot say by how much exactly (or if we can, we do not keep that information). Here too the categories must be exhaustive and mutually exclusive, but in addition you must be able to compare any two categories, and say which one precedes the other (or is bigger, or better, etc.). Table 1.4 displays examples of variables measured at the ordinal level.

Table 1.3 **Examples of variables measured at the nominal level**

Variable	Categories used
Sex	Male Female
Place of birth	The country where the survey is counducted Abroad
Work status	Working full-time Working part-time Temporarily out of work Unemployed Retired Housekeeper Other

Table 1.4 **Examples of variables measured at the ordinal level**

Variable	Ranked Categories
Rating of a restaurant	Excellent Very good Acceptable Poor Very poor
Rank among siblings	First child Second child etc.
Income	High Medium Low

The scale used to write down an ordinal variable is often referred to as a **Likert scale**. It usually has a limited number of ranked categories: anywhere from three to seven categories, sometimes more. For instance, if people are asked to rate a service as:

- ❑ Excellent
- ❑ Very good
- ❑ Good
- ❑ Poor
- ❑ Very poor,

the proposed answers constitute a five-level Likert scale.

Another example of a Likert scale, this time with four levels, is provided by the situations where a statement is given, and respondents are asked to say whether they:

- ❑ Totally agree
- ❑ Agree
- ❑ Disagree
- ❑ Totally disagree.

A variable measured at the ordinal level could be either qualitative or quantitative. In Table 1.4, the variable **Income** is quantitative, and the variable **Rating of a Restaurant** is qualitative, but they are both measured at the ordinal level. For a variable measured at the ordinal level, we can say that one value precedes another, but we cannot give an exact numerical value for the difference between them. For instance, if we know that a respondent is the first child and the other is the second child in the same family, we do not keep track of the age difference between them. It could be one year in one case and five years in another case, but the values recorded under this variable do not give us this information: they only give us the rank.

When recording information about categorical variables, the information is usually *coded*. *Coding* is the operation by which we determine the categories that will be recorded, and the codes used to refer to them. For instance, if the variable is **Sex**, and the two possible answers are:

Male
Female,

we usually code this variable as

1 Male
2 Female.

The numbers 1 and 2 are the *codes*, and the categories Male and Female are the *values* of the variable.

When coding a variable, a code must be given to the cases where no answer has been provided by the respondent, or when the respondent refuses to answer (if the answer is judged too personal or confidential, such as the exact income of a person). We refer to these answers as *missing values* and we give them different codes. Lab 9 explains how to handle them in SPSS.

Finally, some variables are measured by a **numerical scale**. Every observation is measured against the scale and assigned a numerical value, which measures a quantity. These variables are said to be **quantitative**. Table 1.5 displays examples of numerical scale variables.

Table 1.5 **Examples of variables measured at the numerical scale level**

Variable	Numerical Scale
Annual income	In dollars, without decimals (no cents)
Age	In years, with no fractions
Age	In years, with one decimal for fractions of a year
Temperature	In degrees Celsius
Time	In years. A starting point must be specified
Annual income	In dollars, to the nearest thousand

Notice that the same variable can be measured by different scales, as shown in the examples above. So, when we use a numerical scale, we must determine the units used (for instance years or months), and the number of decimals used.

Numerical scales are sometimes subdivided into **interval scales** and **ratio scales**, depending on whether there is an absolute zero to the scale or not. Thus, *temperature* and *time* are measured by interval scales, whereas *age* and *number of children* are each measured by a ratio scale. However, this distinction will not be relevant for most of what we are doing in this course, and we will simply use the term **numerical scale** to talk about this level of measurement. The program SPSS that we are going to use simply uses the term **scale** to refer to such variables.

Most statistical software packages include more specific ways of writing down the observations pertaining to a numerical scale. For instance, SPSS will offer the possibility of specifying that the variable is a currency, or a date.

Moreover, it is also possible to group the values of a quantitative variable into **classes**. Thus, when observing the variable *age*, we can write down the exact age of a person in years, or we can simply write the age group the person falls in, as is done in the following example:

- 18 to 30 years
- 31 to 40 years
- 41 to 50 years
- 50 to 60 years
- Over 60.

When we group a variable such as *age* into a small number of categories as we have just done, we must code the categories as we do for categorical variables. For example,

1 would stand for the category 18 to 30 years
2 would stand for the category 31 to 40 years
etc.

In such situations, we cannot perform the same statistical operations that we do when the values are not grouped. For instance, the mean, or average of the variable *age* is best calculated when the ages are *not* grouped. When we group the values, it is because we want to know the relative importance (that is, the frequency, in percentage) of one group as compared to the others. The information that 50% of the population is under 20 years old in some developing countries is obtained by grouping the ages into *20 years old or less* and *more than 20 years old*. When we collect the data, it is always better to collect it in actual years, since we can easily group it later on in the data file with the help of a statistical software package. In this case, a new column is added to the data file, and it contains the grouped data of the quantitative variable. For example, in the **GSS93 subset** data file that we use in the SPSS labs, you will find two variables for *age*: one is called **age**, and the other one is called **agecat4**. The latter is calculated from the former, by grouping individuals into four age groups. In the column of **agecat4**, the specific age of an individual is not recorded: only the age group of the individual is recorded.

Finally, numerical scales can be either **continuous** or **discrete**. A scale is said to be *continuous* if the observations can theoretically take any value over a certain range, including fractions of a unit. For instance, age, weight, length are continuous variables because they are not limited to specific values, and they can take any value within a certain range. A variable is said to be *discrete* if it can take only a limited number of possible values, but not values in between. For instance, the variable *Number of children* is measured by a discrete scale because it can only be equal to a whole number: 0, 1, 2, etc.

Importance of the Level of Measurement

The level of measurement used for a variable depends on whether it is qualitative or quantitative.

Qualitative variables must be measured at the nominal or ordinal level. They cannot be measured at the numerical scale level, even when their categories are coded with numbers. For instance, as shown above, we usually code the variable *Sex* as follows:

1 Male
2 Female.

In this case, **the numbers 1 and 2 have no numerical value**. They are simply codes. It is shorter to write 1 than *male*, and we could have assigned the numbers differently. If you ask SPSS to compute the mean (or average) for a variable coded in this way, you *will* get a numerical answer. But you must always keep in mind that such a numerical answer is *totally meaningless* because the level of measurement

of that variable is nominal. The numbers used to record the information are simply codes.

Quantitative variables are usually measured by a numerical scale, but they could be measured at the ordinal level also. For instance, if you have the annual income of an individual, you may treat it as a numerical scale, but you could also group the values into Low, Medium and High income and treat the variable at the ordinal level.

When you perform a statistical analysis of data, it is very important to pay attention to the level of measurement of each variable. Some statistical computations are appropriate only to a given level of measurement, and should not be performed if the variable is measured at a different level.

Concepts, Dimensions, and Indicators

We often want to observe social phenomena that are too abstract and complex to be expressed by a single variable. Suppose for instance that we want to observe and measure the degree of *religious inclination* (or the tendency of a person towards religion) in a given social group. Religious inclination can be manifested in many ways: people may have or not have certain *beliefs* about their religion; they may also perform or not certain *rituals* such as attending religious services, fasting, praying, etc.; they may also *seek the advice of the religious leadership* on important decisions, or ignore such leadership; finally, they may seek to look at everything from the point of view of religion, and *apply the teachings* of their religion in their daily lives, or ignore them. All these aspects are not found all the time in all individuals. Some individuals may have strong beliefs, while avoiding the religious services. Other may attend all services while being skeptical about some of the religious dogma. The way to handle this complexity is to subdivide the concept of *religious inclination* into dimensions, which are themselves measured by several indicators. If we were to study religious inclination in the Catholic religion, we would get a set of dimensions and indicators that would look as in Table 1.6 (we are simplifying the issues a little, of course).

The items listed on the right-hand side of Table 1.6 are **indicators** of the concept of *religious inclination*. None of them, taken alone, is a measure of religious inclination, but each of them constitutes *one* aspect of it. Indicators that are seen as similar are grouped together to form one *dimension* of the concept. And finally the various dimensions, taken together, capture the concept as a whole. This way of breaking down a complex concept into dimensions and indicators is called the **operationalization** of the concept. As an illustration, we may want to see how economists operationalize the concept of *cost of living*. They estimate the average cost of most of the standard expenses a family of four is expected to incur. The various expenses are divided into main dimensions such as food, housing, transportation, education, and leisure. Each dimension is then subdivided into smaller dimensions; themselves subdivided further until indicators are reached. For instance, food is

Table 1.6 **Example of how a concept can be broken down into dimensions and indicators**

Concept	Dimensions	Indicators
	I. Beliefs	Belief in God Belief in the Holy Trinity Belief in the main dogma etc.
RELIGIOUS INCLINATION	II. Rituals	Attendance of services Performing prayers Baptizing children etc.
	III. Guidance	Consulting the priest about important decisions Consulting the official opinions of the church on certain issues such as birth control etc.
	IV. Daily life	Being kind and generous to people Not cheating others in commercial transactions etc.

broken down as: meat, vegetables, milk products, etc., themselves subdivided into specific items such as: tomatoes, lettuce, etc. Finally, for each of these indicators, the increase or decrease in the cost of living is measured against the corresponding cost in some year, called the base year. By combining these indicators, economists are able to measure how the cost of living has changed, on the average, for a family of four.

The way a concept is broken down, or operationalized, into dimensions and indicators depends on the theoretical framework adopted for a study. Researchers may not agree on how to operationalize a concept, and you will find in the literature different studies that operationalize concepts in completely different ways, because they rely on different theoretical frameworks.

Summary

Quantitative methods are procedures and techniques for collecting, organizing, describing, analyzing, and interpreting data. In this chapter we have learned the basic vocabulary used to talk about quantitative methods. Data is organized into electronic data files with the help of statistical packages. A data file contains the values taken by a number of cases (which are the units of the population under study) over some variables. Every row represents a case, while every column represents a variable. The units in the data file usually form a sample, which is itself a subset of the whole population. Sometimes, the data file refers to the whole population.

The variables can be either qualitative or quantitative. The system used to record the information is called a measurement scale. There are three levels of measurement: nominal, ordinal and numerical (interval or ratio). The level of measurement of a variable will determine what statistical procedures can be performed, and what kind of graphs must be used to illustrate the data. When a concept is complex, it is not measured directly. It is usually broken down into dimensions and indicators, which are then combined to provide a single measure.

The statistical procedures themselves fall into two broad categories: descriptive statistics and inferential statistics. Descriptive statistical techniques aim at describing the data by summarizing it, while inferential statistical techniques aim at generalizing to a whole population what has been observed on a sample.

Keywords

Students should be able to define and explain *all* the following terms.

Data	Case	Continuous numerical scales
Data file	Unit	Discrete numerical scales
Quantitative methods	Sample	Codes
Variable	Population	Coding
Variable label	Level of measurement	Codebook
Value	Nominal level	Statistics (the two meanings)
Value label	Ordinal level	Descriptive statistics
Variable type	Numerical level (interval or ratio)	Inferential statistics
Quantitative variable	Exhaustive categories	Dimensions of a concept
Qualitative variable	Mutually exclusive categories	Indicators of a concept
	Likert scale	Operationalization of a concept

Suggestions for Further Reading

Blalock Jr., Hubert M. (1982) ***Conceptualization and Measurement in the Social Sciences***. London; Sage Publications.

Norusis, Marija J. (1998) ***SPSS 8.0 Guide to Data Analysis***. Upper Saddle River, NJ: Prentice Hall.

Rosenbaum, Sonia (1979) ***Quantitative Methods and Statistics: A Guide to Social Research***. Beverly Hills: Sage Publications.

Trudel, Robert and Antonius, Rachad (1991) ***Méthodes quantitatives appliquées aux sciences humaines***. Montréal: CEC.