

5

Analyzing Data With Pretest and Posttest Measurements of One Group

*Using the Paired t test, the One-Sample t Test,
and the Binomial Test*

Paul is evaluating a recreational therapy program for people with dementia. The objective of this program is to enhance general life satisfaction. He administered a life satisfaction scale to a group of clients once before service began and once again at the end of 8 weeks of service. Because he did not collect data in a manner to match each person's **pretest scores** and **posttest scores**, he needs to compare the posttest scores of his 21 clients to the mean of the pretest scores. The steps in his data analysis procedure are as follows:

1. He consults Exhibit C.1 (Appendix C) of this book and realizes that his situation falls into the first line on the table (Situation 1) because he is comparing a set of scores (posttest) to a single score (mean of the pretest).
2. He realizes from his consultation of Exhibit C.1 that the one-sample t test is appropriate for his data.
3. He loads into his computer the special Excel file labeled “York, one-sample t test, comparing interval variable to a single score” (as indicated in Table C.1).
4. He enters into one of the columns in this special Excel file each of the posttest scores for his 11 clients.

5. He enters the mean of the pretest scores in the cell of this special file as instructed by it.
6. He examines the data to see whether he has achieved statistical significance and whether the mean of the posttest scores is higher than the mean of the pretest.

In this chapter, you will review how to test your evaluative research hypothesis when you have one group of people in your study and you have measured them before and after receipt of your service. Three situations are included in this chapter:

1. You have pretest and posttest scores for one group of clients, and you can match each person's pretest score with the posttest score (the paired-samples t test).
2. You have pretest and posttest scores for one group of clients, and you cannot match the pretest score with the posttest score (the one-sample t test).
3. You have pretest and posttest measurements on a dichotomous (yes or no) variable for a group of clients.

The first two situations above employ some form of the t test, while the last one employs the binomial test.

Using the t Test

As noted in the list above, two forms of the t test are the paired t test and the one-sample t test. There is a third form, the independent t test, and it will be discussed in Chapter 6, which concerns group research designs. We need to distinguish between the paired and one-sample t tests and between two general types of this test, one-tailed and two-tailed. This chapter will start with an examination of one-tailed and two-tailed tests followed by a review of the paired and one-sample tests. Then you will examine how to employ each of the forms of the t test using Excel and SPSS. Finally, this chapter will examine the binomial test, which you use when you have pretest and posttest measurements with a dichotomous variable. You'll again learn how to use the Excel file to conduct this test, but guidance is not given for the use of SPSS because the complexity of using this software for the binomial test exceeds the scope of this book.

The One-Tailed and Two-Tailed t Test and the Directional Hypothesis

In Chapter 4, you saw an explanation of the normal distribution. Exhibit 5.1 displays the normal distribution of people by IQ scores. As you can see, the mean is 100, and the standard deviation is 15. Therefore, if you have an IQ of 115, you are one standard deviation higher than the mean. IQ scores of 130 or higher represent the high tail of the distribution, which is 2.5% of the total. Another 2.5% of people have an IQ

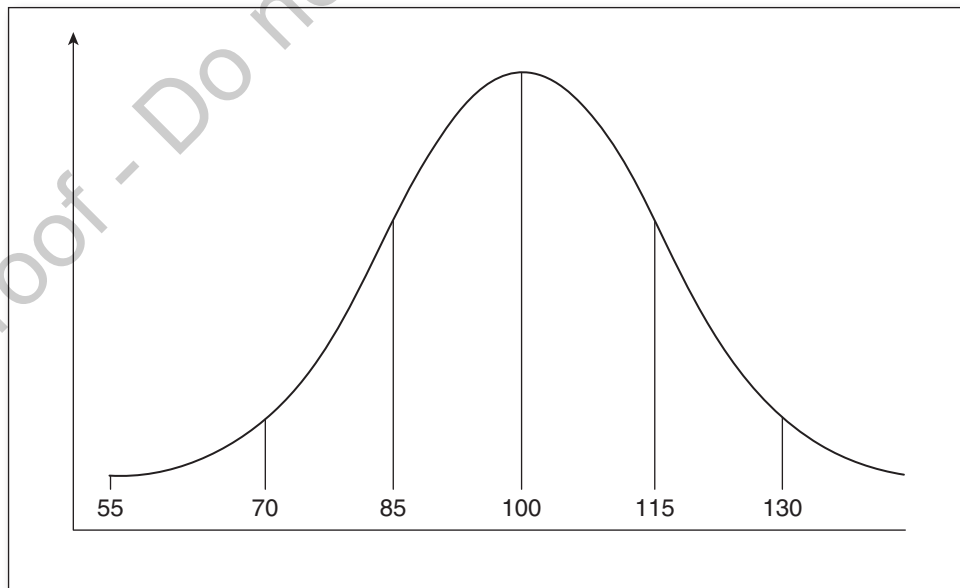
of 70 or lower, which represents the low tail of the distribution. When you combine these two tails, you have 5% of the population. This is related to the concept of “ $p < .05$.” Scores that fall in either tail are different enough from the mean to be statistically significant according to the normal standard in the social sciences.

The tails of the distribution are relevant to whether you have a directional hypothesis or a nondirectional hypothesis. Suppose you are studying the relationship between religiosity and income. If you have no basis for expecting that religious people will have higher incomes than other people, your hypothesis is “Religious people and nonreligious people have *different* levels of income.” This is a **nondirectional hypothesis** because you did not specify which group is expected to have higher incomes. In this case, you want to see whether the difference falls into *either* of the two tails of the distribution. This calls for the use of the **two-tailed t test**.

On the other hand, say you expect the results to fall into only one tail of the distribution. Now you state a **directional hypothesis**: “Religious people have *higher* incomes than nonreligious people.” When we engage in evaluative research, we always state a directional hypothesis because we have a basis for expecting the results will be different in a particular direction (i.e., that scores will show improvement). In this situation, we will use the **one-tailed t test**.

Because this book focuses mainly on evaluative research, examples will usually use the one-tailed t test. However, some people like to use the two-tailed test because it is more conservative, and you will find that SPSS has the two-tailed t test as a default (which you can change for your analysis). So, if you seek statistical significance from

Exhibit 5.1 Normal Distribution (Bell Curve) of IQ Scores



using the two-tailed t test in evaluative research, this is perfectly acceptable. You will just be using a more conservative approach than a one-tailed test, and most people will not argue with that.

However, the philosophical position taken by this author is that the conservative approach tends to support the conclusion that an effective intervention is actually not effective and this error is not necessarily a good thing in evaluative research for the human services. A less conservative approach will tend to err in the opposite way, giving us reason to conclude that an ineffective intervention is in fact effective. Whether it makes sense to us a more or less conservative approach will vary with the situation. If funds are highly limited and you are examining several treatments to determine which one will be given the limited funds, a more conservative approach makes sense. On the other hand, if evidence on treatment effectiveness is limited (and funds are not), it makes more sense to use a less conservative approach. This less conservative approach keeps more interventions in the category of approved practice, giving the practitioner, who may be in an environment that restricts practices based on evidence, more flexibility. The critical point is that the one-tailed t test is the less conservative approach.

Selecting the Appropriate Form of the t Test

Exhibit 5.2 summarizes the criteria, first presented in Chapter 3, that guide your choice of t test. You need to know the level of measurement of your dependent variable, the research design, and whether you have matching data or independent data.

In all cases when you use the t test, your dependent variable must be measured at the interval level. If you are employing the one-group pretest–posttest design with matching scores, you can employ the **paired t test**.

If you have pretest and posttest scores that cannot be matched, you can compute the mean of the pretest scores and use this as the threshold score for the comparison of the posttest scores using the **one-sample t test**. You can also use the one-sample t test if you have some other threshold score for comparison of your posttest scores. For example, you may have data suggesting that the mean pretest score on the Beck Depression Inventory for clients seeking treatment for chronic and severe depression is 32.4 and have posttest scores (but not pretest scores) for your group of 15 clients who are being treated for chronic and severe depression. Maybe you could compare your posttest scores for these clients to that threshold score of 32.4.

If you are comparing gain scores of two groups, you can use the **independent t test**, discussed in the next chapter.

Examining Statistical Significance and Practical Significance With the t Test

When you report your findings, you should provide information that helps the reader to evaluate the issues of practical significance and statistical significance. Statistical significance refers to the extent that your data can be explained by chance. Practical significance refers to the magnitude of the results. Was the client gain noteworthy? Was

Exhibit 5.2 Choosing Among the Three Forms of the *t* Test

Level of measurement of the dependent variable	Research Design		
	One-group pretest–posttest		Comparison group (comparing gain scores of two groups)
	Comparing matching scores	Comparing posttest scores to a threshold score	
Interval	<i>Paired t test</i>	<i>One-sample t test</i>	<i>Independent t test</i>

the difference between the gain of the treatment group and the gain of the comparison group noteworthy? You can have statistical significance without practical significance because you might find the statistically significant amount of gain to be unimpressive.

Statistical Significance

The *p* value is the measure for statistical significance. It reveals the fractional equivalent of the number of times in 100 that your results would occur by chance. A *p* value of .23 means your data would occur by chance 23 times in 100; this result would be deemed statistically insignificant (i.e., $p < .05$) for finding support for your hypothesis. One way to think about statistical significance is in terms of the normal distribution and standard deviation. If your data fall outside of two standard deviations from the mean, you have statistical significance at the 5% level ($p < .05$). Note that in your Excel file, the value of *p* is given as “p,” but SPSS will report this value in the column labeled “Sig (2-tailed)” rather than just labeling it “p.”

As mentioned before, the $p < .05$ standard is arbitrary, with no scientific basis. Thus, one could plausibly argue for a more lenient standard, such as $p < .10$. If you accepted this standard, chance would explain your results 10% of the time. The more lenient your standard, the more likely you will avoid the error of rejecting an effective treatment, but you will increase the likelihood of accepting a treatment that is not effective. So, take your poison. Traditional statisticians are conservative; they try to avoid the second type of error and discount the importance of the first type of error. But the decision is up to you—unless you wish to publish your results. In this case, the suitability of your study for publication will be reviewed by those conservative scholars, and they will look more favorably on the use of a $p < .05$ standard.

Practical Significance

The reason for examining practical significance is to help with professional decision making regarding an intervention. Is the difference between the pretest and posttest

measurements great enough to justify the resources used to implement the treatment? Alternatively, is the difference between the treatment group and the comparison group noteworthy? In other words, did the treatment make enough of a difference to justify its use? We review two important vehicles for examining this question: (1) the extent to which clients achieved a gain that moved them from one threshold of functioning to another and (2) effect size.

Thresholds of Functioning

One guide for this conclusion is whether the difference in scores moved from one threshold of functioning to another. Some scales of depression, for example, have thresholds of functioning that determine the extent of the need for treatment. One level may be declared to be minimal depression (e.g., scores from 0 to 15), another may be moderate depression (e.g., 16–25), and another may be severe depression (e.g., 26+). If your clients' mean score moved from severe depression to moderate depression, you have a basis for suggesting practical significance.

Effect Size

If you are comparing pretest and posttest scores, the **effect size** is the amount of gain measured in standard deviations. You will normally report the amount of difference between the pretest mean and the posttest mean. For instance, the pretest mean for self-esteem might be 15.6 and the posttest mean might be 25.1, for a difference of 9.5. So, your clients typically achieved a nearly 10-point gain. But this figure cannot be easily compared to the outcome of another study that used a different tool for measuring self-esteem. Perhaps your scale has a range of 0 to 30 while the other study used a scale with a range of 0 to 50. A posttest mean of 25.1 would represent a very different level of self-esteem for people taking your scale than for people taking the other scale. And with a much greater range of possible scores for the other scale, a gain of 9.5 points would be a less impressive result in that study than in yours.

So, what do we do? We compute the standard deviation of gain by dividing the gain by the standard deviation of pretest scores. So, if your clients had a standard deviation of gain of 1 and clients in the other study had a standard deviation of gain of 0.5, your clients can be said to have gained more. This is despite the fact that the raw scores of the other group would indicate that they had gained more. In human service research, reporting of effect sizes has received greater emphasis in recent years, for the reason just noted.

The effect size is computed somewhat differently for different forms of the *t* test. When you use the paired *t* test to examine your hypothesis, you compute the effect size by taking the difference between the mean pretest score and the mean posttest score and dividing this figure by the standard deviation of the pretest scores. If you are using the one-sample *t* test, you divide the difference between the threshold score and the posttest mean by the standard deviation of the posttest scores. If you are using the independent *t* test for the comparison group design, you divide the difference in gain

between the two groups by the standard deviation of the comparison group. These will be discussed in more depth later when each form of the t test is illustrated. However, you should know that there are various ways of computing the effect size. We have used one that is simple and easy to understand.

How should the effect size be interpreted in regard to practical significance? First, you can report the effect size knowing that it represents the number of standard deviations of difference (or fractions of a standard deviation) attributed to the intervention (assuming you have also dealt with all the bases for determining causation). As already noted, you can compare your effect size to other known effect sizes demonstrated in other studies or in other data you know. Another way to interpret the effect size is to consider the percentile your effect size represents in regard to people not treated. Jacob Cohen, in his 1988 book *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.; Lawrence Erlbaum) presented information on this idea. The percentile is a figure that shows how one score compares to other scores. If your IQ score is at the 50th percentile for all IQ scores, your score is at the median for all who took the test. One half of all people fall below you and one half are above you. If you are at the 99th percentile, only 1% of all people have an IQ higher than your IQ. You can convert the effect size to a percentile using the data from Cohen. An effect size of 0 would mean there was either no gain for your one group or no difference in gain between your treated group and the comparison group. An effect size of .2 would put your gain at the 58th percentile. Here are more of the effect size figures from Cohen.

Effect size	Percentile
.2	58
.4	66
.6	73
.8	79
1.0	84

SOURCE: Adapted from Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

From these figures, you can see that if you have an effect size of 1.0, you can say that the average person in the treated group had a score that was equal to the 84th percentile of the comparison group (or better off than 83% of the people in that group). If you have an effect size of .4, the average person in the treated group had a score that equaled the 66th percentile of the comparison group.

According to Cohen, an effect size of .2 is small, an effect size of .5 is medium, and an effect size of .8 is large. These figures, however, are a matter of opinion. Cohen was

a mathematician and invented this way of calculating effect size, and we often rely upon the opinions of such people for guidance.

Testing Your Hypothesis With the Paired-Samples t Test When You Have Matching Pretest and Posttest Scores

Perhaps the type of analysis that is most often used for evaluative research by the front-line professional or the student of human services is analysis of matching pretest and posttest scores for one group of clients. While it is often feasible to ask clients to complete a pretest and posttest instrument, it is not often feasible to do the same for a comparison group. Thus, matching pretest and posttest scores are often the data people have available to study.

With matching pretest and posttest scores from the same group of people, you can employ the paired t test as you learned about in Chapter 4. Refer to that chapter if you would like a refresher.

Using Excel for the Paired-Samples t Test

In Chapter 2, you saw a hypothesis tested with the paired t test using the special Excel file designed for this book. The data were shown in Exhibit 2.2, and the Excel results were displayed in Exhibit 2.3. You'll be doing this analysis here.

Your first step in using the Excel file is to load the Excel software on your computer, if it is not already there. Then load and open the special Excel file entitled *York, paired t test, comparing two sets of matched scores*. The screen you'll see is shown in Exhibit 2.3, without the calculated values. Next, enter the pretest and posttest scores from Exhibit 2.2. Then, enter the gain or loss in the third column titled "Gain." Assuming a higher score indicates greater self-esteem, as with these data, subtract the pretest score from the posttest score to calculate the gain or loss. For example, if a client went from a 14 pretest score to a 23 posttest score, then $23 - 14$ equals a 9-point gain. (If your scale were such that a lower score meant greater self-esteem, then you would subtract the posttest score from the pretest score to compute the gain. This is because a drop in score indicates an improvement for the client.) If a case experienced a loss, enter this figure with a negative sign in front—it is "negative gain."

Finally, review the results to see whether your calculated values are the same as those in Exhibit 2.3: The mean posttest score should be 15.55556, the mean pretest score 11.77778, the mean gain score 3.77778, the value of t (rounded) 3.034, and the value of p .01. If your results are different, check your data entry and give it another try.

These results suggest that the data support the hypothesis (posttest scores were higher than pretest scores) because the posttest scores are higher, as hypothesized, and the difference between pretest and posttest scores is statistically significant. The effect size for these data is displayed as -1.30384 . You can ignore the negative sign for the effect size (and you can also ignore it for the value of t when you get such a value in the future).

When reporting all these figures, you would round to the nearest hundredth so that, for example, the mean posttest score of 15.55556 would be reported as 15.56.

Using SPSS for the Paired-Samples t Test

You can use SPSS to do a paired t test on the same pretest and posttest scores from Exhibit 2.2, also used in the Excel example above. Here's how to use SPSS for this task.

1. Load the SPSS software onto your computer.
2. Construct the structure of the SPSS file under the "Variable View" tab as instructed in Chapter 2 and shown in Exhibit 2.1. In other words, you will go to the "Variable View" tab and insert the name *pretest* for the pretest score variable and the name *posttest* for the posttest score variable. (You will not enter *gain* as a variable.)
3. In the "Data View" window, enter the data from Exhibit 2.2.
4. Across the top of the screen, you see various menu items. From this list, select **Analyze**. This will give you a drop-down list with several options.
5. Select **Paired Samples t Test** from the drop-down list.
6. You will see the list of all variables in the box on the left of the screen, an arrow in the middle, and a box on the right with the headings of "Pair 1" and "Pair 2." Click on the variable *pretest* from the box on the left and move it to the box on the right by hitting the arrow in the middle. This variable should appear in the column labeled "Pair 1." Move the variable *posttest* over to the box on the right in the same manner, and you will see this variable listed in the column "Pair 2."
7. Go to the bottom of the screen and click **OK**. You will see the screen in Exhibit 5.3.

In Exhibit 5.3, the first box shows the mean scores for pretest and posttest, the number of people in the sample, the standard deviation, and the standard error of the mean. The main thing to note here is the mean pretest score (11.7778) and the mean posttest score (15.5556).

The middle box shows the paired-samples correlations. This analysis is not a correlation analysis but is an examination of two sets of matched scores, so this information (middle box) is not useful for our basic question and we will ignore it. The box at the bottom reveals the important information we are seeking, the value of t and the value of p . The value of t is given under the heading "t," and the value of p is given under the heading of "Sig (2-tailed)." The value of t in our case is -2.861 , and the value of p is $.021$. The value of t is a negative value, but in your report of this analysis, you would ignore the negative sign, reporting the statistic simply as 2.86 .

Exhibit 5.3 SPSS Output for a Paired t Test

→ **T-Test**

[DataSet2]

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 pretest	11.7778	9	3.07318	1.02439
posttest	15.5556	9	3.74537	1.24846

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 pretest & posttest	9	.338	.374

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	pretest - posttest	-3.77778	3.96162	1.32054	-6.82295	-.73261	-2.861	8	.021

You may have noticed that these values are slightly different from those we got when we used Excel. This is partly because the SPSS file used the two-tailed t test while the Excel file used the one-tailed t test.

Reporting the Findings of Your Analysis

You will report the mean pretest score, the mean posttest score, the effect size, the number of people in the analysis, and the value of p . The special Excel file gives you all these values. If you use SPSS, on the other hand, you'll need to do more work.

With SPSS, you will need to calculate the effect size by dividing the mean difference between the pretest and posttest scores by the standard deviation of the pretest scores. The SPSS output file shows a mean difference of -3.77778 (in the table labeled "Paired Samples Test") and a standard deviation of pretest scores of 3.07 (in the first table, labeled "Paired Samples Statistics"). If you divide 3.77778 by 3.07 , you get the figure 1.23 (again ignoring the negative sign), which is your effect size. (By the way, you do not need to use as many decimal places as we used here for the mean. I only included all the digits so you could make sure you were looking in the right place in the SPSS output file.) You will notice that the effect size is slightly different between Excel (1.30) and SPSS (1.23). With Excel, you have an effect size of 1.30 , while the effect size with SPSS is 1.23 . You will see minor differences like these depending on the software you use to conduct your analysis, but the numbers are similar enough that your conclusions will be the same regardless of which figure you use.

Here is one way you could report the data from Exhibit 5.3.

The mean pretest score for self-esteem is 11.77, while the mean posttest score is 15.55. These data were subjected to the t test for paired samples, with the results showing a statistically significant gain ($t = 2.86$; $n = 9$; $p = .021$). The effect size is 1.23, which means that the posttest scores are slightly more than a standard deviation better than the pretest scores. This is considered a high effect size.

Testing Your Hypothesis With the One-Sample t Test When You Have Pretest and Posttest Scores That Cannot Be Matched

In some situations, you are not able to match clients' pretest scores with posttest scores. For example, you may want to keep each client's score anonymous. In this situation, you would collect pretest and posttest data from all clients in your study without any identifying information. Thus, you cannot enter a pretest and posttest score for a given case on, say, row 3 because you don't know which scores belong to whom. After computing the mean pretest score, enter all the posttest scores into the computer along with this mean pretest score. The statistical question is whether the posttest scores differ significantly from the pretest mean score.

Using Excel for the One-Sample t Test

The example illustrated in Exhibit 5.4 displays posttest scores for self-esteem for eight at-risk middle school students and a threshold score of 10 that represents the pretest mean. You could download the file (*York, one sample t test, comparing interval variable with a single score*), enter these data, and see if you get the same result. First, enter each of the 8 scores in column A. Then enter the number 10 in the cell labeled "enter threshold score here." Review the calculated mean score, effect size, and value of t . You can examine the table below your results, titled "Determining p values," to determine whether your t value is statistically significant (i.e., $p < .05$).

Exhibit 5.4 shows a mean posttest score of 14.875 (next to the heading "Mean score" in the third column). The value of t is 3.808. When you examine the table of t values below, you will see examples for a sample size of 5 and 10 and a few other sample sizes, but there is no information specifically for a sample of 8. So, to be conservative, check whether your t value would be statistically significant if your sample were smaller than it is. In the table, you see that a sample of 5 with a t value of 3.47 is significant at the $p < .05$ level. You have a t value greater than 3.47 and a sample size greater than 5, so you can say that your data are statistically significant ($p < .05$). You can also find a table of t values on the Internet to get more specific data about your p value, but we will not be concerned here about specific values of p . Instead, we simply care whether the standard of .05 has been met. It has, so we will conclude that our data support our hypothesis.

Exhibit 5.4 Using Excel to Compute the One-Sample t Test

	A	B	C	D	E	F	G	H	I	J	
1	York, one sample t test, comparing interval variable with a single score										
2											
3	Use this file if you have a set of scores (e.g., posttest) and a threshold score (e.g., mean of pretest)										
4	for comparison. Enter the threshold score in C16 and your scores in A6 through A46										
5	This file cannot be used for more than 40 scores. See the value of t in cell C18.										
6	12	<== Start Here									
7	18										
8	9		8 count								
9	17		14.875 <-- Mean score								
10	14		3.620687 <-- Standard deviation								
11	12		2.828427 e20								
12	21		1.280106 e21								
13	16										
14			4.875 e22								
15											
16			10 <--enter threshold score here								
17											
18			3.808278 <--This is the value of t								
19			1.34643 <--This is the effect size								
20			See below for determining the p value. If you have at least 5 persons								
21			and a t value at least 2.13, you have a p value of .05								
22											
23			Determining p values								
24											
25			n	t	p						
26			5	1.53	0.1						
27			5	2.13	0.05						
28			5	3.47	0.01						
29			10	1.38	0.1						
30			10	1.83	0.05						
31			10	2.82	0.01						
32			15	1.34	0.1						
33			15	1.76	0.05						
34			15	2.62	0.01						
35			30	1.31	0.1						
36			30	1.69	0.05						
37			30	2.46	0.01						
38			40	1.3	0.1						
39			40	1.64	0.05						
40			40	2.42	0.01						
41											

When you consider practical significance, you can examine the difference between the mean posttest score and the threshold score to which it is being compared and determine whether you believe this level of difference is noteworthy. You can also examine the effect size. For the data in Exhibit 5.4, you can see an effect size of 1.34643. This is considered a large effect size when client gain is being measured, but, of course, this is a matter of opinion—just like your judgment about whether the difference between 14.875 (the mean posttest score) and 10 (the threshold score) is clinically noteworthy.

Using SPSS for the One-Sample t Test

We can use the same data we used for the paired-samples t test to do a one-sample t test in SPSS. Again, Exhibit 2.2 gives matched pretest and posttest scores for one group. In this example, we are simply going to pretend we are not able to match the pretest and posttest scores. Instead, the pretest scores are just the random presentation of pretest scores for these clients. We will use the mean of these pretest scores as the threshold score to which we will compare the posttest scores.

Your first step is to compose the structure of your data file using the “Variable View” tab (shown at the bottom of the screen). You have only one variable—posttest scores. Choose a label for it, such as *score* or *posttest*, and enter this name. You do not need to do anything else to set up the structure of your file. Next, go to the “Data View” tab, where you will see the name you have given your variable in the first column. Enter the 9 scores as follows: 18, 11, 16, 10, 18, 21, 16, 12, 18.

Now you are ready to analyze your data using the one-sample t test. Here are the steps:

1. Go to the top of the screen (in “Data View”) and click on **Analyze**. You will get a drop-down where you will . . .
 - Click on **compare means**. You will get another drop-down menu where you will . . .
 - Click on **One sample t Test**.
2. You will see a box on the left with your variable name in it. Click on this variable to highlight it and then click on the arrow in the middle of the screen to move your variable to the “Test Variable” box on the right of the screen.
 - Go to the box that is labeled “Test Value” and change the number from 0 to 11.77 (the mean pretest score we are using for comparison).
 - Click **OK** at the bottom. You will get your results in the output screen, shown in Exhibit 5.5.

In Exhibit 5.5, you see the mean posttest score (15.556), the value of t (3.032), and the value of p under the column “Sig (2-tailed).” In this case, the value of p is .016. The figures computed in SPSS are not identical to those in Excel, most likely because of differences in rounding of numbers and so forth. These differences do not change the conclusion regarding whether the data support the hypothesis—the critical question from the viewpoint of a practical approach to statistical analysis.

Compute the effect size by dividing the difference between the mean posttest score and the mean pretest score by the standard deviation of the posttest scores. Here, the difference between the mean pretest and mean posttest scores is 3.78, and the standard deviation of the posttest scores is 3.745. When we divide the former by the latter, we get 1.009. This effect size is a little different if you use the standard deviation of the pretest scores rather than that of the posttest scores, and, according

Exhibit 5.5 SPSS Output for a One-Sample t Test

→ **T-Test**

[DataSet0]

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
posttest	9	15.5556	3.74537	1.24846

One-Sample Test

	Test Value = 11.77				
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference
	Lower	Upper			
posttest	3.032	8	.016	3.78556	.9066 6.6645

to Cohen (1988), either of these is appropriate. Indeed, in both cases, the effect size is close to 1.0, which is considered to be a large effect size. We will not, in this book, quibble over fractions.

Reporting the Findings of Your Analysis

You will report the sample size, the mean posttest score, the value of t , and the value of p . Here is one way the data in Exhibit 5.5 could be reported:

The mean posttest score for these 9 clients is 15.56, which was compared to the mean pretest score of 11.78 with the one-sample t test. The results reveal support for the hypothesis ($t = 3.03$; $p < .05$). The effect size is 1.0, which is considered to be a large effect size.

Testing Your Hypothesis With the Binomial Test When You Have Pretest and Posttest Measurements of a Dichotomous Variable

Sometimes you will measure client success in a dichotomous way, meaning that it is recorded as either *yes* or *no*. Suppose you are working with a group of middle school students to help them do a better job of completing their 3-week homework projects. You have data on the 3 weeks prior to the beginning of your special

service and the 3 weeks the students received this service. These case-level data are presented in Exhibit 5.6, which shows 15 students' success in turning in the 3-week homework projects both before the service (pretest) and during the service (posttest). The first student turned in the projects in both the pretest and posttest periods, while the second student failed to turn in the project in the pretest but turned it in during the posttest. The latter demonstrates success. The task is to subject these data to statistical analysis to see whether the posttest figures are significantly better than the pretest ones.

The **binomial test** can be applied to this analysis. With this test, you compare an array of data to a threshold proportion. For the data in Exhibit 5.6, the threshold proportion is the proportion of successful students in the pretest period.

Exhibit 5.6 Successful Completion of 3-Week Homework Projects

Student ID Number	Did this student complete the 3-week homework project?	
	Pretest	Posttest
1	Yes	Yes
2	No	Yes
3	No	Yes
4	Yes	No
5	No	Yes
6	No	No
7	No	Yes
8	No	Yes
9	No	No
10	Yes	Yes
11	Yes	No
12	No	Yes
13	No	Yes
14	No	No
15	No	Yes

These data show that 4 of the 15 students were successful in the pretest. This is a proportion of 0.27 ($4/15 = 0.266$, which rounds to 0.27). In other words, just over 1 student in 4 was successful before your special service began. The task of your service is to raise this proportion. The task of statistical analysis is to determine whether the improvement in the proportion is statistically significant (i.e., cannot easily be explained by chance).

When you employ the Excel file for the binomial test, you need three pieces of data: the proportion that serves as the threshold for comparison (entered as a decimal), the number of people who were successful during the posttest, and the total number of people in the posttest data. In Exhibit 5.6, these figures are as follows:

Threshold decimal proportion = 0.26

Number of successful posttest students = 10

Total number of students with posttest data = 15

Using Excel for the Binomial Test

In Exhibit 5.7, you can see the data from Exhibit 5.6 entered into the special Excel file. The p value associated with these data is displayed as 0.001282. This value could be reported as " $p < .01$." Thus, this group improved at a level that cannot easily be explained by chance (i.e., it is statistically significant).

Exhibit 5.7 Using Excel for the Binomial Test With Dichotomous Pretest–Posttest Data

A	B	C	D	E	F	G	H	I	J	K
York, binomial test, pretest-posttest, dichotomous data										
Use this file when you have pretest and posttest data for one group of people and the dependent variable is measured as a dichotomy (e.g., Yes or No)										
You should determine the proportion of favorable recordings for the pretest data and enter this fraction in A13										
You should enter the number of favorable recordings in the posttest in A11.										
You should enter the total number of posttest recordings in A12.										
You will see the value of p in A14.										
	10	<==enter the number of favorable posttest recordings in A11								
	15	<== enter the total number of posttest recordings in A12.								
	0.27	<== enter the proportion of favorable pretest recordings in A13.								
	0.001282	<== This is your p value								

Using the Binomial Test for the Posttest-Only Design When You Have a Threshold Proportion for Comparison

This chapter is about the analysis of data for the one-group pretest–posttest design. You just saw how to use the binomial test when you have a dichotomous dependent variable. But what if you do not have pretest data? You are not lost as long as you have a threshold proportion to which to compare your **posttest-only design** data. Suppose, for example, that your homeless shelter’s special service has data showing that 20 homeless people who entered your shelter 60 days ago have found permanent homes. You examine your data and find that a total of 30 homeless people were admitted to your shelter 60 days ago, so you have a success rate of 20 out of 30 people or two thirds. Is this good news or bad?

This is not an easy question to answer unless you have a basis for comparing your two-thirds success rate. Suppose that you have found data suggesting that for the nation as a whole, the normal rate of success for 60 days of service is only one third. Is your figure of 20 out of 30 significantly different from a proportion of one third? To answer this question, you can use the binomial test. Using the Excel file shown in Exhibit 5.5, you can enter 20 as the number of favorable posttest recordings, 30 as the total number of posttest recordings, and 0.33 as the threshold for comparison.

Insight Box 5.1

Are Your Pretest and Posttest Data Amenable to Statistical Analysis?

As with other applications of statistical tests, you should examine whether your data are amenable to statistical analysis in a way that would constitute a fair test of your intervention. If your sample size is so small that only a large magnitude of gain will show up as statistically significant, perhaps the wise course would be to declare your data not amenable to statistical analysis. Such a conclusion would leave you with solely the descriptive analysis of your data and eliminate inferential analysis as an option. For example, you could report the pretest and posttest mean scores but not subject the data to a statistical test to determine whether the difference is statistically significant.

When you have pretest and posttest measurements, you can examine the question of whether your data are amenable to statistical analysis by considering magnitude of difference, sample size, and variance. If you consider the issue of magnitude of difference, you would review your pretest data to see whether there is sufficient room for growth that achieving statistical significance is reasonably likely. If there is only a small

(Continued)

(Continued)

difference between pretest and posttest scores, you are not likely to find your data to be statistically significant. In addition, you are unlikely to declare your data to be of practical significance.

Let's examine the magnitude issue with an example. If you give a group of adolescents a self-esteem scale with a possible range of scores of 0 (lowest) to 25 (highest) and find the pretest mean score to be 21.3, you can see that these adolescents have very little room for growth as indicated by this scale. If you engaged in the statistical analysis of the data, you would probably find your treatment was a "failure." You are either treating the wrong behavior, if the adolescents do indeed already have high self-esteem, or you are using the wrong tool for measuring self-esteem. The logical thing to do would be to declare the data not amenable to statistical analysis. You would be well advised to select either a different behavior to measure or a different tool for measuring self-esteem.

Sample size is also a determinant of statistical significance. Samples of fewer than 15 cases are generally considered small for the use of the t test, but statistical significance is found in many studies with this sample size, so a sample size of less than 15 should not be considered a disqualifier for statistical analysis. Instead, approach statistical analysis with caution. Suppose that you believe your mean posttest score is sufficiently better than your mean pretest score to be of practical significance. Suppose further that your data are not statistically significant. What should you do? You are not advised to declare practical significance without statistical significance, because failure to find statistical significance means your data can too easily be explained by chance for you to rely upon them. Instead, you are advised to report your mean pretest and posttest scores, the results of your statistical analysis of data, and the fact that your data fail to support the hypothesis. However, you could add to the discussion that you believe the difference in pretest and posttest mean scores is noteworthy and that the failure of your data to support the hypothesis is likely due to the small sample size. You might suggest that your study be replicated with a larger sample.

Summary

This chapter has demonstrated the application of several statistical tests to data taken from the one-group pretest–posttest design. The t test for paired data and the one-sample t test were illustrated for data measured at the interval level. The binomial test was illustrated with data that are nominal and dichotomous. There are, of course, other situations not covered in this chapter, but it is believed that these applications will cover the majority of situations in which the student of the human services or human service professional will need to examine data for the one-group pretest–posttest design.

Quiz

1. Which of the following statements is/are true?
 - a. You employ the one-tailed t test when you have a directional hypothesis.
 - b. You use a form of the t test for data measured at the nominal level.
 - c. Both of the above are true.
 - d. Neither of the above is true.
2. You can use one of the forms of the t test if you have what kind of data?
 - a. Matched pretest and posttest scores for a single group of people
 - b. Gain scores for two groups of people
 - c. Both of the above
 - d. Neither of the above
3. Statistical significance is determined by:
 - a. The magnitude of the difference in pretest and posttest data
 - b. The size of the sample
 - c. Both of the above
 - d. Neither of the above
4. What does *practical significance* mean?
 - a. A statistically significant gain is considered noteworthy from a practical viewpoint.
 - b. Your data are statistically significant.
 - c. Your data are not amenable to statistical analysis.
 - d. All of the above.
5. What does *effect size* refer to?
 - a. Whether the data have statistical significance
 - b. Whether the data went in the hypothesized direction
 - c. The amount of gain (or difference between groups) as measured in standard deviations
 - d. None of the above
6. What is effect size relevant to?
 - a. Standard deviation
 - b. Practical significance
 - c. Both of the above
 - d. Neither of the above

Practice Exercise

In this exercise, you will be given some research examples, and it will be your job to test the hypothesis with the data given. First, you will need to determine the statistic that would be appropriate (you can consult Chapter 4). Then you'll load the appropriate Excel file and enter the data. Finally, you will interpret the results and answer some questions.

Case: Improving Parental Attitudes

The first task is to subject a set of data to a test of the hypothesis. This study involved improvement of parental attitudes as measured with pretest and posttest scores on the Index of Parental Attitudes (IPA). You've seen these data before—in Chapter 3, you were tested on your ability to select the appropriate statistic. Now, you are to test the hypothesis. The table with the data is reproduced here for your convenience.

The following are the IPA scores for the 8 clients of this training program, both at the beginning (pretest) and the end (posttest) of the training. Higher scores represent more problems, so the intent is to reduce the scores.

Client ID number	Pretest Score	Posttest Score
1	72	69
2	66	57
3	46	52
4	66	47
5	77	61
6	72	65
7	42	31
8	54	56

Review Questions

1. What is the hypothesis being tested in the case?
 - a. Parental attitudes will improve because of the program.
 - b. Posttest scores on the IPA will be higher than pretest scores.
 - c. Pretest scores on the IPA will be higher than posttest scores.
 - d. Most clients will achieve a gain on parental attitudes.

2. What statistic is appropriate for testing this hypothesis?
 - a. The one-sample t test
 - b. The paired-samples t test
 - c. The independent-samples t test
 - d. The McNemar test for the significance of changes
3. What does the statistic selected in question 2 do?
 - a. It compares the proportion of pretest scores that are over a certain threshold to the number that are not over this threshold in order to determine whether this difference would likely occur by chance.
 - b. It compares matched scores of a single group to see whether there is a difference not easily explained by chance.
 - c. It compares the mean posttest score to a threshold score that represents the pretest.
 - d. It compares the mean scores of two groups to see whether they are different at a level that cannot easily be explained by chance.
4. What were the mean pretest and posttest scores?
 - a. The mean pretest score was 54.75, and the mean posttest score was 61.875.
 - b. The mean pretest score was 61.875, and the mean posttest score was 54.75.
 - c. The mean pretest score was 6.875, and the mean posttest score was 0.587.
 - d. The mean pretest score was 2.46, and the mean posttest score was 6.87
5. What were the value of t , the value of p , and the effect size?
 - a. The value of t was 2.52, the value of p was .025, and the effect size was 0.587.
 - b. The value of t was 6.875, the value of p was .025, and the effect size was 0.77.
 - c. The value of t was 0.025, the value of p was 2.46, and the effect size was 3.22.
 - d. The value of t was 0.58, the value of p was .025, and the effect size was 0.587.
6. Do the data support the hypothesis?
 - a. Yes, because the mean pretest score was higher than the mean posttest score.
 - b. Yes, because the mean posttest score was higher than the mean pretest score.
 - c. Yes, because the mean posttest score was lower than the mean pretest score and the p value revealed statistical significance.
 - d. No, because the data failed to go in the hypothesized direction.
7. How should the results be presented?
 - a. The mean posttest score (54.75) is lower than the mean pretest score (61.875), representing an improvement in parental attitudes. These data were subjected to statistical analysis, which revealed that the difference is statistically significant ($t = 2.52$; $p = .025$). Thus, the hypothesis is supported.
 - b. The p value of .025 is significant because $t = 2.46$, which means the hypothesis is supported.

- c. The hypothesis is supported, showing that the intervention was successful.
 - d. The mean gain of 6.875 is lower than the mean pretest score (61.875), suggesting that the data fail to support the hypothesis.
8. The effect size for this study could be best characterized as:
- a. An effect size that is not statistically significant but is of practical significance
 - b. An effect size that is very small
 - c. An effect size that is medium
 - d. An effect size that is very large

KEY TERMS

Binomial test. A statistical test used to compare data measured dichotomously (e.g., yes or no) to a threshold proportion. For example, you have 7 clients who were successful in regard to marital reconciliation and 3 who were not, and you want to compare this proportion of success (3/10) to a normal rate of success of 30% for couples in similar situations.

Directional hypothesis. A hypothesis that predicts the direction of the evaluative data, for example, “Posttest scores for self-esteem will be higher than pretest scores.”

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.

Independent *t* test. A form of the *t* test used to compare the scores of two groups.

Nondirectional hypothesis. A hypothesis that does not predict the direction of the data, for example, “There is a relationship between extroversion and SAT scores.”

One-sample *t* test. A form of the *t* test used to compare an array of scores to a single threshold score.

One-tailed *t* test. A general type of the *t* test (e.g., independent *t* test, paired *t* test, one-sample *t* test) used with interval data when you have a directional hypothesis.

Paired *t* test. A form of the *t* test used with matched scores, normally from a study using a one-group pretest–posttest design.

Posttest-only design. A research design where data are collected only at the end of the treatment period and is compared to a threshold.

Posttest scores. Scores that are taken from clients at the end of the treatment period.

Pretest scores. Scores that are taken from clients before treatment begins.

Two-tailed *t* test. A statistical test used with interval data and a nondirectional hypothesis.