2

THE LOGIC OF MEASUREMENT

Measurement is one of the most fundamental elements of science. In the case of social research, the task of measurement is typically one of characterizing individuals in terms of the issues under study. Thus, a study of voting will characterize respondents in terms of the candidate for whom they plan to vote, and a study of abortion will describe people in terms of their attitudes on that topic.

VALIDITY PROBLEMS

Validity is a term used casually in everyday language, but it has a precise meaning in social research. It describes an indicator of a concept. Most simply, an indicator is said to be valid if it really measures the concept it is intended to measure; it is invalid if it doesn't.

As a simple example, let's consider political orientation, ranging from very liberal to very conservative. For an example of a clearly valid measure of this concept, here's the way the General Social Survey (GSS) asked about it.

> POLVIEWS: We hear a lot of talk these days about liberals and conservatives. I'm going to show you a seven-point scale on which political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?
>
> 1. Extremely liberal
> 2. Liberal
> 3. Slightly liberal
> 4. Moderate, middle-of-the-road
> 5. Slightly conservative
> 6. Conservative
> 7. Extremely conservative

At the opposite extreme, a simple question about respondent gender would not be a valid measure of political orientation because political orientation and gender are different concepts. But now, let's consider another questionnaire item that does not come from the GSS. This item lies somewhere in between these two extremes of validity with regard to measuring political orientation.

> Question: Which of the two major political parties do you most identify with?
>
> 1. Democratic Party
> 2. Republican Party
> 3. Neither

This second item is another reasonable measure of political orientation. Moreover, it is related to the first because Democrats are, on the whole, more liberal than Republicans. On the other hand, there are conservative Democrats and liberal Republicans. If our purpose is to tap into the liberal–conservative dimension, the initial item that asks directly about political orientation is obviously a more valid indicator of the concept than is the item about political party.

This particular example offers us a clear choice as to the most valid indicator of the concept at hand, but matters are not always so clear-cut. If we were measuring levels of prejudice, for example, we could not simply ask, "How prejudiced are you?" both because no one is likely to admit to being prejudiced and because it's possible for someone to be prejudiced without knowing it. As we search for workable indicators of a concept such as prejudice, the matter of validity becomes something to which we must pay careful attention.

Validity is a concern not only when you collect and analyze your own data—a process known as ***primary research***—but also when you reanalyze data previously collected by someone else, as we do in this book. The process of reanalyzing someone else's data is referred to as ***secondary analysis***. Even if you can think of a survey question that would have captured your concept perfectly, the original researchers might not have asked it. Hence, you often need to use ingenuity in constructing measures that capture the quality in which you are interested. In the case of political orientation, for example, you might combine the responses to several questions—asking for attitudes about civil liberties, past voting behavior, political party identification, and so forth. We'll return to the use of multiple indicators shortly.

In large part, the question of validity is settled on *prima facie* grounds: We judge an indicator to be relatively valid or invalid on the face of it. It was on this basis that you had no trouble seeing that asking directly about political orientation was a valid indicator of that concept, whereas asking about a person's gender was definitely not a valid measure of political orientation. Later in this book, we'll explore some simple methodological techniques that are also used to test the validity of measures.

RELIABILITY PROBLEMS

Reliability is a different but equally important quality of measurement. ***Reliability*** refers to the quality of a measuring instrument that would cause it to report the same value in successive observations of a given case (provided the phenomenon being measured has not changed). For instance, if you step on a bathroom scale five times in a row and each time it gives you a different weight, the scale has a reliability problem. Conversely, if you step on a bathroom scale five times in a row and the scale gives you the same weight each time (even if the weight is wrong), the scale is reliable.

Similarly, if your statistics instructor administered the same test to you three times and each time you got a different score, even though your knowledge of statistics had not changed in the interim, the test would have a reliability problem. Conversely, if your instructor administered the same test three times and your score was the same each time, the test would be reliable.

In the context of survey research, reliability also concerns whether we can trust the answers that people give us even when their misstatements are honest ones. For instance, in medical research, some patients report in one survey that they have had a particular organ removed, only to indicate in subsequent surveys that they still have that organ.[1] Similarly, students of voting behavior regularly encounter individuals who claim on one survey that they did vote in the past presidential election and then claim in subsequent surveys that they either did not vote or do not remember whether they voted. As noted previously, these statements are often honest ones because it is difficult enough for most of us to recall what we did a few months ago, let alone several years ago.

Conceptually, the test of reliability is whether respondents would give the same answers repeatedly if the measurement could be made in such a way that (a) their situations had not changed (e.g., they hadn't had additional surgery to remove organs) and (b) they couldn't remember the answer they gave before.

As we suggested, empirically testing the reliability of an item requires multiple measures (e.g., your instructor must administer the statistics test at least two times in order to determine its reliability).

However, we can sometimes assess the reliability of a single item based on its practicality. For example, years ago, one of us was asked to help administer a survey to teenage drivers in California. Over researcher objections, the client insisted on asking the question, "How many miles have you driven?" and providing a space for the teenager to write in their response. Perhaps you can recognize the problem in this question by attempting to answer it yourself. Unless you have never driven an automobile, we doubt that you can report how many miles you have driven with any accuracy. In the survey mentioned, some teenagers reported driving hundreds of thousands of miles—highly questionable results for people who had only been driving for a few years. Unless they have been keeping a driving log their whole life, most people cannot report with any accuracy how many total miles they have driven.

In this situation, it would be better to provide respondents with a set of categories realistically reflecting the number of miles they are likely to have driven: fewer than 1,000 miles; 1,000 to 4,999 miles; 5,000 to 9,999 miles; and so on. Such a set of categories gives respondents a framework within which to place their own situations. Even though they still may not know exactly how much they have driven, there is a fair likelihood that the categories they choose will contain their correct answers. The success of this technique depends on our having a good idea in advance of what constitutes reasonable categories, perhaps as determined by previous research. As an alternative, we might ask respondents to volunteer the number of miles they have driven but limit the time period to something they are likely to remember. Thus, we might ask how many miles they drove during the preceding week or month.

DISTINGUISHING BETWEEN VALIDITY AND RELIABILITY

Perhaps the difference between validity and reliability can be seen most clearly in our previous example of weighing oneself on a bathroom scale. If you step on the scale repeatedly and it gives you a different weight each time, then the scale has a reliability problem. On the other hand, if the scale tells you that you weigh 125 pounds every time you step on it, then it's pretty reliable. However, if you actually weigh 225 pounds, then the scale, though reliable, has a problem in the validity department; it doesn't indicate your weight accurately.

Both validity and reliability are important in the analysis of data. If you are interested in learning why some people have deeply held religious beliefs and others do not, then asking people how often they attend church will be problematic. This question doesn't really provide a valid measure of the concept that interests you (religious belief) because anything you learn from asking the question will explain the causes of church attendance, not religious belief. Plus, if you asked people how many times they had attended church in the past year, any answers you received would probably not be reliable since, as with the driver survey mentioned previously, most people can't remember that far back. (It would be better, once again, to give them categories from which to choose.) Therefore, anything you might think you had learned about the causes of church attendance might be only a function of the errors people made in answering the question. You would have no assurance that another study would yield the same result, so your study would not be reliable.

MULTIPLE INDICATORS

Often, the solution to the problems just discussed lies in the creation of ***composite measures***, measurements of variables that combine two or more indicators of the variable, and the use of ***multiple indicators***, or several questions about the same concept. As a simple example, to measure the degree to which a sample group of church members holds Christian beliefs, you might ask the members of that group questions about several issues, each dealing with a particular belief:

- Belief in God
- Belief that Jesus was divine
- Belief in the existence of the Devil

- Belief in an afterlife: heaven and hell
- Belief in the literal truth of the Bible

The various answers respondents give to these questions could be used to create an overall measure of their Christian religious belief or lack thereof. In the simplest procedure, you could give respondents 1 point for each belief with which they agreed, allowing you to score them from 0 to 5 on the index. Notice that this is the same logic by which you may earn 1 point for each correct answer on an exam, with the total score being taken as an indication of how well you know the material.

Some social science concepts are implicitly multidimensional. Consider the concept of *social class*, for example. Typically, this term is used in reference to a combination of education, income, and occupation and, sometimes, dimensions such as social class identification and prestige. For the purpose of data analysis, this concept would therefore be measured through the use of multiple indicators. When it becomes appropriate in the analyses we are going to undertake together, we'll show you how to create and use some simple composite measures.

LEVELS OF MEASUREMENT

As we convert the concepts in our minds into empirical measurements in the form of variables, we sometimes have options as to those variables' level of statistical sophistication. Specifically, there are a number of different possibilities regarding the relationships among the categories constituting a variable. Therefore, in social research, we commonly speak of four ***levels of measurement***: nominal, ordinal, ratio, and interval.

Nominal Variables

Some variables simply distinguish different kinds of people. Gender is a good example of such a variable in the General Social Survey; the two categories of the variable, male and female, distinguish between men and women. Of course, such a treatment of this variable may not suit all research purposes, as transgender and intersex individuals are achieving more visibility, and there is more scholarly interest in them. The way we conceptualize and study gender will, no doubt, evolve in the coming years. Similarly, political party distinguishes Democrats from Republicans and from members of other parties; religious affiliation distinguishes Protestants from Catholics, Jews, and so forth. We refer to these variables as *nominal* because the categories constituting the variables are based on names.

Nominal variables simply name the different attributes constituting them. The attributes constituting a nominal variable (e.g., religious affiliation composed of Protestant, Catholic, Jewish, etc.) are simply different on their face. Republicans and Democrats are different from each other, as are Protestants and Catholics. In other cases, however, we can say more about the attributes making up variables.

Ordinal Variables

Many social scientific variables go a step beyond simply naming the different attributes constituting a variable. ***Ordinal variables*** arrange those attributes in some order: from low to high, from more to less, and so on. Whereas the nominal variable *religious affiliation* classifies people into different religious groups, *religiosity* might order them into groups such as very religious, somewhat religious, and not-at-all religious. And whereas the nominal variable *political party identification* simply distinguishes different groups (e.g., Democrats and Republicans), an ordinal measure of political philosophy might rank-order the very liberal, the somewhat liberal, the middle-of-the-road, the somewhat conservative, and the very conservative. Ordinal variables share the nominal variable's quality of distinguishing differences among people, but they add the quality of *rank-ordering* those differences.

At the same time, it is not meaningful to talk about the distances separating the attributes that make up an ordinal variable. For example, we have no basis for talking about the amount of liberalism separating the very liberal from the somewhat liberal or the somewhat liberal from the middle-of-the-road. We can say that the first group in each comparison is more liberal than the second, but we can't say by how much.

Ratio Variables

Some variables allow us to speak more precisely about the distances between the attributes constituting them. Consider age for a moment. The distance (in time) between 10 years old and 20 years old is exactly the same as that between 60 years old and 70 years old. Thus, it makes sense to talk about the distance (in time) between two ages (e.g., they are 10 years apart). Moreover, variables such as age have the quality of containing a genuine zero point—in this case, 0 years old. This quality is what allows us to examine ratios among the categories constituting such variables. Thus, we can say that a 20-year-old is twice as old as a 10-year-old. By comparison, we would have no grounds for saying one person is twice as religious as another, so religion can only be measured as a nominal or ordinal variable.

Ratio variables, then, share all of the qualities associated with nominal and ordinal variables, but they also have a zero point, and their categories can be separated by fixed units of distance and can thus be compared as ratios. Other examples of ratio measures include "annual amount of personal of household income," "years of schooling completed," and "hours worked per week."

Interval Variables

Somewhat rare in social research are ***interval variables***, which have the quality of standard intervals of measurement but lack a genuine zero point. One example is intelligence quotient, or IQ. Although IQ is calculated in such a way as to allow for a score of zero, such a score would not indicate a complete lack of any intelligence because the person would at least have been able to take the test.

Now, try to warm up to the idea of measuring temperature. The Celsius and Fahrenheit temperature scales both have 0° marks, but neither represents a total lack of heat, given that temperatures below zero are possible. The Kelvin scale, by contrast, is based on an absolute zero, which does represent a total lack of heat (measured in terms of molecular motion).

For most statistics used by social scientists, interval and ratio scales may be considered the same. When we start using SPSS Statistics, we'll see that its creators have lumped interval and ratio variables into a single category called ***scale***. Although these variables may be combined for practical purposes, the distinction between them helps us understand why a negative income might be interpreted as debt and why a negative age is impossible!

MEASUREMENT AND INFORMATION

Knowing a variable's level of measurement is important for selecting an appropriate statistic. Variables of different levels of measurement contain different amounts of information. The only information we have about nominal variables is the number of cases that share a common attribute. With ordinal variables, in addition to knowing how many cases fall into a given category, we know a greater-than, less-than relationship between the cases. Variables measured at the interval level have points that are equidistant from one another, so we know how much greater or less each case is than the others. Finally, with ratio variables, we have all of the characteristics of nominal, ordinal, and interval variables, plus the knowledge that zero is not arbitrary but means an absence of the phenomenon.

The statistics that SPSS Statistics has been programmed to compute are designed to make maximum use of the information preserved in a level of measurement. Using the mode on a sample of grade point averages ignores information used by the mean. Conversely, using the mean for a sample of religious preferences assumes information (equidistant points) not contained in a nominal measure. Responsible use of statistics requires selecting a statistic that matches the data's level of measurement. We'll talk about this more later. Right now, we want you to know that being able to identify a variable's level of measurement is essential for selecting the right statistical tool. We don't want to see you using a screwdriver when you need a hammer.

Table 2.1 displays the three primary levels of measurement that we discuss in this book: nominal, ordinal, and interval/ratio (scale). We purposefully designed the table as a series of steps to remind you that there is a hierarchy implied in the levels-of-measurement idea. Variables at the nominal level (the bottom step) contain the least amount of information, variables at the ordinal level (the middle step) contain more information, and variables at the interval/ratio level (the highest step) contain the most

TABLE 2.1 ⬢ Levels of Measurement

		INTERVAL/RATIO Distance between categories is meaningful. Examples: income (measured in thousands of dollars); age (measured in years)
	ORDINAL Categories can be rank ordered. Examples: social class (lower, working, middle, upper); attitudes toward gun control (strongly oppose, oppose, favor, strongly favor)	
NOMINAL Categories differ in name. Examples: gender (male, female); party identification (Democrat, Republican)		

information. You should also note that as you move from the nominal to the ordinal and finally to the interval/ratio level, each level has the qualities of the level(s) below it, plus a new trait.

MEASUREMENT OPTIONS

Sometimes, you will have options regarding the levels of measurement to be created in variables. For instance, although age can qualify as a ratio variable, it can be measured as ordinal (e.g., young, middle-aged, and old) or even as nominal (baby boomer or not baby boomer).

The significance of these levels of measurement will become more apparent when we begin to analyze the variables in our data set. As you'll discover, some analytic techniques are appropriate to nominal variables, some to ordinal variables, and some to ratio variables. On the one hand, you will need to know a variable's level of measurement to determine which analytic techniques are appropriate. On the other hand, where you have options for measurement, your choice of measurement level may be determined by the techniques you want to employ.

CLASSIFYING VARIABLES AS DISCRETE OR CONTINUOUS

In addition to distinguishing variables by their level of measurement, researchers sometimes classify variables as discrete or continuous. Just as distinguishing between levels of measurement helps us choose appropriate statistics, so too does knowing whether variables are discrete or continuous.

Discrete variables are variables whose values are completely separate from one another. In the GSS data, these include variables such as NUMBER OF SIBLINGS and SEX. These are variables with a limited number of distinct (i.e., discrete) values or categories that cannot be reduced or subdivided into smaller units or numbers. Discrete variables can be nominal (sex), ordinal (class rank), or interval/ratio (number of siblings). All of these variables are discrete because their values cannot be subdivided or reduced. A respondent may, for instance, have 1 dog, but they cannot have 0.5 or 0.25 dogs. Dogs, like people, come in discrete units of 1; they simply cannot be subdivided into smaller units.

(Note that it is possible to have an average number of dogs per family in a town that is not a whole number, say 1.5 dogs. This is an average; it does not and cannot indicate that any single family has

a fractional physical dog—dogs, like people, are whole creatures only, like Mina in the photo to the right posing with a previous edition of *Adventures in Social Research*!)

Continuous variables, on the other hand, are variables whose values can be infinitely subdivided, such as AGE or EDUC (education measured in years). These variables are continuous both because they have a time dimension and because time can be infinitely subdivided (i.e., years, months, weeks, days, hours, minutes, seconds, etc.). The level of measurement for continuous variables can be either interval/ratio (age measured in years) or ordinal (age measured as infant, toddler, adolescent, preteen, teenager, etc.).

Unlike levels of measurement, which tell us the amount of information provided by a measure, the discrete versus continuous distinction gives us information about the underlying characteristics of a variable. In particular, it refers to the phenomenon's divisibility, or whether the values of a variable can be subdivided into ever-smaller units, as described in Table 2.2.

Source: Photo by Billy Wagner.

TABLE 2.2 ⬢ Tips for Distinguishing Between Discrete and Continuous Variables

Discrete Variable	Example: Number of Siblings
	• Units CANNOT be reduced to ever-smaller units.
	• There is NOT an infinite number of other possible categories between any two categories of this variable (i.e., 1 and 2 siblings, or 2 and 3 siblings, etc.).
	• Variables can be nominal, ordinal, or interval/ratio (scale).
Continuous Variable	**Example: Age Measured in Years**
	• Units CAN be further reduced to smaller units (i.e., months, weeks, days, hours, minutes, seconds, etc.).
	• There ARE an infinite number of other possible categories between any two categories of this variable (i.e., 19 and 20 years old, 20 and 21 years old, etc.).
	• Variables can be interval or ratio (scale).

CONCLUSION

Measurement is a fundamental aspect of social science research. It may be seen as the transition from concepts to variables—from sometimes ambiguous mental images to precise, empirical measures. Whereas we often speak casually about such concepts as prejudice, social class, and liberalism in everyday conversation, social scientists must be more precise in their use of these terms.

Chapters 1 and 2 have given you a brief overview of two important issues in social scientific inquiry that are directly relevant to our primary focus—computerized statistical analysis. The chapters that follow build on this discussion of theory and measurement and show you the concrete techniques you need to engage in data analysis.

Main Points

- Measurement is a vital component of social scientific research.
- In designing and evaluating measurements, social scientists must pay particular attention to the problems of validity and reliability.
- A common remedy for problems of validity and reliability is the construction of composite measures using multiple indicators.
- Level of measurement signifies the different amounts and types of information obtained about a variable, and consideration of level of measurement is essential for selecting appropriate statistical tools.
- The four levels of measurement are nominal, ordinal, ratio, and interval.
- Variables of different levels of measurement contain different amounts of information.
- There is an implied hierarchy of the various levels of measurement, with the nominal level being at the bottom (meaning it can provide the least information), the ordinal level in the middle, and the interval and ratio levels being at the top (meaning they can provide the most information).
- In addition to classifying variables by their level of measurement, researchers distinguish between discrete and continuous variables.
- Discrete variables cannot be infinitely subdivided into ever-smaller units, whereas continuous variables can.

Key Terms

Composite measures 17
Continuous variables 21
Discrete variables 20
Interval variables 19
Levels of measurement 18
Multiple indicators 17
Nominal variables 18
Ordinal variables 18
Primary research 16
Ratio variables 19
Reliability 16
Scale 19
Secondary analysis 16
Validity 15

Review Questions

1. A researcher sets out to measure drug use on U.S. college campuses by asking a representative sample of undergraduates whether they are currently receiving federal grants or loans. What is the problem with this measure?
2. A researcher asks a representative sample of baby boomers how much alcohol they consumed during their college years and includes on the survey form a space for them to write in the actual number of drinks they consumed. One month later, the researcher administers the same questionnaire to the same respondents, and more than half the respondents report consuming much less alcohol during their college years than they reported just a month earlier. What is the problem with this measure?
3. Multiple indicators are useful in dealing with what types of problems?
4. Name one reason why it is important to know or be able to identify a variable's level of measurement.
5. Ordinal variables have all the qualities of variables at which other level of measurement?
6. Ratio variables have all the qualities of variables at which other levels of measurement?
7. The creators of SPSS Statistics have combined ratio and interval variables into one category that they refer to as _____.
8. A variable whose values can be infinitely subdivided is called a _____ variable.

Identify the level of measurement of each of the following variables, and classify the variables as either discrete or continuous (Questions 9–11).

9. A researcher measures respondents' attitudes toward premarital sex by asking the following question: "If a man and woman have sexual relations before marriage, do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?"
10. A researcher measures the amount of television viewing by asking the following question: "On the average day, how many hours do you personally watch television?" The researcher then asks respondents to fill in the actual number of hours in a space provided.
11. A researcher measures marital status by asking respondents whether they are currently married, widowed, divorced, separated, or never married.

Indicate whether the following statements (Questions 12–17) are true or false.

12. Certain variables can be measured at both the nominal and ordinal levels.
13. You are invited to watch a screening of a new movie and then rate it as excellent, good, fair, or poor. The level of measurement is nominal because the ratings differ in name.
14. A researcher asks respondents to indicate the last four digits of their Social Security numbers. The level of measurement for this variable is interval/ratio because the distance between categories is meaningful.
15. A researcher asks respondents how many siblings they have. This variable can be categorized as continuous.
16. A researcher asks respondents how long they have lived at their current residence. This variable can be categorized as continuous.
17. Discrete variables can be either nominal or ordinal but not interval/ratio.

Complete the activities below (Questions 18–20).

18. Construct measures of *annual income* at two levels of measurement.
19. Classify the variables you constructed in response to Question 18 as either discrete or continuous.
20. Construct measures of *individual age* at two levels of measurement.

Note

1. The authors are grateful to Professor Randall MacIntosh, California State University–Sacramento, for this suggestion.