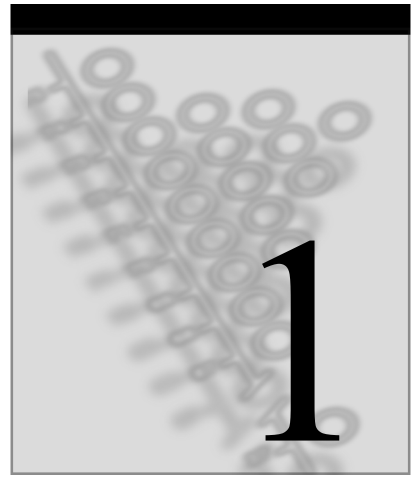


Basic Concepts in Research and DATA Analysis

Overview. This chapter reviews basic concepts and terminology with respect to research design and statistics. Different types of variables that can be analyzed are described as well as the scales of measurement with which these variables are assessed. The chapter reviews the differences between nonexperimental and experimental research and the differences between descriptive and inferential analyses. Finally, basic concepts in hypothesis testing are presented. After completing this chapter, you should be familiar with the fundamental issues and terminology of data analysis. You will be prepared to begin learning about SAS in subsequent chapters.

Introduction: A Common Language for Researchers

Research in the social sciences is a diverse topic. In part, this is because the social sciences represent a wide variety of disciplines, including (but not limited to) psychology, sociology, political science, anthropology, communication, education, management, and economics. A further complicating matter is the fact that, within each discipline, researchers can use a number of very different *quantitative methods* to conduct research. These methods can include unobtrusive observation, participant observation, case studies, interviews, focus groups, surveys, *ex post facto* studies, laboratory experiments, and field experiments to name but a few.

Despite this diversity in methods and topics investigated, most social science research still shares a number of common characteristics. Regardless of field, most research involves an investigator gathering data and performing analyses to determine what the data mean. In addition, most social scientists use a common language when conducting and reporting their research findings. For instance, researchers in both psychology and management speak of "testing null hypotheses" and "obtaining statistically significant *p* values."

The purpose of this chapter is to review some of the fundamental concepts and terms that are shared across social science disciplines. You should familiarize (or refamiliarize) yourself with this material before proceeding to the subsequent chapters, as most of the terms introduced here are referred to repeatedly throughout the text. If you are currently taking your first course in statistics, this chapter provides an elementary introduction; if you have already completed a course in statistics, it provides a quick review.

Steps to Follow When Conducting Research

The specific steps to follow when conducting research depend, in part, on the topic of investigation, where the researchers are in their overall program of research, and other factors. Nonetheless, much research in the social sciences follows a systematic course of action that begins with the statement of a research question and ends with the researcher drawing conclusions about a null hypothesis. This section describes the research process as a planned sequence that consists of the following six steps:

1. developing a statement of the research question;
2. developing a statement of the research hypotheses (i.e., specific questions to be tested);
3. defining the instruments (e.g., questionnaires, unobtrusive observation measures);
4. gathering the data;
5. analyzing the data;
6. drawing conclusions regarding the null and research hypotheses.

The preceding steps are illustrated here with reference to a fictitious research problem. Imagine that you have been hired by a large insurance company to find ways of improving the productivity of its insurance agents. Specifically, the company would like you to find ways to increase the amount of insurance policies sold by the average agent. You will therefore begin a program of research to identify the determinants of agent productivity.

The Research Question

The process of research often begins with an attempt to arrive at a clear statement of the **research question** (or questions). The research question is a statement of what you hope to learn by the time you have completed the study. It is good practice to revise and refine the research question several times to ensure that you are very explicit and precise.

For example, in the present case, you might begin with the question, "What is the difference between agents who sell a lot of insurance compared to those who sell very little insurance?" An alternative question might be, "What variables have a causal effect on the amount of insurance sold by agents?" Upon reflection, you might realize that the insurance company really only wants to know what things management can do to help agents to sell more. This might eliminate from consideration certain personality traits or demographic variables that are not under management's control, and substantially narrow the focus of the research program. Upon further refinement, a more specific statement of the research question might be, "What variables under the control of management have a causal effect on the amount of insurance sold by agents?" Once you define the research question(s) clearly, you are in a better position to develop a good hypothesis that provides an answer to the question(s).

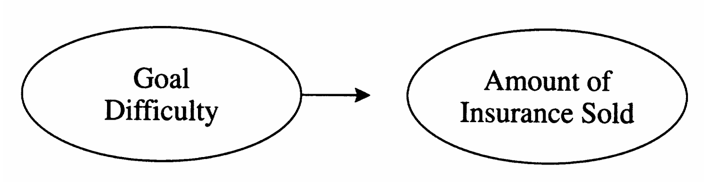
The Hypothesis

A **hypothesis** is a statement about the predicted relationships among observed events or factors. A good hypothesis in the present case might identify which *specific variables* will have a causal effect on the amount of insurance sold by agents. For example, a hypothesis might predict that agents' level of training will have a positive effect on the amount of insurance sold. Or, it might predict that agents' level of motivation will positively affect sales.

In developing the hypothesis, you might be influenced by any number of sources: an existing theory; some related research; or even personal experience. Let's assume that you have been influenced by **goal-setting theory** that states, among other things, that higher levels of work performance are achieved when employees have difficult work-related goals. Drawing on goal-setting theory, you now state the following hypothesis: "The difficulty of goals that agents set for themselves is positively related to the amount of insurance they sell."

Notice how this statement satisfies our definition for a hypothesis as it is a statement about the assumed causal relationship between two variables. The first variable can be labeled *Goal Difficulty*, and the second can be labeled *Amount of Insurance Sold*. This relationship is illustrated in Figure 1.1:

Figure 1.1 Hypothesized Relationship between Goal Difficulty and Amount of Insurance Sold



The same hypothesis could be restated in a number of other ways. For example, the following hypothesis makes the same basic prediction: "Agents who set difficult goals for themselves sell greater amounts of insurance than agents who do not set difficult goals."

Notice that these hypotheses are stated in the present tense. It is also acceptable to state hypotheses in the past tense. For example, the preceding could have been stated: "Agents who set difficult goals for themselves sold greater amounts of insurance than agents who did not set difficult goals." The verb tense of the hypothesis depends on whether the researcher will be examining data already collected or will undertake data collection at some future point.

You should also note that these two hypotheses are quite broad in nature. In many research situations, it is helpful to state hypotheses that are more specific in the predictions they make. A more specific hypothesis for the present study might be: "Agents who score above 60 on the Smith Goal Difficulty Scale will sell greater amounts of insurance than agents who score below 40 on the Smith Goal Difficulty Scale."

Defining the Instrument, Gathering Data, Analyzing Data, and Drawing Conclusions

With the hypothesis stated, you can now test it by conducting a study in which you gather and analyze relevant data. *Data* can be defined as a collection of scores obtained when participants' characteristics and/or performance are assessed. For instance, you might decide to test your hypothesis by conducting a simple correlational study. As an example, you might identify a group of 100 agents and determine:

- the difficulty of the goals that have been set for each agent;
- the amount of insurance sold by each agent.

Different types of instruments are used to obtain different types of data. For example, you can use a questionnaire to assess goal difficulty, but rely on company records for measures of insurance sold. Once the data are gathered, each agent will have a score indicating the difficulty of his or her goals and a second score indicating the amount of insurance that he or she has sold.

With the data in hand, you analyze these data to determine if the agents with the more difficult goals did, as hypothesized, sell more insurance. If yes, the study provides support for your hypothesis; if no, it fails to provide support. In either case, you draw conclusions regarding the tenability of your hypotheses. This information is then considered with respect to your research question. These findings might stimulate new questions and hypotheses for subsequent research and the cycle would repeat. For example, if you found support for your hypothesis with the current correlational study, you might choose to follow up with a study using a different method, such as an experimental study. (The difference between these methods is described later.) Over time, a body of research evidence would accumulate, and researchers would be able to review study findings to draw conclusions about the determinants of insurance sales.

Variables, Values, and Observations

Variables

When discussing data, social scientists often speak in terms of variables, values, and observations. For the type of research discussed here, a **variable** refers to some specific participant characteristic that can assume different values (i.e., the values vary). For the participants in the study just described, Amount of Insurance Sold is an example of a variable: some participants had sold a lot of insurance whereas others had sold less. A second variable was Goal Difficulty: some participants set more difficult goals while others set less difficult goals. Participant Age and Sex (male or female) were other variables, though not considered in our initial hypothesis.

Values

A **value** is a specific quantity or category of a variable. For example, Amount of Insurance Sold is a variable that can assume a range of values. One agent might sell $2,000,000 worth of insurance in one year, one might sell $100,000, whereas another may sell nothing (i.e., $0). Age is another variable that can assume a variety of values. In our example, these values range from a low of 22 years (the youngest agent) to a high of 64 years (the oldest agent). In other words, age is a variable that is comprised of a wide array of specific values.

Quantitative Variables versus Classification Variables

For both types of these variables, a given value is the specific score that indicates participants' standing with respect to variable of interest. Amount of Insurance Sold and Age are **quantitative variables** since numbers serve as values. The word "score" is an appropriate substitute for the word value in these cases.

A different type of variable is a **classification variable** or, alternatively, **qualitative variable** or **categorical variable**. With classification variables, different values represent different groups to which participants can belong. Sex is a good example of a classification variable as it (generally) assumes only one of two values (i.e., participants are classified as either male or female). Race is an example of a classification variable that can assume a larger number of values: participants can be classified as Caucasian; African American; Asian American; or as belonging to a large number of other groups. Notice why these variables are classification variables and not quantitative variables. The values represent only group membership; they do not represent a characteristic that some participants possess in greater quantity than others.

Observational Units

In discussing data, researchers often make reference to **observational units** that can be defined as individual participants (or objects) that serve as the source of the data. (Observational units are also referred to as *units of analysis*.) Within the social sciences, a person usually serves as the observational unit under study (although it is also possible to use some other entity such as an individual school or organization as the observational unit). In this text, the person is used as the observational unit in all examples. Researchers often refer to the number of **observations** (or **cases**) included in their datasets. This simply refers to the number of participants studied.

For a more concrete illustration of the concepts discussed so far, consider the following dataset:

Table 1.1

Insurance Sales Data

Observation	Name	Sex	Age	Goal Difficulty Scores	Rank	Sales
1	Bob	M	34	97	2	$598,243
2	Pietro	M	56	80	1	$367,342
3	LeMont	M	36	67	4	$254,998
4	Susan	F	24	40	3	$80,344
5	Saleem	M	22	37	5	$40,172
6	Mack	M	44	24	6	$0

The preceding table reports information about six research participants: Bob; Pietro; LeMont; Susan; Saleem; and Mack. Therefore, the dataset includes six observations. Information about a given observation (participant) appears as a **row** running left to right across the table. The first **column** of the dataset (running vertically) indicates the observation number, and the second column reports the name of the participant. The remaining five columns report information on the five research variables under study. The “Sex” column reports participant sex, which can assume one of two values: “M” for male and “F” for female. The “Age” column reports participants’ age in years. The “Goal Difficulty Scores” column reports participants’ scores on a fictitious goal difficulty scale. (Assume that each participant completed a 20-item questionnaire that assessed the difficulty of his or her work goals.)

Depending on their responses, participants receive a score that can range from a low of 0 (meaning that work goals are quite easy) to a high of 100 (meaning that they are quite difficult). The “Rank” column shows how participants were ranked by their supervisor according to their overall effectiveness as agents when compared to one another. A rank of 1 represents the most effective agent whereas a rank of 6 represents the least effective. Finally, the “Sales” column indicates the amount of insurance sold by each agent (in dollars) during the past year.

This example illustrates a very small dataset with six observations and five research variables (i.e., Sex, Age, Goal Difficulty, Rank and Sales). One variable is a classification variable (Sex), while the remainder are quantitative variables. The numbers or letters that appear within a column represent some of the values that can be assumed by that variable.

Scales of Measurement

One of the most important schemes for classifying a variable involves its **scale of measurement**. Data generally fall within four different scales of measurement: nominal; ordinal; ratio; and interval. Before analyzing a dataset, it is important to determine which scales of measurement were used because certain types of statistical procedures require specific scales of measurement. In this text, each chapter that deals with a specific statistical procedure specifies what scale of measurement is required. The researcher must determine scales of measurement for study variables before selecting which statistical procedures to use.

Nominal Scales

A **nominal scale** is a classification system that places people, objects, or other entities into mutually exclusive categories. A variable measured along a nominal scale is a *classification variable*; it simply indicates the group to which each participant belongs. The examples of classification variables provided earlier (e.g., Sex and Ethnicity) also are examples of nominal-level variables; they tell us to which group a participant belongs but they do not provide any quantitative information. That is, the Sex variable identifies participants as either male or female but it does not tell us that participants possess more or less of a specific characteristic relative to others. However, the remaining three scales of measurement—ordinal, interval, and ratio—provide some quantitative information.

Ordinal Scales

Values on an **ordinal scale** represent the rank order of participants with respect to the variable being assessed. For example, the preceding table includes one variable called Rank that represents the rank ordering of participants according to their overall effectiveness as agents. Values for this ordinal scale represent a *hierarchy of levels* with respect to the construct of “effectiveness.” We know that the agent ranked 1 was perceived as being more effective than the agent ranked 2, that the agent ranked 2 was more effective than the one ranked 3, and so forth.

The information conveyed by an ordinal scale is limited because equal differences in scale values do not necessarily have equal quantitative meaning. For example, notice the following rankings:

Rank	**Name**
1	Bob
2	Pietro
3	Susan
4	LeMont
5	Saleem
6	Mack

Notice that Bob was ranked 1 while Pietro was ranked 2. The difference between these two rankings is 1 (because 2 – 1 = 1), so there is one unit of difference between Bob and Pietro. Now notice that Saleen was ranked 5 while Mack was ranked 6. The difference between these two rankings is also 1 (because 6 – 5 = 1), so there is also 1 unit of difference between Saleem and Mack. Putting the two together, the difference in ranking between Bob and Pietro is equal to the difference in ranking between Saleem and Mack.

But, does this mean that the difference in overall effectiveness between Bob and Pietro is equal to the difference in overall effectiveness between Saleem and Mack? Not necessarily. It is possible that Bob was significantly superior to Pietro in effectiveness, while Saleem might have been only slightly superior to Mack. In fact, this appears to be the case. Whereas Bob had sold policies totaling $598,243, Pietro had sold $367,342 for a difference of $230,830 between the two. In contrast, the difference in sales between Saleem ($40,170) and Mack ($0) was only a faction of the difference between Bob and Pietro (i.e., $40,170 vs. $230,830). This example indicates that these rankings reveal very little about the quantitative differences between participants with regard to the underlying construct (effectiveness, in this case). An ordinal scale simply provides a rank order. Other scales of measurement are required to provide this added level of measurement.

Interval Scales

With an **interval scale**, equal differences between values have equal quantitative meaning. For this reason, it can be seen that an interval scale provides more quantitative information than an ordinal scale. A good example of interval measurement is the Fahrenheit scale used to measure temperature. With the Fahrenheit scale, the difference between 70 degrees and 75 degrees is equal to the difference between 80 degrees and 85 degrees. In other words, the units of measurement are equal throughout the full range of the scale.

However, the interval scale also has a limitation; it does not have a true zero point. A **true zero point** means that a value of zero on the scale represents zero quantity of the variable being assessed. It should be obvious that the Fahrenheit scale does not have a true zero point; when the thermometer reads 0 degrees Fahrenheit, that does not mean that there is absolutely no heat present in the environment.

Social scientists often assume that many of their man-made variables are measured on an interval scale. In the preceding study involving insurance agents, for example, you would probably assume that scores from the goal difficulty questionnaire constitute an interval-level scale (i.e., you would likely assume that the difference between a score of 50 and 60 is approximately equal to the difference between a score of 70 and 80). Many researchers would also assume that scores from an instrument such as an intelligence test are also measured at the interval level of measurement.

On the other hand, some researchers are skeptical that instruments such as these have true equal-interval properties and prefer to refer to them as **quasi-interval** scales (e.g., Likert-type scales to which respondents indicate their degree of agreement to a series of statements with a fixed number of response alternatives such as strongly disagree, disagree, neutral, agree, and strongly agree). Disagreements concerning the level of measurement with such instruments continue to be a controversial topic within the social sciences (i.e., whether scale responses ranging from *strongly disagree* to *strongly agree* constitute ordinal- or interval-level measurement).

It is clear that there is no true zero point with either of the preceding instruments. A score of 0 on the goal difficulty scale does not indicate the complete absence of goal difficulty, and a score of 0 on an intelligence test does not indicate the complete absence of intelligence. A true zero point can be found only with variables measured on a ratio scale.

Ratio Scales

Ratio scales are similar to interval scales in that equal differences between scale values have equal quantitative meaning. However, ratio scales also have a true zero point which gives them an additional property. With ratio scales, it is possible to make meaningful statements about the ratios between scale values. For example, the system of inches used with a common ruler is an example of a ratio scale. There is a true zero point with this system in which zero inches does, in fact, indicate a complete absence of length. With this scale, therefore, it is possible to make meaningful statements about ratios. It is appropriate to say that an object four inches long is twice as long as an object two inches long. Age, as measured in years, is also on a ratio scale as a 10-year old house is twice as old as a 5-year old house. Notice that it is not possible to make these statements about ratios with the interval-level variables discussed above. One would not say that a person with an IQ of 160 is twice as intelligent as a person with an IQ of 80.

Although ratio-level scales might be easiest to find when one considers the physical properties of objects (e.g., height and weight), they are also common in the type of research discussed in this text. For example, the study discussed previously included the variables for age and amount of insurance sold (in dollars). Both of these have true zero points and are measured as ratio scales.

Basic Approaches to Research

Nonexperimental Research

Much research can be categorized as being either experimental or nonexperimental in nature. In **nonexperimental research** (also called **nonmanipulative** or **correlational** research), the investigator simply studies the association between two or more naturally occurring variables. A **naturally occurring variable** is one that is not manipulated or controlled by the researcher; it is simply observed and measured (e.g., the age of insurance salespersons).

The insurance study described previously is a good example of nonexperimental research since you simply measured two naturally occurring variables (i.e., goal difficulty and amount of insurance sold) to determine whether they were related. If, in a different study, you investigated the relationship between IQ and college grade point average (GPA), this would also be an example of nonexperimental research.

With nonexperimental research designs, social scientists often refer to criterion variables and predictor variables. A **criterion variable** is an outcome variable that might be predicted by one or more other variables. The criterion variable is generally the main focus of the study; it is the outcome variable mentioned in the statement of the research problem. In our example, the criterion variable is Amount of Insurance Sold.

The **predictor variable**, on the other hand, is that variable used to predict or explain values of the criterion. In some studies, you might even believe that the predictor variable has a causal effect on the criterion. In the insurance study, for example, the predictor variable was Goal Difficulty. Because you believed that Goal Difficulty positively affects insurance sales, you conducted a study in which Goal Difficulty is identified as the predictor and Sales as the criterion. You do not necessarily have to believe that there is a causal relationship between Goal Difficulty and Sales to conduct this study. You might simply be interested in determining whether there is an association between these two variables (i.e., as the values for the predictor change, a corresponding change in the criterion variable is observed).

You should note that nonexperimental research that examines the relationship between just two variables generally provides little evidence concerning cause-and-effect relationships. The reasons for this can be seen by reviewing the study on insurance sales. If the social scientist conducts this study and finds that the agents with the more difficult goals also tend to sell more insurance, does that mean that having difficult goals *caused* them to sell more insurance? Not necessarily. You can argue that selling a lot of insurance increases the agents' self-confidence and that this, in turn, causes them to set higher work goals for themselves. Under this second scenario, it was actually the insurance sales that had a causal effect on Goal Difficulty.

As this example shows, with nonexperimental research it is often possible to obtain a result consistent with a range of causal explanations. Hence, a strong inference that "variable A had a causal effect on variable B" is seldom possible when you conduct simple correlational research with just two variables. To obtain stronger evidence of cause and effect, researchers generally either analyze the relationships between a larger number of variables using sophisticated statistical procedures that are beyond the scope of this text, or drop the nonexperimental approach entirely and, instead, use experimental research methods. The nature of experimental research is discussed in the following section.

Experimental Research

Most experimental research can be identified by three important characteristics:

- Participants are randomly assigned to experimental and control conditions.
- The researcher manipulates one or more variables.
- Participants in different experimental conditions are treated similarly with regard to all variables except the manipulated variable.

To illustrate these concepts, assume that you conduct an experiment to test the hypothesis that goal difficulty positively affects insurance sales. Assume that you identify a group of 100 agents who will serve as study participants. You randomly assign 50 agents to a "difficult-goal" condition. Participants in this group are told by their superiors to make at least 25 cold calls (sales calls) to potential policyholders per week. The other 50 agents assigned to the "easy-goal" condition have been told to make just five cold calls to potential policyholders per week. The design of this experiment is illustrated in Figure 1.2.

Figure 1.2 Design of the Experiment Used to Assess the Effects of Goal Difficulty

Group		Treatment Conditions Under the Independent Variable (Goal Difficulty)		Results Obtained with the Dependent Variable (Amount of Insurance Sold)
Group 1 (n = 50)	→	Difficult-Goal Condition	→	\$156,000 in Sales
Group 2 (n = 50)	→	Easy-Goal Condition	→	\$121,000 in Sales

After 12 months, you determine how much new insurance each agent has sold that year. Assume that the average agent in the difficult-goal condition sold \$156,000 worth of new policies while the average agent in the easy-goal condition sold just \$121,000 worth.

It is possible to use some of the terminology associated with nonexperimental research when discussing this experiment. For example, it would be appropriate to continue to refer to Amount of Insurance Sold as being a criterion variable because this is the outcome variable of central interest. You could also continue to refer to Goal Difficulty as the predictor variable because you believe that this variable will predict sales to some extent.

Notice, however, that Goal Difficulty is now a somewhat different variable. In the nonexperimental study, Goal Difficulty was a naturally occurring variable that could take on a variety of values (whatever score participants received on the goal difficulty questionnaire). In the present experiment, however, Goal Difficulty is a **manipulated variable**, which means that you (as the researcher) determined what value of the variable would be assigned to both participant groups. In this experiment, Goal Difficulty could assume only one of two values. Therefore, Goal Difficulty is now a classification variable, assessed on a nominal scale.

Although it is acceptable to speak of predictor and criterion variables within the context of experimental research, it is more common to speak in terms of independent and dependent variables. The independent variable (IV) is that variable whose values (or levels) are selected by the experimenter to determine what effect the independent variable has on the dependent variable. The **independent variable** is the experimental counterpart to a predictor variable. A dependent variable (DV) is some aspect of the study participant's behavior that is assessed to reflect the effects of the independent variable. The **dependent variable** is the experimental counterpart to a criterion variable. In the present experiment, Goal Difficulty is the independent variable while Sales is the dependent variable. Remember that the terms predictor variable and criterion variable can be used with almost any type of research, but that the terms independent variable and dependent variable should be used only with experimental research.

Researchers often refer to the different **levels of the independent variable**. These levels are also referred to as **experimental conditions** or **treatment conditions** and correspond to the different groups to which participants can be assigned. The present example includes two experimental conditions: a difficult-goal condition and an easy-goal condition.

With respect to the independent variable, you can speak in terms of the experimental group versus the control group. Generally speaking, the **experimental group** receives the experimental treatment of interest while the **control group** is an equivalent group of participants who do not receive this treatment. The simplest type of experiment consists of just one experimental group and one control group. For example, the present study could have been redesigned so that it consisted of an experimental group that was assigned the goal of making 25 cold calls (the difficult-goal condition) and a control group in which no goals were assigned (the no-goal condition). Obviously, you can expand the study by creating more than one experimental group. You could do this in the present case by assigning one experimental group the difficult goal of 25 cold calls and the second experimental group the easy goal of just 5 cold calls.

Descriptive versus Inferential Statistical Analysis

To understand the difference between descriptive and inferential statistics, you must first understand the difference between populations and samples. A **population** is the *entire collection* of a carefully defined set of people, objects, or events. For example, if the insurance company in question employed 10,000 insurance agents in the European Union, then those 10,000 agents would constitute the population of agents hired by that company. A **sample**, on the other hand, is a subset of the people, objects, or events selected from a population. For example, the 100 agents used in the experiment described earlier constitute a sample.

Descriptive Analyses

A **parameter** is a descriptive characteristic of a population. For example, if you assessed the average amount of insurance sold by all 10,000 agents in this company, the resulting average would be a parameter. To obtain this average, of course, you would first need to tabulate the amount of insurance sold by each and every agent. In calculating this average, you are engaging in descriptive analysis. Descriptive analyses organize, summarize, and identify major characteristics of the population.

Most people think of populations as being very large groups, such as all of the people in the United Kingdom. However, a group does not have to be large to be a population, it only has to be the entire collection of the people or things being studied. For example, a teacher can define all twelfth-grade students in a single school as a population and then calculate the average score of these students on a measure of class satisfaction. The resulting average would be a population parameter.

Inferential Analyses

A **statistic**, on the other hand, is a numerical value that is computed from a sample and either describes some characteristic of that sample such as the average value, or is used to make inferences about the population from which the sample is drawn. For example, if you were to compute the average amount of insurance sold by your sample of 100 agents, that average would be a statistic because it summarizes a specific characteristic of the sample. Remember that the word "statistic" is generally associated with samples while "parameter" is generally associated with populations.

In contrast to descriptive analyses, **inferential statistics** involve information from a sample to make inferences, or estimates, about the population (i.e., infer from the sample to the larger population). For example, assume that you need to know how much insurance is sold by the average agent in the company. It might not be possible to obtain the necessary information from all 10,000 agents and then determine the average. An alternative would be to draw a random (and ideally representative) sample of 100 agents and determine the average amount sold by this subset. A **random sample** is a subset of the population in which each member of that population has an equal chance of selection. If this group of 100 sold an average of $179,322 worth of policies last year, then your best guess of the amount of insurance sold by all 10,000 agents would likewise be $179,322 on average. Here, you have used characteristics of the sample to make inferences about characteristics of the population. This is the real value of inferential statistical procedures; they allow you to examine information obtained from a relatively small sample and then make inferences about the overall population. For example, pollsters conduct telephone surveys to ascertain the voting preferences of Canadians leading up to, and between, federal elections. From randomly selected samples of approximately 1,200 participants, these pollsters can extrapolate their findings to the population of 20 million eligible voters with considerable accuracy (i.e., within relatively narrow limits).

Hypothesis Testing

Most of the procedures described in this text are inferential procedures that allow you to test specific hypotheses about the characteristics of populations. As an illustration, consider the simple experiment described earlier in which 50 agents were assigned to a difficult-goal condition and 50 other agents to an easy-goal condition. After one year, the agents with difficult goals had sold an average of $156,000 worth of insurance while the agents with easy goals had sold $121,000 worth. On the surface, this would seem to support your hypothesis that difficult goals cause agents to sell more insurance. But can you be sure of this? Even if goal setting had no effect at all, you would not really expect the two groups of 50 agents to sell exactly the same amount of insurance; one group would sell somewhat more than the other due to chance alone. The difficult-goal group did sell more insurance, but did it sell a sufficiently greater amount of insurance to suggest that the difference was due to your manipulation (i.e., random assignment to the experimental group)?

What's more, it could easily be argued that you don't even care about the amount of insurance sold by these two relatively small samples. What really matters is the amount of insurance sold by the larger populations that they represent. The first population could be defined as "the population of agents who are assigned difficult goals" and the second would be "the population of agents who are assigned easy goals." Your real research question involves the issue of whether the first population sells more than the second. This is where hypothesis testing comes in.

Types of Inferential Tests

Generally speaking, there are two types of tests conducted when using inferential procedures: tests of group differences and tests of association. With a **test of group differences,** you typically want to know whether populations differ with respect to their scores on some criterion variable. The present experiment would lead to a test of group differences because you want to know whether the average amount of insurance sold in the population of difficult-goal agents is different from the average amount sold in the population of easy-goal agents. A different example of a test of group differences might involve a study in which the researcher wants to know whether Caucasian-Americans, African-Americans, and Asian-Americans differ with respect to their scores on an academic achievement scale. Notice that in both cases, two or more distinct populations are being compared with respect to their scores on a single criterion variable.

With a **test of association** on the other hand, you are working with a single group of individuals and want to know whether or not there is a relationship between two or more variables. Perhaps the best-known test of association involves testing the significance of a correlation coefficient. Assume that you have conducted a simple correlational study in which you asked 100 agents to complete the 20-item goal-difficulty questionnaire. Remember that with this questionnaire, participants could receive a score ranging from a low of 0 to a high of 100 (interval measurement). You could then correlate these goal-difficulty scores with the amount of insurance sold by agents that year. Here, the goal-difficulty scores constitute the predictor variable while the amount of insurance sold serves as the criterion. Obtaining a strong positive correlation between these two variables would mean that the more difficult the agents' goals, the more insurance they tended to sell. Why would this be called a test of association? That is because you are determining whether there is an association, or *relationship*, between the predictor and criterion variables. Notice also that only one group is studied (i.e., there is no random assignment or experimental manipulation that creates a difficult-goal sample versus an easy-goal sample).

To be thorough, it is worth mentioning that there are some relatively sophisticated procedures that also allow you to perform a third type of test: whether the association between variables is the same across multiple groups. Analysis of covariance (ANCOVA) is one procedure that enables such a test. For example, you might hypothesize that the association between self-reported Goal Difficulty and insurance sales is stronger in the population of agents assigned difficult goals than it is in the population assigned easy goals. To test this assertion, you might randomly assign a group of insurance agents to either an easy-goal condition or a difficult-goal condition (as described earlier). Each agent could complete the 20-item self-report goal difficulty scale and then be exposed to the appropriate treatment. Subsequently, you would record each agent's sales. Analysis of covariance would allow you to determine whether the relationship between questionnaire scores and sales is stronger in the difficult-goal population than it is in the easy-goal population. (ANCOVA would also allow you to test a number of additional hypotheses.)

Types of Hypotheses

Two different types of hypotheses are relevant to most statistical tests. The first is called the null hypothesis, which is generally abbreviated as H_0. The **null hypothesis** is a statement that, in the population(s) being studied, there are either (a) *no difference* between the groups or; (b) *no relationship* between the measured variables. For a given statistical test, either (a) or (b) will apply, depending on whether one is conducting a test of group differences or a test of association, respectively.

With a **test of group differences**, the null hypothesis states that, in the population, there are no differences between groups studied with respect to their mean scores on the criterion variable. In the experiment in which a difficult-goal condition is being compared to an easy-goal condition, the following null hypothesis might be used:

> H_0: In the population, individuals assigned difficult goals do not differ from individuals assigned easy goals with respect to the mean amount of insurance sold.

This null hypothesis can also be expressed mathematically with symbols in the following way:

$$H_0\text{:}\ \ M_1 = M_2$$

where:

H_0 represents null hypothesis

M_1 represents mean sales for the difficult-goal population

M_2 represents mean sales for the easy-goal population

In contrast to the null hypothesis, you will also form an alternative hypothesis (H_1) that states the opposite of the null. The **alternative hypothesis** is a statement that there *is a difference* between groups, or that there *is a relationship* between the variables in the population(s) studied.

Perhaps the most common alternative hypothesis is a **nondirectional alternative hypothesis** (often referred to as a 2-sided hypothesis). With a test of group differences, a no-direction alternative hypothesis predicts that the various populations will differ, but makes no specific prediction as to how they will differ (e.g., one outperforming the other). In the preceding experiment, the following nondirectional null hypothesis might be used:

H_1: In the population, individuals assigned difficult goals differ from individuals assigned easy goals with respect to the amount of insurance sold.

This alternative hypothesis can also be expressed with symbols in the following way:

H_1: $M_1 \neq M_2$

In contrast, a directional or 1-sided alternative hypothesis makes a more specific statement regarding the expected outcome of the analysis. With a test of group differences, a directional alternative hypothesis not only predicts that the populations differ, but also contends which will be relatively high and which will be relatively low. Here is a directional alternative hypothesis for the preceding experiment:

H_1: The amount of insurance sold is higher in the population of individuals assigned difficult goals than in the population of individuals assigned easy goals.

This hypothesis can be symbolically represented in the following way:

H_1: $M_1 > M_2$

Had you believed that the easy-goal population would sell more insurance, you would have replaced the "greater than" symbol (>) with the "less than" symbol (<), as follows:

H_1: $M_1 < M_2$

Null and alternative hypotheses are also used with tests of association. For the study in which you correlated goal-difficulty questionnaire scores with the amount of insurance sold, you might have used the following null hypothesis:

H_0: In the population, the correlation between goal-difficulty scores and the amount of insurance sold is zero.

You could state a nondirectional alternative hypothesis that corresponds to this null hypothesis in this way:

H_1: In the population, the correlation between goal-difficulty scores and the amount of insurance sold is not equal to zero.

Notice that the preceding is an example of a nondirectional alternative hypothesis because it does not specifically predict whether the correlation is positive or negative, only that it is not zero. On the other hand, a directional alternative hypothesis might predict a positive correlation between the two variables. You could state such a prediction as follows:

H_1: In the population, the correlation between goal-difficulty scores and the amount of insurance sold is greater than zero.

There is an important advantage associated with the use of directional alternative hypotheses compared to nondirectional hypotheses. Directional hypotheses allow researchers to perform 1-sided statistical tests (also called *1-tail tests*), which are relatively powerful. Here, "powerful" means that *one-sided* tests are more likely to find statistically significant differences between groups when differences really do exist. In contrast, nondirectional hypotheses allow only *2-sided* statistical tests (also called *2-tail tests*) that are less powerful.

Because they lead to more powerful tests, directional hypotheses are generally preferred over nondirectional hypotheses. However, directional hypotheses should be stated only when they can be justified on the basis of theory, prior research, or some other acceptable reason. For example, you should state the directional hypothesis that "the amount of insurance sold is higher in the population of individuals assigned difficult goals than in the population of individuals assigned easy goals" only if there are theoretical or empirical reasons to believe that the difficult-goal group will indeed score higher on insurance sales. The same should be true when you specifically predict a positive correlation rather than a negative correlation (or vice versa).

The *p* or Significance Value

Hypothesis testing, in essence, is a process of determining whether you can reject your null hypothesis with an acceptable level of confidence. When analyzing data with SAS, you will review the output for two pieces of information that are critical for this purpose: the obtained statistic and the probability (p) or significance value associated with that statistic. For example, consider the experiment in which you compared the difficult-goal group to the easy-goal group. One way to test the null hypothesis associated with this study would be to perform an independent samples t test (described in detail in Chapter 8, "t Tests: Independent Samples and Paired Samples"). When the data analysis for this study has been completed, you would review a t statistic and its corresponding p value. If the p value is very small (e.g., $p < .05$), you will reject the null hypothesis.

For example, assume that you obtain a t statistic of 0.14 and a corresponding p value of .90. This p value indicates that there are 90 chances in 100 that you would obtain a t statistic of 0.14 (or larger) if the null hypothesis were true. Because this probability is high, you would report that there is very little evidence to refute the null hypothesis. In other words, you would fail to reject your null hypothesis and would, instead, conclude that there is not sufficient evidence to find a statistically significant difference between groups (i.e., between group differences might well be due only to chance).

On the other hand, assume that the research project instead produces a t value of 3.45 and a corresponding p value of . 01. The p value of . 01 indicates that there is only one chance in 100 that you would obtain a t statistic of 3.45 (or larger) if the null hypothesis were true. This is so unlikely that you can be fairly confident that the null hypothesis is not true. You would therefore reject the null hypothesis and conclude that the two populations do, in fact, appear to differ. In rejecting the null hypothesis, you have tentatively accepted the alternative hypothesis.

Technically, the p value does not really provide the probability that the null hypothesis is true. Instead, it provides the probability that you would obtain the present results (the present t statistic, in this case) if the null hypothesis were true. This might seem like a trivial difference, but it is important that you not be confused by the meaning of the *p* value.

Notice that you were able to reject the null hypothesis only when the *p* value was a fairly small number (.01, in the above example). But how small must a *p* value be before you can reject the null hypothesis? A *p* value of .05 seems to be the most commonly accepted cutoff. Typically, when researchers obtain a *p* value *larger* than .05 (such as .13 or .37), they will fail to reject the null hypothesis and will instead conclude that the differences or relationships being studied were not statistically significant (i.e., differences can occur as a matter of chance alone). When they obtain a *p* value *smaller* than .05 (such as .04 or .02 or . 01), they will reject the null hypothesis and conclude that differences or relationships being studied are statistically significant. The .05 level of significance is not an absolute rule that must be followed in all cases, but it should be serviceable for most types of investigations likely to be conducted in the social sciences.

Fixed Effects versus Random Effects

Experimental designs can be represented as mathematical models and these models can be described as fixed-effects models, random-effects models, or mixed-effects models. The use of these terms refers to the way that the levels of the independent (or predictor) variable were selected.

When the researcher arbitrarily selects the levels of the independent variable, the independent variable is called a **fixed-effects factor** and the resulting model is a **fixed-effects model**. For example, assume that, in the current study, you arbitrarily decide that participants in your easy-goal condition would be told to make just 5 cold calls per week and that participants in the difficult-goal condition would be told to make 25 cold calls per week. In this case, you have *fixed* (i.e., arbitrarily selected) the levels of the independent variable. Your experiment therefore represents a fixed-effects model.

In contrast, when the researcher randomly selects levels of the independent variable from a population of possible levels, the independent variable is called a **random-effects factor**, and the model is a **random-effects model**. For example, assume that you have determined that the number of cold calls that an insurance agent could possibly place in one week ranges from 0 to 45. This range represents the population of cold calls that you could possibly research. Assume that you use some random procedure to select two values from this population (perhaps by drawing numbers from a hat). Following this procedure, the values 12 and 32 are drawn. When conducting your study, one group of participants is assigned to make at least 12 cold calls per week, while the second is assigned to make 32 calls. In this instance, your study represents a random-effects model because the levels of the independent variable were randomly selected.

Most research in the social sciences involves **fixed-effects models**. As an illustration, assume that you are conducting research on the effectiveness of hypnosis in reducing anxiety among participants who suffer from test anxiety. Specifically, you could perform an experiment that compares the effectiveness of 10 sessions of relaxation training versus 10 sessions of relaxation training plus hypnosis. In this study, the independent variable might be labeled something like Type of Therapy. Notice that you did not randomly select these two treatment conditions from the population of all possible treatment conditions; you knew which treatments you wished to compare and designed the study accordingly. Therefore, your study represents a fixed-effects model.

To provide a nonexperimental example, assume that you were to conduct a study to determine whether Hispanic-Americans score significantly higher than Korean-Americans on academic achievement. The predictor variable in your study would be Ethnicity while the criterion variable would be scores on some index of academic achievement. In all likelihood, you would not have arbitrarily chosen "Hispanic American" versus "Korean American" because you are particularly interested in these two ethnic groups; you did not randomly select these groups from all possible ethnic groups. Therefore, the study is again an example of a fixed-effects model.

Of course, random-effects factors do sometimes appear in social science research. For example, in a repeated-measures investigation (in which repeated measures on the criterion variable are taken from each participant), *participant group* is viewed as a random-effects factor (assuming that they have been randomly selected). Some studies include both fixed-effects factors and random-effects factors. The resulting models are called **mixed-effects models**.

This distinction between fixed versus random effects has important implications for the types of inferences that can be drawn from statistical tests. When analyzing a fixed-effects model, you can generalize the results of the analysis only to the specific levels of the independent variable that were manipulated in that study. This means that if you arbitrarily selected 5 cold calls versus 25 cold calls for your two treatment conditions, once the data are analyzed you can draw conclusions only about the population of agents assigned 5 cold calls versus the population assigned 25 cold calls.

If, on the other hand, you randomly selected two values for your treatment conditions (say 12 versus 32 cold calls) from the population of possible values, your model is a random-effects model. This means that you can draw conclusions about the entire population of possible values that your independent variable could assume; these inferences would not be restricted to just the two treatment conditions investigated in the study. In other words, you could draw inferences about the relationship between the population of the possible number of cold calls to which agents might be assigned and the criterion variable (insurance sales).

Conclusion

Regardless of discipline, researchers need a common language when discussing their work. This chapter has reviewed the basic concepts and terminology of research that will be referred to throughout this text. Now that you can speak the language, you are ready to move on to Chapter 2 where you will learn how to prepare a simple SAS program.