1

Why Do We Use Statistics?

Basic Concepts We Need to Know

Jamie has been collecting data on client outcomes for her Employment Support Program, which is designed to improve the chances of employment for a high-risk population. She compared clients who had completed the program with those on the waiting list for service, looking at whether the client had obtained employment. A year ago, she found that 60% of her 10 clients had obtained employment as compared to 40% of the 10 clients on the waiting list. Recently, she did another study and found that the 9 clients who had completed service did not have a higher rate of employment than those on the waiting list. She then compared her results with those of Parker, who runs a different employment program. He found a year ago that 65% of his 45 clients had obtained employment as compared to 45% of those on the waiting list.

Do you want to guess how well Parker's clients are doing now? If his clients are still doing better than people who have not participated in his program, then why might his results have stayed the same while Jamie's seem to have changed? Is there a difference between the effectiveness of their work over the last year?

You may have noticed that Parker had a much bigger number of clients in his study. Therefore, he can be more confident that his recent data will show a better outcome for his clients than those on the waiting list. Because Jamie's sample sizes are smaller, her data are more likely to fail to be statistically significant. In other words, her initial results could too easily be explained by chance rather than something we can depend on to make predictions. If you have a large number of clients rather than a small number of clients and you find a modest difference that favors the clients, then when you conduct another study later, you are more likely to find your client group still has a better outcome than nonclients. Your results are more likely to be consistent because you have to a greater degree ruled out chance as the explanation for your data.

Why We Use Statistics

Statistics refers to the collection and analysis of data in a way that helps us make decisions. Should we continue the Support Group for Victims of Abuse? What about Jamie's Employment Support Program? To answer these questions, we need to collect data on client growth and draw conclusions about the future effectiveness of the program. We might find that it is a great success, which means we should tell others about it. We might also find that it is a modest success or no success at all.

Where can statistics help us in this endeavor? We might find that a small number of clients had a small amount of growth. If so, can we be confident that a repeat of this program would bring modest success to another small number of clients? It would probably not be wise to bet our money on repeat success in this situation. Why? Because a small amount of growth experienced by a small number of people is not likely to be statistically significant. This means that we cannot adequately say that our treatment, rather than chance, is the explanation of our data.

Let's examine the issue of **chance**. Suppose someone came into a room with a number of people and announced that he could tell whether or not someone was right-handed or left-handed by looking intensely into their eyes. Suppose further that someone else in the room said "Show me," to which this person looked into her eyes and said, "You are right-handed." Would you be a believer in this person's ability to use intense eye inspection to detect whether someone was right-handed or left-handed? I would think not. Why? Because only about 10% to 15% of people are left-handed, so someone who guesses that another person is right-handed has the odds on his or her side. In other words, such a person could be right just by chance.

Suppose that you are examining a counseling program for at-risk youth at your middle school. You have been providing counseling to help students improve their confidence in the belief that they will someday graduate from high school. You have been collecting data using a special scale that measures confidence in continuing in school. To examine the effectiveness of this program, you have collected data at the beginning and end of the counseling. Suppose that the mean (average) score for confidence is 21.2 at the beginning of the program and the mean score after 3 months is 25.6. Given that higher scores on this scale mean more confidence, you have some evidence that your program is effective, especially if you can adequately rule out chance as the explanation for the increase in mean score. You rule out chance as a good explanation by subjecting your data to statistical analysis using a statistical test.

Now, what if you measured a single client on anxiety once a week during a period before treatment and did the same for the client during the treatment period? Suppose you found the following scores on your anxiety scale (where higher scores mean more anxiety): 24, 25, 21, 19, 18. These scores were going down before treatment began, indicating that the client's anxiety was lessening. Suppose further that the weekly anxiety scores during the treatment period were 18, 19, 17, 17, and 16. The mean of these treatment scores is 17.4, while the mean of the previous scores is 21.4. This might seem to indicate that your treatment was a success, but don't forget that the scores were going down before treatment. If we were to project the baseline scores (taken before treatment) into the treatment period, we would see a pattern that looks a lot like the treatment scores, suggesting that the treatment did not make a difference. To demonstrate otherwise, we would need to see a pattern of treatment scores that was distinctly different from the projected trend. In this situation, perhaps we would conclude either that treatment was not needed (the client was already improving) or that a different target behavior should be the focus of the service provided.

When we evaluate human services, we could make a mistake: We could conclude that our treatment has been effective because our clients have had a measured improvement in scores, when, in fact, we should conclude that our data can be easily explained by chance. In this event, we should not be confident that applying this treatment with other clients will bring about similar success. In other words, before continuing to use the treatment, we should use statistics to test whether chance is the best explanation of our data.

And let us not forget the issue of accountability in the administration of human services. Increasingly, those who hold the "purse strings" expect data to be used to justify the continued expenditure of funds. Statistics will naturally be a big part of this task.

What You Will Find in the Rest of This Chapter

This chapter will give a preview of what this book is about and discuss some basic concepts that guide the statistical analysis of data. While the focus is upon the analysis of data for evaluative studies, the information in this book will also be useful for descriptive and explanatory studies in which you are, respectively, describing your clients or examining the relationships between client characteristics (e.g., Do older clients experience more growth?). In addition, statistics useful for evaluation of human services are useful for studies in other fields. However, the primary focus here is on evaluation research, in which you evaluate the outcomes of an intervention, and the selection of statistics for examination is guided by this interest.

There is a quiz at the end of this chapter which you can use for two purposes. First, you could take it right now to see whether you are already familiar with the concepts in this chapter. Second, you could test yourself after you have read the chapter to see whether you need further review. There is also a chapter glossary, which you can use to review and test your understanding of key terms.

This book is designed to be an extremely user-friendly approach to using statistics for answering research questions. The goal is that you will learn how to use statistics by engaging in concrete tasks related to data of interest to you. If you employ the statistical concepts and methods demonstrated in these pages to conduct studies, you will learn about the essentials of statistics in a way that will likely stay with you after this learning expedition is done.

Illustrated in this chapter is a process for data analysis:

1. What is our research question?
2. What is the structure of our data?
3. What statistical test will we use?

4. How do we use the computer for statistical analysis?
5. What are our conclusions?

This process will be illustrated in this chapter and should reveal the user-friendly approach to statistics that is employed throughout this book. But first, let's examine some basic concepts to get familiar with the nature of statistics.

Two Key Issues Addressed by Data Analysis—Practical Significance and Statistical Significance

The statistical examination of data addresses two fundamental issues:

1. Are my data clinically noteworthy?
2. Are my data easily explained by chance?

The first question refers to *practical significance* while the second one refers to *statistical significance*. We have discussed the basic concept of statistical significance in the first section of this chapter. Now we will expand on that discussion of statistical significance and add the theme of practical significance.

Statistical Significance

As previously illustrated, a major issue in regard to **statistical significance** is the extent to which your data can be explained by chance. How much can you depend on your data to be telling you about something real, rather than a phenomenon that occurred just by chance?

The statistical tests you'll learn about in this book will yield a value of $\boldsymbol{p}$. The value of p can range from 0 to 1.0. A value of 0 means that chance does not explain your data at all, while a value of 1.0 means that chance explains your data completely. In other words, a p value of 1.0 means all you have is chance, not data that can be treated seriously. When you do studies and when you read other people's research, you will not likely see p values of 0 or 1.0. Indeed, these values are so rare that you should question such a result if it pops up on your computer. Instead, you will typically see values like .23 or .67 or .03 and so forth.

You have probably seen the expression "$p < .05$" in research reports. This statement means that these data would be expected to occur by chance fewer than 5 times in 100, and such data are normally held to have met the standard for statistical significance in the social sciences. So, if you employ this standard in your research, you will be hoping for a p value of less than .05 in order to say that your data support your expectations (i.e., your hypothesis).

Your **hypothesis** is your educated guess about what you will find when you analyze your data. You might state your hypothesis as "Posttest scores for depression will be lower than pretest scores." You would then collect pretest and posttest scores and subject these data to statistical analysis. If the posttest scores were lower and the difference between pretest and posttest scores were statistically significant, you could say your data supported your hypothesis.

The standard of .05 for stating that data support a hypothesis has been advocated by those who write about research and statistics and has been widely accepted. However, a decision to use the standard of .10 (10 times in 100) instead of .05 would be just as scientifically valid because the standard is a matter of judgment. When making this judgment, we should consider the relative advantage of being more or less conservative in the interpretation of our data. If we are engaging in evaluative research and wish to make sure we do not conclude that an ineffective treatment is effective, we will be more conservative and employ a low value of p as our standard. We might use the standard of .05 or an even more conservative standard of .01. On the other hand, if we want to make sure we do not erroneously conclude that an effective treatment is not effective, we would choose a less conservative standard, like .10. Be aware, however, that the further you depart from the standard of .05, the less likely you will find agreement among those who read your reports.

As indicated above, the standard you select should be guided by the kind of risk that is more important, given what you are studying. In research about drugs, you are likely to find a conservative standard, maybe .01 or even .001. That is because of the danger inherent in a false positive (i.e., believing an ineffective drug is actually effective).

If you are interested in what determines statistical significance, read Insight Box 1.1, where sample size, magnitude of data, and variance are discussed.

Insight Box 1.1

What Determines the Level of Statistical Significance?

There are three things that determine the level of statistical significance for your data—sample size, magnitude of the data, and variance. Sample size is the number of people you have collected data on (or the number of pieces of data). Magnitude deals with the question of how much change has occurred. Variance measures the extent that the scores on your key variables are either similar to each other or different from each other.

Sample Size

Sample size is the number of people you have in your study. The larger the sample size, the better the level of statistical significance. You can have data showing that 60% of your clients got better while only 30% of those in the comparison group got better, yet fail to have statistical significance because your sample is very small. However, take these same proportions (60% vs. 30%) with a much larger sample, and you will likely have statistical significance.

(Continued)

(Continued)

Suppose, for example, that we use a statistic called chi square (which will be discussed later in this book) to compare data on a group of clients to that on a group of similar people who have not yet had the service. We compare them on the basis of whether each person in each group got better during the treatment period (Yes or No). Some in each group got better, so the question is whether the difference between the two groups is statistically significant. Suppose 6 people in our treatment group got better and 4 did not, and 3 people in our comparison group got better and 7 did not. Thus, we find that 60% of those in our treatment group improved while only 30% of those in the comparison group improved. We might see this difference as noteworthy because our clients were twice as likely to improve. But when we submit these data to statistical analysis, we find that chance is a good explanation of the difference between the treated group and the comparison group. In other words, these data failed to achieve statistical significance.

So how does sample size make a difference? Let's say our data give us the same proportions to compare (60% vs. 30%), but now we have a much larger sample size. Suppose that we had 60 clients who got better and 40 who did not. Thus, 60% got better just as in the previous example. And, suppose we had 30 people in the comparison group who got better and 70 who did not. For the comparison group, 30% got better, again just as in the previous example. Now when we compare these data using chi square, we find the data to be statistically significant.

Magnitude

Magnitude refers to the amount of difference or the strength of the relationship. How much difference is there between two groups? In the previous example, we compared the improvement rate of 60% to the improvement rate of 30%. If 80% of the clients had gotten better while only 30% of the comparison group people did the same, we would have a greater magnitude. There is a bigger difference when we compare 80% to 30% than when we compare 60% to 30%. Magnitude measures difference in a number of relationships: the extent of the difference between two groups (e.g., the number who are Democrats vs. the number who are Republican), the extent of change from pretest to posttest for clients receiving a service, or the strength of the relationship between variables. It answers the question, *How much change has occurred?*

Variance

The extent that data fluctuates from one score to another is known as variance. Suppose you are serving middle school students at risk of dropping out and your service is designed to improve their grades, because poor grades have been found to influence dropping out. Before you start your service, you collect the grades (given numerically) for all six of the students in your caseload for each of the previous 6 months. The average grades for these clients for the prior 6 months are 56, 61, 55, 63, 49, and 55. As you can see, the grades have ranged from a low of 49 to a high of 61, but the differences from one month to another are not dramatic. A small amount of fluctuation is normal. The average of these numerical grades is 56.5. But, more importantly, the average amount that each score differs from the average overall score is 3.66. In other words, if you subtracted each score from the average and got the average of these deviations, you would have a value of 3.66. You might say that the variance in these scores is small. Now what if you had these grades from six students—48, 52, 63, 77, 84, and 88? Right away, you can see that the variance is greater. The average of the deviations of scores for this group from the average of all scores is 14.33, which is much higher than the 3.66 for the previous set of grades.

The important consideration in regard to statistical significance is that the first set of grades, with low variance, is more likely to generate a finding of statistical significance than is the case of the second set of grades with higher variance. So, the less the variance in scores, the more you can depend on any difference between data sets as not being the result of chance.

Let's summarize what influences statistical significance by holding two of these three factors constant and comparing change on the third. So here goes.

- If you have a larger sample, you are more likely to achieve statistical significance.
- If your data show a higher magnitude, you are more likely to achieve statistical significance.
- If you have less variance, you are more likely to find statistical significance.

And, of course, if all three of these things are in your favor compared to another study, you will have a better level of statistical significance than that study.

Practical Significance

Practical significance refers to whether the data are noteworthy in your professional opinion. Going back to our previous example, in which your hypothesis is "Posttest scores for depression will be lower than pretest scores," suppose that 60% of your clients improved during the treatment period. Is this good news or bad news? What if you also measured scores of a group of people who did not receive the treatment, and only 30% of them improved? Now can you conclude that your treatment works? Or suppose your agency placed 30% of all people served in your homeless shelter into permanent homes: Good news or bad news? Practical significance focuses on the **magnitude** of the data and considers such issues as the amount of resources expended to achieve the result.

Practical significance is a matter of informed opinion and is not as concrete as statistical significance. There is no clear guide for determining whether or not practical significance has been achieved. Thus, reasonable people can have different opinions about practical significance, and it can be the subject of debate.

In this book, you will be given some guidance that you can use to develop your own opinion about practical significance. One standard to consider is whether the gain in client functioning moved the typical client in your study from one threshold of functioning (e.g., being severely depressed) to a better threshold (e.g., being mildly depressed). Another consideration is the **effect size**, which is the amount of gain measured for your clients expressed in terms of standard deviations of gain. Effect size, which will appear throughout this book, can be used to compare studies to one another, even if they are of different kinds and used different measurement tools. There also are suggested guidelines for what effect size value is "high," and even though these determinations are only the opinions of certain people, they can be useful.

Because practical significance is a matter of opinion, your critical task in writing your research report is to articulate how you came to your opinion. There is no clear basis for someone to argue with you about whether you are correct in your assessment of a study's practical significance, but people may argue with you about what guidelines you used to come to your conclusion.

You can explore the concept of practical significance by considering the following situation. First, examine the data. Then determine on what basis you will decide whether practical significance was achieved. Finally, share your information with someone else who has done the same analysis. The two of you can either agree and confirm why you agree or debate your differences.

What Change in Scores Would Be of Practical Significance?

Let's suppose we have scale of nonphysical partner abuse that measures the extent to which one person believes his or her partner is abusive in nonphysical ways. On this scale, the respondent answers either yes or no to a set of 10 items such as these:

- My partner belittles me so much that it is really irritating.
- My partner becomes hostile if my work is not done when he or she thinks it should be.
- My partner is hostile about new friendships I develop.
- My partner becomes very angry if I disagree with him or her.

All items are worded in the negative direction, as illustrated by the four example items above. The respondent gets 1 point for each answer of yes. Therefore, the highest possible score is 10, and the lowest possible score is 0.

The question here is how much gain would represent practical significance for a couple being counselled for the problem of nonphysical abuse. Let's suppose that in this couple, the husband has been accused of being abusive and the couple has been provided with 10 therapy sessions over a period of 10 weeks. Suppose you had administered this scale at the beginning of the treatment period and again once per week during the treatment period. The pretest score was 7. How much change would you need to see to view the difference as being of practical significance: a change of 1 point, 2 points, 3 points, 4 points . . . ? After you decide on a number, turn your attention to the more important issue for this exercise: What is the basis for your decision that the number for practical significance is the one you selected?

Again, your final task is to compare your answers with those of someone else who has done the same thing. Compare both how much change you think is needed to determine the difference is of practical significance and your rationale for your choices. Does one of you have a more cogent rationale? What can you learn from articulating your standards for practical significance?

You should now have an idea of the two key things that data analysis, facilitated by what you learn in this book, can do for you: It will help you to answer the questions of practical significance and statistical significance. When you see the magnitude of your results, you will be able to determine whether the results are more likely due to chance than to your program. And if the results are due to the treatment, you can reasonably say whether the outcome is of practical value.

A key reason to examine both issues is that you can have statistical significance without having practical significance. That is, your data may show a small gain that is statistically significant. Statistical significance means you can treat the results as real, not due to chance. But if the gain is small, you have to make a judgment about whether it is really noteworthy. Practical significance, though a matter of opinion, is informed by what you know about the tools being used and the target behavior being treated. Thus, you should be prepared to offer a good rationale for your opinion.

Using Statistics to Describe Clients, Evaluate Services, and Explain Client Behavior

Human services research can be divided into three categories based on the purpose of the inquiry. These purposes are description, explanation, and evaluation.

The *descriptive purpose* is achieved by simply describing social phenomena. What is the distribution of our clients by race, age, and gender? How many of our clients have a preschool child?

The *explanation purpose* focuses on the relationships between variables. The researcher examines the data to see whether one variable might explain another. Do females achieve greater gain in our support group program than males?

The *evaluation purpose* seeks to determine whether human services were successful. How much gain did our clients achieve in regard to the target behavior?

Thus, we can say that statistics can be useful for descriptive research, evaluative research, and explanatory research. Regardless of the type of research being done, a key concept in the analysis of data is "What are our study variables?"

Variables

For all three purposes of research, you will be using statistics to examine variables. A **variable** is something that varies, whereas a **constant** is something that does not vary. If you measure depression using the Beck Depression Inventory, you will have depression as a variable. If you measure it both before and after treatment, you will have two variables, the pretest score and the posttest score. If you ask respondents to indicate their gender, you will have gender as a variable. If you do not measure whether or not a respondent is a student, you will not have the concept of student as a variable. And, of course, if all of your respondents are students, you could not possibly have student as a variable because all participants are all alike in that way.

Some Examples of Types of Research

Let's look at some examples of descriptive research, evaluative research, and explanatory research. We might want to know the mean age of our clients or the proportion of our families that have a preschool age child or the number of clients who have expressed a need for our women's support group service. The purpose of such research is descriptive, so we could label this **descriptive research**, and we would apply descriptive statistics.

Another purpose for research in the human services is to explain social phenomena by examining the relationship between variables. We might want to know whether female clients in our afterschool tutoring program achieve better gain in grades than the male clients. We might want to know whether clients who attend more support group meetings have greater gain in feelings of support. This is research that could be labeled **explanatory research** because it is used to explain something. For this research, we would normally employ inferential statistics.

A third purpose for research would be the evaluation of the success of our services. We may want data on the level of gain for our clients in a given service. Do the clients in our afterschool tutoring program achieve better grades in school, and if so, how much better? We could call this **evaluative research** because we use it to evaluate our services. Technically, we could say that evaluative research is a form of explanatory research because it is designed to see whether treatment explains client gain. But we will distinguish these two types of research in this book because of the special role that evaluative research plays in the human services. As a type of explanatory research, evaluative research also employs inferential statistics, because we need to infer something about a larger population of clients based on data from a sample of our clients. Thus, we ask, *To what extent can we say our findings are relevant to a population greater than our sample?*

This book will help you to examine data that are relevant to each of these three types of research. You will learn how to engage in descriptive research by computing statistics for client variables like age, gender, race, and so forth. You will learn how to execute explanatory research by examining the correlation between variables, such as number of sessions attended and gain on outcome. The primary purpose of this book, however, is to provide you with a user-friendly guide to conducting evaluative research. You will learn how to test your hypotheses when you evaluate human service outcomes. You will examine, for example, how to compare pretest and posttest scores for depression to test the hypothesis that posttest scores will be lower.

Descriptive and Inferential Statistics

One of the things that a statistic helps us do is summarize the data from a sample of people, like all the clients served by our agency. We could calculate the mean age of our clients or the proportion of the families we serve who have a preschool-age child. These data are called **descriptive statistics**. The purpose is to describe a sample (i.e., the people in your study). The various characteristics that we want to describe are called variables. So, we might have the variables of age and race and grade in school and so forth.

Another thing that a statistic can help us to do is to estimate the likelihood that our data would occur by chance when we want to infer from a sample to a population. The **sample** is the people from whom data were collected, whereas the **population** is a larger group of people that includes the individuals in the sample. When we want to use a sample to infer the characteristics of a population, we use **inferential statistics**. Are the data from our sample a good estimate of the characteristics of a larger population from which the sample was drawn? If we draw a sample of 15 of the clients currently served by our agency and find they achieved a 30% gain on school grades during the service period, how likely is it that these gains occurred, not due to chance, but rather due to the program? If the change is due to the program, then we can infer that other clients of our agency (or other agencies) could be expected to experience similar gains if they received the same service.

We use inferential statistics when we are testing a hypothesis in evaluative research. The hypothesis is a prediction of our results, for example, "The clients' grades during the treatment period will be higher than their grades prior to treatment." To use another variable as an example, suppose we measure the gain of our clients on self-esteem and compare it to data from a comparison group of similar people who did not receive our service. We would first want to know whether our clients had a greater gain. If our clients had a greater gain, then we want to know how likely it is that the difference is due to mere chance. Traits like self-esteem fluctuate naturally in populations, whether or not the individuals in the population have received a service designed to improve it. This is where chance comes into play. Can our measured data be best explained as normal fluctuations in behaviors? If so, we would say that our data is explained by chance. However, if our data are not easily explained by chance, we can feel confident that the self-esteem gains indicated by our data would likely occur in similar studies of other people in the larger population from which our sample was drawn. In other words, we are in a better position to bet that our results would be repeated with others.

A key difference then between descriptive statistics and inferential statistics is that descriptive statistics are typically used to portray a sample in its entirety, such as all the clients in our study. Inferential statistics, on the other hand, relate to a sample of a larger population, and we are trying to infer something about the population from studying a sample of those people. It is with inferential statistics that we encounter the idea of statistical significance.

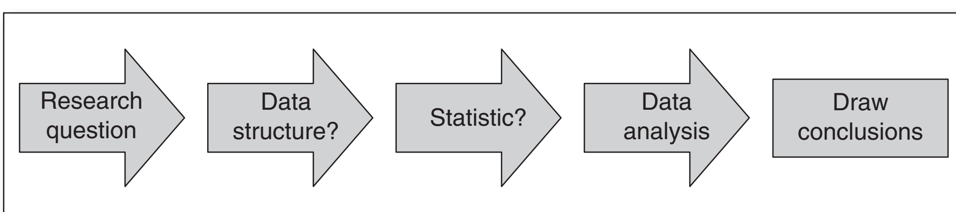
How Do We Analyze Data the User-Friendly Way?

Let's examine the process for statistical analysis in simple terms. First, you decide what you need to know. The answer to this question will be derived from the purpose of your study or your basic research question. Next, you need to know what your data looks like. We will refer to this as the structure of your data. This means, of course, that you have collected data after deciding how to measure the variables in your study. Third, you will find a statistic that will help you to answer your research question based on your knowledge of the structure of your data. Fourth, you will apply your statistic to your data using the computer. In this book, you will be given several special files using the Excel software that you can use in a very user-friendly fashion. Finally, you draw conclusions that do not outrun your data. Let's see this in graphic form in Exhibit 1.1.

We will examine each of these steps starting with the nature of our research question, or the basic purpose of our study. In our example, we will be using *evaluative research* because we want to know whether a service is effective. To illustrate the analysis of data on homelessness, we have, hypothetically, collected data on the 15 people who entered our homeless shelter 60 days ago. Each of these people were given services designed to overcome personal obstacles to getting a home and services designed to help them make arrangements to get a home. Everyone was homeless upon entry, meaning that homelessness did not vary upon entry into our shelter, so using a pretest and posttest measure of homelessness is not warranted. **Variance** is a critical issue in the application of statistics.

We only have data at the end of a service period, which means we only have posttest data. At first glance, this may not seem to be a situation in which statistical analysis could be useful. But at least one statistic can help, if we have a **threshold** score for the comparison of our posttest data. Let's suppose, for example, that we find the proportion of homeless people nationally who get a home within 60 days of entry into a homeless shelter is 15%. We can use a statistic that can compare our posttest data to this figure. Let's see how this works in regard to the process of data analysis, starting with the research question.

Exhibit 1.1 The Steps of Data Analysis



1. Our Research Question

Is the proportion of our clients who received a permanent home within 60 days of entry into our homeless shelter greater than the national average?

2. The Structure of Our Data

The *structure of data* refers to what data look like for the purposes of selecting a statistical test. Among the things to consider are these: (1) At what level are our variables measured? (2) Do we have **related data** (**matched data**) or **independent data**? (3) What is our research design? These questions will be addressed in full in later chapters, but we will quickly illustrate them here for our particular example.

- We will be collecting data that is measured at the nominal level because study subjects are simply placed into categories (Yes or No for finding a home in 60 days) that have no order from low to high.
- This is not an example of matched data because we did not give each person two scores on an instrument, at two points in time, that can be compared. Instead, we have data independence.
- We are using the posttest-only research design, something you will not often see in evaluative research.

3. The Statistic That We Can Use

In Appendix A, you will find a comprehensive discussion of how to find a statistic for your data. Exhibit A.3 shows how to find a statistic for posttest-only data when you have a threshold for comparison. In our example, we will be comparing our data to the national average proportion of homeless people who find a home on their own. Later, you will go through this process in more detail, but for now we will show the process in brief to illustrate using statistics.

If you had gone through the decision tree in Exhibit A.3, you would indicate that your threshold is nominal and dichotomous (i.e., only has two categories) and that your dependent variable (your client data) is also measured as a dichotomous variable. By following this path, you would come to the binomial test as the statistic for your data. The binomial test compares the proportions of people in two categories to a threshold to see whether the people in your study have significantly different values than that threshold. If you assumed that the people nationwide were distributed equally in the two categories, you would compare your data to the 50% threshold. Your hypothetical data, however, show that about 15% of people who enter a homeless shelter find a home within 60 days. So, 15% is your threshold. First you collect your data. Then you see whether the proportion of clients finding homes is greater than 15%. If the proportion is greater than 15%, you use the binomial test to determine how likely the better results for your clients are due to mere chance. If you find that the difference between your clients' results and the outcomes for homeless people nationwide is statistically significant, then you know the answer to your research question (Step 1) is *yes*.

4. Data Analysis With the Excel File

Later in this book, you will learn how to use a special Excel file (designed just for this book) named *York, binomial, posttest compared to a threshold, dichotomous data*. You will simply load Excel on your computer, if it is not already there; download this special file; and enter your data. You are encouraged to load this file onto your computer now and enter the data that follows in order to get a preview of how using Excel to analyze data works. The data are as follows: Of the 15 people who entered the shelter 60 days ago, 10 have found homes after 60 days.

The special Excel file asks you to enter the following data:

1. The number of people who had favorable results, 10
2. The total number of people in the sample, 15
3. The threshold fraction for comparison (the national proportion), 0.15

From the previous discussion of statistical significance, you learned that the value of p indicates the number of times out of 100, or the percentage of times, the data would be explained by chance. Thus, lower scores are better in the sense that they mean our research hypothesis is more likely to be true. As noted previously, the usual standard in the social sciences for ruling out chance as a good explanation of the data is .05 (5 times in 100). When you enter your data, you find that the p value is less than .001, which clearly meets the standard in the social sciences for statistical significance ($.001 < .05$). So, you can conclude that your 66% success rate is significantly higher than a 15% rate, given the sample size. This comparison of proportions, however, is relevant only to this specific data. A comparison of a 66% rate and a 15% rate may not be significant with a smaller sample size.

5. Conclusions

We found that our success rate was better than the national average. Because we achieved statistical significance, we can rule out chance as the explanation of our superior success. Whether we have achieved practical significance, of course, is a matter of opinion.

The remainder of this book employs the same user-friendly approach to statistical analysis as was demonstrated in this chapter. This book will help you to analyze your data when you have data to analyze. It will teach you what various statistics do by getting you involved in using the statistics. You will be given examples and asked to enter data into the computer and complete the statistical analysis. Then you will explain your findings. To explain your results, you will need to understand the particular statistic you used and the theme of that chapter. So, you will learn by doing. It is not assumed that you really enjoy examining statistical formulas as a math student would likely do. Instead, this text assumes that you have a need to know how to use statistics to test whether your evaluative data are significant.

In Chapter 2 is an introduction to the use of the computer for data analysis. You have seen from the above material how using the computer plays an important role in the overall process of data analysis.

What You Will Learn From This Book

At the completion of this book, you will be able to use statistical analysis to answer your research questions. On the path to this achievement, you will learn how to select a statistical measure in light of your research question. You will also learn how to enter your data into the computer to facilitate statistical analysis and reporting of the results. In your reports, you will show an understanding of what statistical results mean in regard to both practical significance and statistical significance.

The following are among the specific competencies you will acquire:

- Report the descriptive characteristics of your study subjects. Examples include the frequencies and proportions of people in the category of gender or the mean, median, mode, and standard deviation of variables like age.
- Test your explanatory research hypotheses, in which you examine the relationship between variables, and report the results, demonstrating your understanding of how to use statistical tests to address the issue of chance.
- Test your evaluative research hypotheses, in which you examine client progress to evaluate the outcomes of your interventions, demonstrating your understanding of relevant statistical tests.

Several principles guide the material in this book, as discussed in Insight Box 1.2.

Insight Box 1.2

Some Principles That Guide the User-Friendly Approach to Statistical Analysis

This book employs a user-friendly approach to statistical analysis of data for the evaluation of human services. Therefore, it is designed for work on the front line of human services—to conduct the kind of analysis that might be undertaken by a student of human services or a professional in the field. Thus, priority is given to the statistics and statistical concepts most likely to be useful for this person. Following are a number of principles that guide this practical approach.

(Continued)

(Continued)

1. A simple statistical test will normally do the job.

This book will focus on the simple statistics that normally are useful for the human service professional *and* are likely to be understandable to that person. There are many statistical tests that vary in mathematical sophistication. The more sophisticated ones are less prone to error in the estimation of probability. But these tests are more useful for the technical expert whose job it is to pinpoint data with great precision. Statistical precision has little meaning to the front-line professional whose goal is simply to move beyond professional intuition in evaluative endeavors to the use of concrete data analysis.

We could use the example of comparing gain scores between a treated group and a comparison group. You could use the statistic known as the analysis of covariance for this analysis, but it is more complicated both in execution and understanding than the independent *t* test. Therefore, the *t* test is included in this book, while analysis of covariance is not.

This practical approach to statistical analysis assumes that using a more sophisticated test is not likely to change the basic conclusions that are drawn from a study. The more sophisticated test will provide slightly different *p* values than the less sophisticated one, but it is unlikely that the difference in *p* values will be so great as to change the important conclusion: whether the data can be too easily explained by chance to be taken seriously as evidence supporting the treatment being evaluated.

2. Conservatism in the choice of a standard for statistical significance is more useful for the academician than the human service professional.

The experts in statistics tend to take a conservative approach to their trade. This leads them to favor one type of statistical error over another, meaning that an evaluative study is more likely to conclude that the treatment was not effective when, in fact, it was effective. It is as though these experts believe that we must be really sure before we draw any conclusions from data. It is better, according to this view, to say we did not find success when our intervention was in fact successful than it would be to do the opposite.

This is best explained by the examination of two types of error in data analysis. Type I error is the failure to see success when it is present. Type II error is the conclusion that something is successful when it is not. Type I error is known as the false positive: you think you have discovered something when, in fact, you have not. Type II error is known as the false negative: you think you did not find something when, in fact, you did (but have failed to see it).

These two types of error are relevant to the standard we use to determine whether our data support our hypothesis (or cause us to reject our null hypothesis). In the social sciences, the normal standard is $p < .05$, which means that our data could be explained by chance less than 5 times in 100. If our data can be explained by chance less than 5 times in 100, we say we have achieved statistical significance and, therefore, our data cannot be easily explained by chance. If our clients gained at a level that meets this standard, we conclude that our intervention was effective.

But what if our *p* value is .09 instead of .05 or less? If we accept the normal standard in the social sciences, we have to conclude that our data are not statistically significant; therefore, we have failed to find evidence of the success of our intervention. With a *p* value of .09, we are saying that our data would be explained by chance 9 times in 100. But maybe this is good enough.

The standard of .05 is arbitrary. There is no basic scientific basis for it. It is a standard selected by conservative statisticians who want to err on the side of being too cautious rather than err on the side of not being cautious enough. There is nothing unscientific about using the level of .10 rather than .05 as the standard. Nevertheless, in these pages you can assume that we are using the generally accepted standard, unless otherwise notified.

3. The distinction between the ordinal and interval levels of measurement is more useful for the theoretician than the practitioner.

Level of measurement is a critical theme in selecting a statistic for your data analysis. Variables that are measured at the nominal level place people into categories that have no order (e.g., smoker, nonsmoker), while variables measured at the ordinal level place people into categories that have an order (e.g., strongly agree, agree, disagree, strongly disagree). Variables measured at the interval level have a numerical value, and one measurement could be said to be a specific amount more or less than another (e.g., age measured in years; a 33-year-old is 3 years old than a 30-year-old). The statistics we select are predicated upon assumptions about the level of measurement. For example, the chi square test is for variables measured at the nominal level, while the Pearson correlation coefficient is for the examination of two variables measured at the interval level. It makes no sense to treat a nominal variable as though it is measured at a higher level because there is no numerical meaning to the categories it contains. Men are neither higher nor lower than women. Republicans are neither higher nor lower than Democrats. Human service professionals who select the field of aging as a specialization are neither higher nor lower than those who select mental health or physical health or child welfare. These are all just categories that distinguish one person from another. So, if studying the relationship between gender and one's field of specialization, you would not give each specialization and each gender an arbitrary number when you code your data for the computer and then compute a Pearson correlation between the two variables. This would make no sense because the Pearson correlation assumes that you are working with interval variables, such as age and number of years of schooling. Instead, you would need a statistic that assumes the variables are nominal.

But what about the distinction between the ordinal and interval levels of measurement? How important is this? Now we have come to the point of this segment. The distinction is more theoretical than practical. Suppose we assume that the ordinal variable is measured at the interval level and employ a statistic that makes this assumption. Are we doing something really stupid? I think not. For example, we might enter data into the

(Continued)

(Continued)

computer in regard to the variable "the extent to which clients agree they would recommend our agency to another person" as follows:

1 = Not at all
2 = To a small extent
3 = To some extent
4 = To a great extent
5 = To a very great extent

The theoretical argument is that we cannot assume that the numerical distance between "not at all" and "to a small extent" is the same as the numerical distance between "to a small extent" and "to some extent." But we violate this distinction every day when we compute the score for a scale that contains items that are ordinal in nature (e.g., the response to question 4 on a depression scale). We give values to these ordered categories in accordance with the hierarchy and we add up the score for a score for anxiety or depression or whatever. So, if we can do that, why can't we treat an ordinal variable as though it is interval for the purposes of statistical analysis? I cannot think of a good reason not to.

4. Violate the assumption of data independence with caution and transparency.

Statistics are typically based on the assumption that the data are drawn from different people rather than being repeated measures of a variable for a single person. If the data are drawn from different people, they are said to be independent. The statistics employed in this book for data taken from group research designs are appropriate for these statistical measures because they do not violate this assumption. But the same cannot be said about the statistical approaches taken in this book for single-subject designs. These designs call upon the researcher to measure the same person repeatedly over time. If you measure one client four times in a baseline period and compute the standard deviation of baseline scores as your first step in the procedure (to be explained later), you may be violating the assumption of independence.

The statistical issue, whose full explanation is beyond the scope of this book, is autocorrelation. If these scores are autocorrelated, they will pose a threat to data independence. If not, there is less cause for worry. The importance of this issue is of debate in the literature (see, for example, Huitema, 1985). The cautious researcher will simply avoid using the statistics in this book for the single-subject research design and will, therefore, be able to ignore the issue of statistical significance altogether. The argument of this book is that statistical significance is an important enough issue to be included in the analysis of single-subject research, providing that we acknowledge the limitation imposed by matched data. To avoid the issue of statistical significance would be the greater sin.

Reference

Huitema, B. E. (1985). Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment, 7*(2), 107–118.

Quiz

1. To what does *practical significance* refer?
   a. The same thing as *statistical significance*—it is just another name for it
   b. Whether the results of a statistical study are noteworthy from a practical (or clinical) viewpoint
   c. Whether the data could easily be explained by chance
   d. The relevance of the sample selection procedures to the issue of chance
2. What is true of descriptive research?
   a. It explains one variable by examining its relationship to another variable.
   b. It characterizes a single variable.
   c. It always has two or more variables in the analysis.
   d. It explores unknown theoretical territory.
3. What is true of explanatory research?
   a. It explains one variable by examining its relationship to another variable.
   b. It describes a single variable.
   c. It always examines whether an intervention is effective.
   d. It explores unknown theoretical territory.
4. What is true of evaluative research?
   a. It examines the outcomes of an intervention in regard to a research question.
   b. It describes a single variable.
   c. It explores unknown theoretical territory.
   d. It examines what causes target behaviors to be as they are.
5. Statistical software, such as Excel or SPSS, displays a data file in which of the following formats?
   a. Paragraphs of narrative about the nature of the intervention and the target behavior
   b. A chart (or table) with variables in the rows and cases in the columns
   c. A chart (or table) with cases in the rows and variables in the columns
   d. A graph showing the relationship between two variables
6. This book will help you to undertake statistical analysis of data for what kind of research?
   a. Descriptive research
   b. Evaluative research
   c. Explanatory research
   d. All of the above
7. The concept of statistical significance refers to what issue?
   a. Causation
   b. Generalization
   c. Chance
   d. The connection between analysis and attribution

8. What does a *p* value of .34 mean?
   a. The data would be expected to occur by chance 34 times in 100.
   b. The data would be expected to occur by chance 0.34 times in 100.
   c. You have achieved statistical significance according to the normal standard in the social sciences.
   d. None of the above

9. Which of the following statements about descriptive and inferential statistics is/are true?
   a. Inferential statistics are used to test an explanatory or evaluative hypothesis.
   b. An example of a descriptive statistic is a mean.
   c. Both of the above are true.
   d. Neither of the above is true.

10. When you report your findings from the test of your hypothesis, you should present data for what purpose?
    a. To show whether statistical significance has been achieved
    b. To give the reader information on which to draw conclusions about practical significance
    c. Both of the above
    d. Neither of the above

11. Which of the following would be more difficult to defend in a debate?
    a. The achievement of statistical significance in a situation where practical significance was not apparent
    b. The achievement of practical significance where statistical significance was clearly not achieved
    c. Both (a) nor (b) are equally difficult to defend.
    d. Neither (a) nor (b) could be defended at all.

12. What do you need to know in order to use a guide to find a statistic for your data?
    a. The variables you have measured
    b. What your data look like in regard to structure, such as level of measurement and whether the data are related or independent
    c. Both of the above
    d. Neither of the above

KEY TERMS

Chance. The issue addressed by statistical tests. If the data can easily be explained by chance, you cannot depend upon them.

Constant. Something that does not vary. For example, if all the people in your study sample are male, the concept of gender is a constant rather than a variable in your study.

Descriptive research. Research that has the purpose of describing a sample of people in regard to such things as frequencies, proportions, means, and standard deviations.

Descriptive statistics. Statistics that describe a variable in regard to such things as frequencies, proportions, means, and standard deviations.

Effect size. The amount of client gain for a single group, or the difference in gain between two groups, as represented by units of standard deviations. Effect size facilitates the comparison of results across different studies. An effect size of 1.0 means the gain (or difference in gain) is equal to one standard deviation of scores.

Evaluative research. Research that has the purpose of determining the outcome of an intervention.

Explanatory research. Research that has the purpose of explaining a variable by examining its relationship to other variables.

Hypothesis. A statement of what you expect your data to reveal if your theoretical position is supported. In evaluative research, your hypothesis would suggest that you will find that your intervention was effective in achieving a certain outcome (e.g., "Posttest scores for self-esteem will be higher than pretest scores for self-esteem").

Independent data. The measurement of a variable for different groups of people rather than for the same group of people at two points in time. (The latter case refers to *related data*.)

Inferential statistics. Statistics that are used to infer the characteristics of a population based on data from a sample. These statistics are commonly used to test the hypothesis, with a focus on the extent to which the data can be explained by chance.

Magnitude. The amount of difference or the strength of the relationship. It refers to quantity. Higher magnitude means a greater difference or stronger relationship between variables. It is an indicator of practical significance.

***p*.** The number of times in 100 that chance would explain the data. The p value of .05 (or less) is generally considered the standard for finding support for the hypothesis.

Population. The study population is the group of people from whom the study sample was selected. Researchers may seek to infer conclusions about the population based on data from their sample.

Practical significance. The extent to which the data are noteworthy in your professional opinion. The greater the magnitude of the results, the more reason for declaring data to be of practical significance.

Related data (matched data). The measurement of a variable at two points in time for one group of people, as with the one-group pretest–posttest design.

Sample. The study sample is the people from whom data were collected. Compare with *population*.

Statistical significance. Statistical significance means that chance has been ruled out as a plausible explanation of the data. Statistical significance is measured by the value of p, which refers to the decimal equivalent of the number of times in 100 (the percentage) that the data would occur by chance. The normal standard in the social sciences is $p < .05$, which means the data would occur by chance less than 5 times in 100.

Threshold. A number that represents a level of client status that could be used for comparison of data from a study.

Variable. Something that varies. It is the key concept in measurement because research studies examine data in regard to variables like gender, age, score on a depression scale, and so forth.

Variance. The extent to which data fluctuate from one score to another, or the level of diversity in the data. One measure of variance is the standard deviation.