Chapter 1

BASIC CONCEPTS IN RESEARCH AND DATA ANALYSIS

Overview. This chapter reviews basic concepts and terminology from research design and statistics. It describes the different types of variables that may be analyzed, as well the scales of measurement with which these variables are assessed. The chapter reviews the differences between nonexperimental and experimental research, as well as 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, and will be prepared to begin learning about the SAS System 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. Further complicating matters is the fact that, within each discipline, researchers can use a number of very different *methods* to conduct research. These methods may include unobtrusive observation, participant observation, case studies, interviews, focus groups, surveys, ex post facto studies, laboratory experiments, and field experiments.

Despite this diversity in methods used 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 in conducting and reporting their research: researchers in psychology and management speak of "testing null hypotheses" and "obtaining significant *p* values."

The purpose of this chapter is to review some of the fundamental concepts and terms that are shared across the social sciences. You should familiarize (or refamiliarize) yourself with this material before proceeding to the subsequent chapters, as most of the terms introduced here will be referred to again and again 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, it is accurate to say that 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 hypothesis
3. Defining the instrument (e.g., questionnaire, unobtrusive measures)
4. Gathering the data
5. Analyzing the data
6. Drawing conclusions regarding the hypothesis

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 dollar 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 have learned by the time you have completed the program of research. It is good practice to revise and refine the research question several times to ensure that you are very clear about what it is you *really* want to know.

For example, in the present case, you might begin with the question "What is the difference between agents who sell much insurance and agents who sell little insurance?" An alternative question might be "What variables have a causal effect on the amount of insurance sold by agents?" Upon reflection, you may realize that the insurance company really only wants to know what things *management* can do to cause the agents to sell more insurance. This may 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. This narrowing, in turn, leads to a more specific statement of the research question such as "What variables under the control of management have a causal effect on the amount of insurance sold by agents?" Once you have defined the research question more clearly, you are in a better position to develop a good hypothesis that provides an answer to the question.

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

In developing the hypothesis, you may be influenced by any of a 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**. This theory states, among other things, that higher levels of work performance are achieved when difficult work-related goals are set for employees. Drawing on goal-setting theory, you now state the following hypothesis: "The difficulty of the goals that agents set for themselves is positively related to the amount of insurance they sell." Notice how this statement satisfies our definition for an hypothesis, as it is a statement about the relationship between two variables. The first variable can be labelled Goal Difficulty, and the second can be labelled Amount of Insurance Sold.

The same hypothesis could also be stated 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 have been 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."

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 may now test it by conducting a study in which you gather and analyze some relevant data. **Data** may be defined as a collection of scores obtained when a subject's characteristics and/or performance are assessed. For example, you may choose to test your hypothesis by conducting a simple correlational study: you may identify a group of 100 agents and determine

(a) the difficulty of the goals that have been set for each agent

(b) the amount of insurance sold by each agent.

Different types of instruments may be used to obtain different types of data. For example, you may 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 one 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 gathered, you analyze them to see if the agents with the more difficult goals did, in fact, sell more insurance. If yes, the study lends some support to your hypothesis; if no, it fails to provide support. In either case, you could draw conclusions regarding the tenability of your hypotheses, and would have made some progress toward answering your research question. The information learned in the current study may stimulate new questions or new hypotheses for subsequent studies, and the cycle would repeat. For example, if you obtained support for your hypothesis with the current correlational study, you may choose to follow it up with a study using a different method, perhaps 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 this body to draw general conclusions about the determinants of insurance sales.

Variables, Values, and Observations

Variables. When discussing data, one often speaks in terms of variables, values, and observations. For the type of research discussed here, a **variable** refers to some specific characteristic of a subject which may assume one or more different values. For the subjects in the study just described, Amount of Insurance Sold is an example of a variable: some subjects had sold a lot of insurance, and others had sold less. A different variable was Goal Difficulty: some subjects had more difficult goals, while others had less difficult goals. Subject Age was a third variable, while subject Sex (male or female) was yet another.

Values. A **value**, on the other hand, refers to either a subject's relative standing on a quantitative variable, or a subject's classification within a classification variable. For example, Amount of Insurance Sold is a quantitative variable which may assume a large number of values: one agent may sell $2,000,000 worth of insurance in one year, one may sell $100,000 worth, and another may sell $0 worth. Age is another quantitative variable which may assume a wide variety of values. In the sample studied, these values may have ranged from a low of 22 years (for the youngest agent) to a high of 64 years (for the oldest agent).

Quantitative variables versus classification variables. You can see that, in both of these examples, a given value is a type of score that indicates where the subject stands on the variable of interest. The word "score" is an appropriate substitute for the word value in these cases because both Amount of Insurance Sold and Age are **quantitative variables**: variables in which numbers serve as values.

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 the subject may belong. Sex is a good example of a classification variable, as it may assume only one of two values: a subject is classified as either male or female. Race is an example of a classification variable which may assume a larger number of values: a subject may 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 only represent group membership; they do not represent a characteristic that some subjects possess in greater quantity than others.

Observational units. In discussing data, researchers often make references to **observational units**, which may be defined as the individual subjects (or other objects) which serve as the source of the data. 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 data set, and this simply refers to the number of subjects who were studied.

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

Table 1.1

Insurance Sales Data

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

The preceding table reports information about six research subjects: Bob, Walt, Jane, Susan, Jim, and Mack; therefore, the data set includes six observations. Information about a given observation (subject) appears as a **row** running left to right across the table. The first **column** of the data set (running vertically) indicates the observation number, and the second column reports the name of the subject who constitutes that observation. The remaining five columns report information on the five research variables under study. The "Sex" column reports subject sex, which may assume one of two values: "M" for male and "F" for female. The "Age" column reports the subject's age in years. The "Goal Difficulty Scores" column reports the subject's score on a fictitious goal difficulty scale: assume that each participant completed a 20-item questionnaire which assessed the difficulty of his or her work goals. Depending on how they respond to the questionnaire, subjects receive a score which may range from a low of 0 (meaning that the subject's work goals are quite easy) to a high of 100 (meaning that they are quite difficult). The "Rank" column shows how the subjects were ranked by their supervisor according to their overall effectiveness as agents. A rank of 1 represents the most effective agent, and a rank of 6 represents the least effective. Finally, the "Sales" column reveals the amount of insurance sold by each agent (in dollars) during the most recent year.

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

Scales of Measurement

One of the most important schemes for classifying a variable involves its **scale of measurement**. Researchers generally discuss four different scales of measurement: nominal, ordinal, ratio, and interval. Before analyzing a data set, it is important to determine which scales of measurement were used, because certain types of statistical procedures require certain scales of measurement. For example, one-way analysis of variance generally requires that the independent variable be a nominal-level variable and the dependent variable be an interval-level or ratio-level variable. In this text, each chapter that deals with a specific statistical procedure indicates what scale of measurement is required with the variables under study. Then, you must decide whether your variables meet these requirements.

Nominal scales. A nominal scale is a classification system that places people, objects, or other entities into mutually exclusive categories. A variable measured using a nominal scale is a classification variable: it simply indicates the group to which each subject belongs. The examples of classification variables provided earlier (e.g., Sex and Race) also serve as examples of nominal-level variables: they tell us which group a subject belongs to but they do not provide any quantitative information about the subjects. That is, the Sex variable might tell us that some subjects are males and other are females, but it does not tell us that some subjects possess more of a specific characteristic relative to others. The remaining three scales of measurement, however, provide some quantitative information.

Ordinal scales. Values on an ordinal scale represent the rank order of the subjects with respect to the variable being assessed. For example, the preceding table includes one variable called Rank which represents the rank-ordering of the subjects according to their overall effectiveness as agents. The values on 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.

Caution: An ordinal scale has a serious limitation in that equal differences in scale values do not necessarily have equal quantitative meaning. For example, notice the following rankings:

Rank	Name
1	Walt
2	Bob
3	Susan
4	Jane
5	Jim
6	Mack

Notice that Walt was ranked 1 while Bob was ranked 2. The difference between these two rankings is 1 (because 2 – 1 = 1), so there is one unit of difference between Walt and Bob. Now notice that Jim 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 Jim and Mack. Putting the two together, the difference in ranking between Walt and Bob is equal to the difference in ranking between Jim and Mack. But does this mean that the difference in *overall effectiveness* between Walt and Bob is equal to the difference in overall effectiveness between Jim and Mack? Not necessarily. It is possible that Walt was just barely superior to Bob in effectiveness, while Jim was substantially superior to Mack. These rankings reveal very little about the quantitative differences between the subjects with regard to the underlying construct (effectiveness, in this case). An ordinal scale simply provides a rank order of who is better than whom.

Interval scales. With an interval scale, equal differences between scale values do have equal quantitative meaning. For this reason, it can be seen that the interval scale provides more quantitative information than the ordinal scale. A good example of an interval scale 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: The units of measurement are equal throughout the full range of the scale.

However, the interval scale also has an important 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 construct being assessed. It should be obvious that the Fahrenheit scale does not have a true zero point: when the thermometer reads 0 degrees, that does not mean that there is absolutely no heat present in the environment.

Researchers in the social sciences often assume that many of their man-made variables are measured on an interval scale. For example, in the preceding study involving insurance agents, you would probably assume that scores from the goal difficulty questionnaire constitute an interval-level scale; that is, 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. Disagreements concerning the level of measurement achieved with such instruments continues to be a controversial topic within the social sciences.

At any rate, 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 may 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 that zero inches does in fact indicate a complete absence of length. With this scale, 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: 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, as there is no true zero point with that scale.

Although ratio-level scales may 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 manual. 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 described as being either nonexperimental or experimental in nature. In **nonexperimental research** (also called **nonmanipulative** or **correlational research**), the researcher simply studies the naturally occurring relationship between two or more naturally occurring variables. A **naturally occurring variable** is a variable which is *not* manipulated or controlled by the researcher; it is simply measured it as it normally exists.

The insurance study described previously is a good example of nonexperimental research in that you simply measured two naturally occurring variables (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 designs, researchers often refer to criterion variables and predictor variables. A **criterion variable** is an outcome variable which may be predicted from one or more predictor variables. The criterion variable is often the main focus of the study in that 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 values on the criterion. In some studies, you may 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 may positively affect insurance sales, you conducted a study in which Goal Difficulty was the predictor and Sales was the criterion. You do not necessarily have to believe that there is a causal relationship between two variables to conduct a study such as this; however, you may simply be interested in determining whether it is possible to predict one variable from the other.

You should note here that nonexperimental research that investigates the relationship between just two variables generally provides relatively weak evidence concerning cause-and-effect relationships. The reasons for this can be seen by reviewing the study on insurance sales. If the psychologist 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 causes them to set higher work goals for themselves. Under this second scenario, it was actually the insurance sales which had a causal effect on Goal Difficulty.

As this example shows, with nonexperimental research it is often possible to obtain a single finding which is consistent with a number of contradictory causal explanations. Hence, a strong inference that "variable A had a causal effect 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 (such as path analysis), 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:

- Subjects are randomly assigned to experimental conditions.
- The researcher manipulates an independent variable.
- Subjects in different experimental conditions are treated similarly with regard to all variables except the independent 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 subjects. You randomly assign 50 agents to a "difficult goal" condition. Subjects in this group are told by their superiors to make at least 25 cold calls (sales calls) to potential policy-holders per week. The other 50 agents have been randomly assigned to the "easy goal" condition. They have been told to make just 5 cold calls to potential policy holders per week. .

After one year, 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 which could take on a wide variety of values (whatever score the subject 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 each subject. In the experiment, Goal Difficulty could assume only one of two values: subjects were either in the difficult goal group or the easy goal group. 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 variables and dependent variables. The **independent variable** (abbreviated 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** (abbreviated DV), on the other hand, is some aspect of the subject's behavior which 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 may be used with almost any type of research, but that the terms independent 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 a subject may 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 subjects that does 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 U.S., 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 that 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 mean, you are engaging in descriptive statistical analysis. **Descriptive** statistical analysis focuses on the exhaustive measurement of population characteristics: You define a population, assess each member of that population, and arrive at some form of summary value (such as a mean or standard deviation).

Most people think of populations as being very large groups, such as "all of the people in the U.S." 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 may define as a population all 23 students taking an English course, and then calculate the average score of these students on a measure of class satisfaction. The resulting average would be a parameter.

Inferential analyses. A **statistic**, on the other hand, is a numerical value which is computed from a sample, and either describes some characteristic of that sample such as the mean, 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 statistics, **inferential** statistical analysis involves using information from a sample to make inferences, or estimates, about the population. For example, assume that you need to know how much insurance is sold by the average agent in the company. It may not be possible to obtain the necessary information from all 10,000 agents and then calculate a mean. An alternative would be to draw a random (and ideally representative) sample of 100 agents and determine the average amount sold by this subset. 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. Here, you have used characteristics of the sample to make inferences about characteristics of the population. Using some simple statistical procedures, you could even put confidence intervals around the estimate which would allow you to make statements such as "there is a 95% chance that the actual population mean lies somewhere between $172,994 and $185,650." This is the real value of inferential statistical procedures: they allow you to review information obtained from a relatively small sample, and then make inferences about a much larger population.

Hypothesis Testing

Most of the procedures described in this manual 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. Assume that, after one year, the difficult-goal agents had sold an average of $156,000 worth of insurance, while the easy-goal agents had sold only $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 *enough* more to make you confident that the difference was due to your manipulation?

What's more, one could easily argue that you don't really even *care* about the amount of insurance sold by these two small samples. What really matters is the amount of insurance sold by the larger populations which 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 two populations differ with respect to their mean 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 mean scores on a locus of control scale. Notice that in both cases, two or more distinct populations are being compared with respect to their mean scores on a single criterion variable.

With a **test of association**, on the other hand, you are working with a single population of individuals and wish to know whether there is a relationship between two or more variables within this population. 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, subjects could receive a score which may range from a low of 0 to a high of 100. You could then correlate these goal difficulty scores with the amount of insurance sold by the 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. Notice why this would be called a test of association: you are determining whether there is an association, or *relationship*, between the predictor and criterion variables. Notice also that there is only one population being studied; there is no experimental manipulation which creates a difficult-goal population versus an easy-goal population.

For the sake of completeness, 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 two variables is the same across two or more populations. Analysis of covariance (ANCOVA) is one procedure that allows such a test. For example, you may form a hypothesis 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 hypothesis, 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 could 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). Although analysis of covariance is beyond the scope of this text, interested readers are referred to Littell, Freund, and Spector (1991) for instructions on how it may be performed using the SAS System.

Types of hypotheses. Two different types of hypotheses are relevant to most statistical tests. The first is called the null hypothesis, which is often abbreviated as H_0. The **null hypothesis** is a statement that, in the population(s) being studied, there are either (a) *no differences* between the group means, or (b) *no relationships* 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.

With a test of group differences, the null hypothesis states that, in the population, there are no differences between any of the groups being 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 with symbols in the following way:

$$H_0: M_1 = M_2$$

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) which states the opposite of the null. The **alternative hypothesis** is a statement that there *is a difference* between the means, or that there *is a relationship* between the variables, in the population(s) being studied.

Perhaps the most common alternative hypothesis is a **nondirectional alternative hypothesis**. With a test of group differences, a nondirection alternative hypothesis predicts that the means for the various populations will differ, but makes no specific prediction as to which mean will be relatively high and which will be relatively low. 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 mean 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 alternative hypothesis** makes a more specific prediction regarding the expected outcome of the analysis. With a test of group differences, a directional alternative hypothesis not only predicts that the population means differ, but also predicts which population means will be relatively high and which will be relatively low. Here is a directional alternative hypothesis for the preceding experiment:

H_1: The average 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 may be symbolically represented with the following:

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. A directional alternative hypothesis, on the other hand, 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 *one-sided* statistical tests (also called one-tail tests), which are relatively powerful. Here, "powerful" means that one-sided tests are more likely to find significant differences between group means (for example) when differences really do exist. In contrast, nondirectional hypotheses allow only *two-sided* statistical tests (also called two-tail tests) which 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 grounds. For example, you should state the directional hypothesis that "The average 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 **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 the SAS System, you will review the SAS output for two pieces of information which are critical for this purpose: the obtained statistic, and the probability (p) 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. When the data analysis for this study has been completed, you would review a t statistic and its corresponding p value. The p value indicates the probability that one would obtain the present results if the null hypothesis were true. If the p value is very small, you will reject the null hypothesis.

For example, assume that you obtain a t statistic 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 so 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 the two groups.

On the other hand, assume that the research project instead produces a t value of 8.45 and a corresponding p value of .001. The p value of .001 indicates that there is only one chance in 1000 that you would obtain a t value of 8.45 (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, differ on mean sales. 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 may 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 (.001, 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. When they obtain a *p* value *smaller* than .05 (such as .04 or .002 or .0001), they will reject the null and conclude that differences or relationships being studied were 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 may 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 variable (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 decided that the subjects in your easy-goal condition would be told to make just 5 cold calls per week, and that the subjects in the difficult-goal condition would be told to make 25 cold calls per week. In this case, you have *fixed* (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. In conducting your study, one group of subjects 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, not fixed.

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 subjects who suffer from phobias. Specifically, you wish to 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 labelled 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 Caucasian Americans score significantly higher than African Americans on internal locus of control. The predictor variable in your study would be Race, while the criterion variable would be scores on some index of locus of control. In all likelihood, you would have arbitrarily chosen "Caucasian American" versus "African American" as the groups under your predictor variable because you are particularly interested in these two races; you did not randomly select these groups from all possible races. 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 subject), subjects are 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 may 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 may draw conclusions only about the population of agents assigned 5 cold calls versus the population assigned 25 cold calls.

On the other hand, if 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 may 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 that agents may be assigned, and the criterion variable (insurance sales).

Conclusion

Regardless of discipline, researchers need a common language to use when discussing their work with others. 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 submit a simple SAS program.

References

Littell, R. C., Freund, R. J. & Spector, P. C. (1991). *SAS system for linear models, third edition.* Cary, NC: SAS Institute Inc.

Yaremko, R. M., Harari, H., Harrison, R. C., & Lynn, E. (1982). *Reference handbook of research and statistical methods in psychology: For students and professionals.* New York: Harper & Row.