CHAPTER W3

APPLYING STATISTICAL METHODS IN YOUR OWN RESEARCH PROJECT

Chapter Outline

- Designing Your Study: Determining the Appropriate Statistical Test
- Figuring Power and Needed Sample Size
- Conducting the Study
- Entering Scores into the Computer
- Data Screening
- Carrying Out the Major Analyses
- Writing Up Your Results

Are You Ready? What You Need to Have Mastered Before Starting This Chapter

As much of *Statistics for Psychology* as possible; Mini-Web Chapter W1; SPSS procedures from the *Study Guide and Computer Workbook* (or familiarity with another statistics program or even with just the statistical functions in a spreadsheet program)

So you're going to carry out your own research project! This chapter helps you use what you have learned in *Statistics for Psychology* to explore an idea on your own. You have a fairly clear understanding of the major statistical methods used in psychology and can make sense of them when they are used in a research article. However, you'll need some additional points to apply this to your own research. In particular, in this Mini-Web chapter we consider designing your study (determining whether there is an appropriate statistical method and determining needed sample size), conducting the study, entering the scores into the computer, checking for and dealing with missing scores (for example, where a participant did not answer all the questions), checking whether the scores on each variable meet the assumptions for the procedure you want to use (and what to do about it if they don't), carrying out the analyses, and writing up your results.

DESIGNING YOUR STUDY: DETERMINING THE APPROPRIATE STATISTICAL TEST

Once you have a possible idea for a research question, the next step is to develop a specific research plan to address that question—one that does as good a job as possible of approximating the ideal research design in terms of equivalence of experimental and control groups, equivalence of circumstances, generalizability, and measurement (see Mini-Web Chapter W1).

Before starting out on a new study, experienced researchers think out what statistical method they will use when the study is done. Otherwise you may find there are no methods available or the ones that are will be less than ideal for some reason. You also need to know what statistical tests you will use in order to figure out sample size and power (see Chapter 8).

WHAT TEST TO USE IN THE USUAL SITUATION OF EQUAL-INTERVAL MEASUREMENT

In the most common research situation, your scores are approximately equal interval (see Chapter 1 for a discussion of "levels of measurement".). Figure W3-1 diagrams a decision tree for deciding on an appropriate statistical technique in this standard situation. (Figure W3-1 also assumes you have just one dependent or criterion variable; we'll consider what to do when you have more than one at the end of this overall section on selecting an appropriate test.)

The first decision is whether you are testing differences between means on variables or associations among variables. A study comparing the effect of different colors of printing on reading time focuses on the mean reading time for each color; thus, this is a study comparing means of variables. A study looking at the relation of mother's age to her oldest child's school grades is about associations among variables.

If your study focuses on differences between means, the next question is whether there are two means being compared (for example, if you were looking at red printing vs. black printing) or whether there are more than two (for example, if you were looking at differences among printing in red, black, green, and blue). Within each of these, you next must decide whether your design is between subjects (for example, some participants will read a story in black printing and other participants will read the story in red printing) or within-subjects (for example, each participant will read one story in black printing and another story in red printing).

If your study focuses on associations, the next question is whether there is one predictor (for example, mother's age) or more than one predictor (for example, mother's age and father's age).

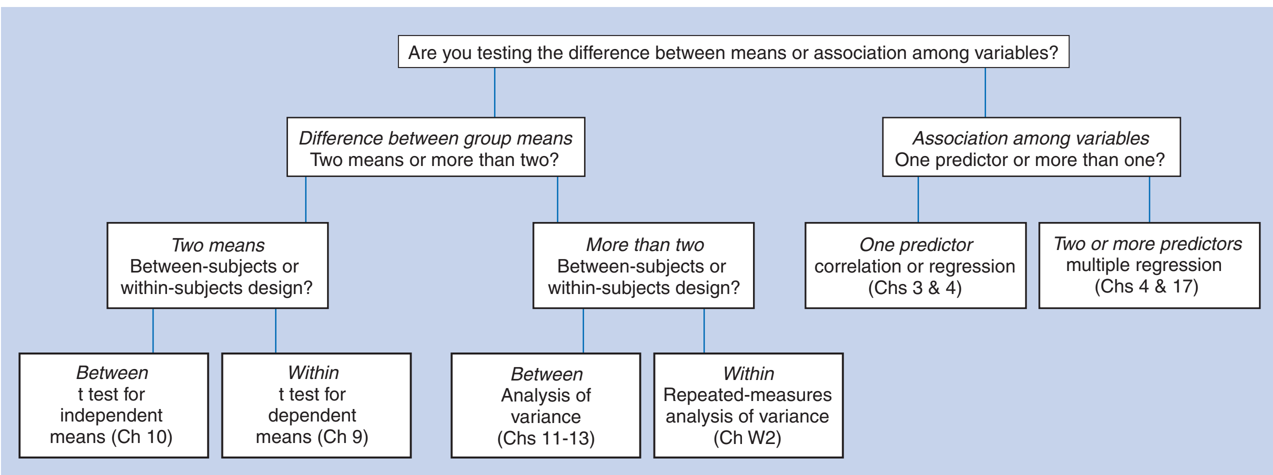

FIGURE W3-1 *Decision Tree for Identifying Statistical Tests for Studies with an Equal-Interval Dependent or Criterion Variable*

The decision tree in Figure W3-1 leads you to the correct statistical test to use in each of these situations.

WHEN YOU HAVE RANK-ORDER SCORES

Suppose your dependent or criterion variable scores are on a rank-order variable, such as place finished in a race or birth order in one's family? With rank-order scores, you may want to use one of the special rank-order tests discussed briefly in Chapter 15. Most of these are available in SPSS (under *Analyze / Nonparametric Tests*). You can use Figure W3-1 to find what you would use if you had equal interval scores, and then find the appropriate equivalent rank-order test using Table 15-4. However, as we explain in Chapter 15, in most cases you get reasonably accurate results if you just use the ordinary equal-interval statistical procedures, treating the ranks (1, 2, 3, etc.) as if they were equal-interval scores. If you want to be more accurate you can use the ordinary tests with the adjustment described in Footnote 3 of Chapter 15.

WHEN YOUR SCORES ARE CATEGORIES

Suppose your dependent or criterion variable scores are on a nominal variable, such as which of several candidates a person most favors, or which of several possible responses a child gives to a teacher's question. The standard chi-square tests for goodness of fit and for independence (see Chapter 14) cover most such situations. However, as noted in Chapter 14, you can only use these tests when each person is in a different category (that is you can not use them when you have a within-subjects design). Also, if you have a three- or more-way contingency table, you need to use procedures that are quite advanced (such as *log-linear chi-square*) and in any case often can not answer very directly the questions you might pose. Thus, if you have designed a study with a nominal dependent or criterion variable and you have repeated measures or more than a two-way table, we strongly urge you to find a way to re-design your study, perhaps by using a measure that is equal-interval instead.

However, if your nominal dependent or criterion variable has only two categories (or if you can combine categories so you end up with only two), you can then give the categories any two arbitrary numbers, such as 1 and 2, and then treat it as an ordinary equal-interval variable. But this does not work for more than two categories. (This situation and the issues involved are similar to *nominal coding*, described in Chapter 16.)

WHEN YOU HAVE MORE THAN ONE CRITERION OR DEPENDENT VARIABLES.

Most psychology research studies have a single criterion or dependent variable and the standard methods you learned in Chapters 1-16 (correlation, regression, t test, etc.) are designed for such situations. However, you will sometimes want do to studies that use more than one such variable. For example, an experiment on the effect of color of printing on reading a story might measure both time to read the story and also comprehension of the story. Or a survey might look at how mother's age predicts four criterion variables, her oldest child's grades in elementary school, junior high school, high school, and college.

Method of separate analyses. In studies like these, one solution is to use separate ordinary tests for each variable. Thus, you might run one t test comparing the effects of different colors of printing on reading time and another t test comparing the effect of the different colors on reading comprehension. Similarly you could conduct one regression analysis with mother's age predicting her oldest child's elementary school grades, another with mother's age predicting her oldest child's junior high school grades, and so on.

Method of averaging measures. Another solution is to combine the several dependent or criterion measures into a single overall measure. (This is particularly appropriate if there are high correlations among the variables being combined.) For example, you could take the average of the four kinds of grades. However, in some situations, it is not so simple. Consider what would happen if you just did a simple average of each participant's reading time and reading comprehension score. One problem is that shorter reading times are better, but higher comprehension scores are better. So before averaging, you would need to reverse one of them. For example, if reading times go from 200 to 300 seconds, you can subtract each time from 300; then a high score would mean a better time and a low score, a worse reading time. Another problem in this example is that the two variables might be on quite different scales—for example, reading time (after reversing) goes from 0 to 100, comprehension scores might go from 1 to 7. Thus, if you combined them, the reading time would probably have a bigger influence on the mean. A solution to this problem is to convert all the scores on each scale to Z scores, then average the Z scores. In this way the different measures are put on the same scale.

Multivariate tests. Yet another approach would be to carry out an overall analysis that considers all the dependent or criterion variables together. As discussed in Chapter 17, such procedures are called *multivariate statistical tests*. After reading Chapter 17 you should be able to understand what such tests do and in most cases select an appropriate procedure (for example, for the effect of colors of printing on reading time and reading comprehension, you could use a multivariate analysis of variance), carry it out in SPSS, and make some sense of the results. At the same time, however, without more advanced courses you really don't know enough to evaluate whether you have met assumptions or how to figure effect sizes. Also, Chapter 17 does not cover some multivariate procedures you might need (for example, because it is rarely used, Chapter 17 does not introduce canonical correlation, the procedure you would need for handling the regression example of mother's and father's age predicting their oldest child's various school grades).

Thus, until you take more advanced courses, it is best to use the procedures you know, either doing separate tests for each variable or combining the variables by averaging.

FIGURING POWER AND NEEDED SAMPLE SIZE

As discussed in some detail in Chapter 8, a crucial issue when planning a study is deciding if there is sufficient statistical power—that is, if the research hypothesis is true, what is the probability this study will produce a significant result supporting it? To figure the statistical power of a planned study you first need to know what statistical test you will use (see previous section). Then you have to make some estimate of the expected effect size. We discuss estimating effect size in Chapter 8—you may have a rough idea of whether the effect will be small, medium, or large based on previous similar research or a minimum effect size that would be important. If your study will use one of the major procedures we cover (t tests, correlations, etc.) there are power tables in each chapter. (Table A-5 is an index to these power tables.) If you are using a more advanced or unusual procedure that is not covered in the book, you may be able to find the power in Cohen's (1988) book of tables. If your planned study has power below about 80%, our Table 8-5 suggests several ways you can alter the study to increase the power—such as by increasing the number of participants.

You can also start with an expected effect size and use one of the tables that tell you how many participants you need for 80% power. (Table A-5 includes an index to such tables.) It often turns out that the number you need is too large to be practical. This is particularly the case since participants are often hard to recruit, not everyone you recruit will qualify, and if you need to test participants on more than one occasion, you are likely to lose many after the first testing. Thus, you may need to use one of the other methods listed in Table 8-5 to increase power or change the basic study itself.

CONDUCTING THE STUDY

From the point of view of the statistical analyses, there are three main things to keep in mind during the data gathering: (a) When you set up the lay out of the questionnaires or data recording sheets, consider how easy it will be to type the scores into the computer; (b) keep all of the questionnaires or data recording sheets in a safe place; and (c) make sure you have recorded everything important on them—such as date and time of the study, experimental condition, and anything else that is not completed by the participants. Now and then a student conducts a study and when it is done discovers that there is no record of which participant was in which condition! It is also important to record anything that happens with a participant that is unusual.

ENTERING SCORES INTO THE COMPUTER

As for entering the scores into a computer, here are some suggestions from our long experience to make the process as straightforward and accurate as possible.

Setting up the spreadsheet. With SPSS and most statistical programs you type the scores on to a spreadsheet with one line per participant and one column per variable. Give some thought in advance as to how you will set it up. For example, put the columns for the variables in the same order as they are on the raw sheets that you are typing them in from.

Variable names. It is important to type at the top of each column the descriptive name for each variable (rather than leave the variable named by the column number)—for example, "Age" is better than "Variable 1"! In SPSS you have to keep your variable name to 8 characters, so use abbreviations you will remember later (sometimes much later).

Sums and codes. To avoid work and errors, enter the scores directly and let the computer do any needed combining or changing. For example, if you have a 10-item questionnaire, it is better to enter the person's score on each item and let the computer figure the total or average. Similarly, if the computer needs number codes for a variable that was originally entered as words or letters, enter the words or letters—for example, "f" or "m" for gender, and then instruct the computer to make an additional column with number codes it creates for these. That way you will make fewer mistakes and also you won't forget later which number goes with which letter!

Typing in the scores. Have all material organized and at hand. Type it in all at once. Save often. Keep notes of any irregularities (for example, a participant who answers 8 on a 1-7 scale) or any decisions you make (for example, that if a person marks between two scale points, you will always score it as the higher scale point). When you are done, make a printout of the data file (and a copy on a disk)

Check your work. Accuracy is extremely important. If possible have someone else double check the entries for at least a few of your participants.

DATA SCREENING

Experienced researchers resist the urge to jump right into the analysis as soon as all the scores are entered. First, they "screen" the data for accuracy, missing values, and whether the assumptions for the planned statistical tests are met. That way they don't get excited (or depressed) about their conclusions, then find later that all their conclusions had no relation to the real results and it all has to be done again.

CHECKS FOR ACCURACY

Even after you have double checked your typing, errors are still common. With so many numbers, it is just too easy to make mistakes (even big ones that drastically change results—such as hitting the 0 key too long so you enter 100 instead of 10!). The most important additional check is based on a listing of each variable's number of cases, mean, standard deviation, and maximum and minimum score. (In SPSS, you can get such a listing with *Analyze / Descriptive Statistics / Descriptives*, and then selecting all of your numeric variables.) If you have more than one grouping, such as an experimental and a control group, you may want a listing for each group separately. (In SPSS, you can get separate outputs for each grouping with *Data / Split File*, checking "Organize output by groups," and selecting the grouping variable.)

Once you have this listing, for each variable you look at (a) the number of cases, to be sure that all or most of the participants have scores on the variable; (b) the maximum, to be sure that none of the scores are higher than is possible (such as a 70 on a scale that goes from 1 to 7); (c) the minimum, to be sure that none of the score are lower than possible (such as a minus value for number of children); and (d) the mean and standard deviation, to be sure they seem reasonable in light of other similar variables in your study and what you know about these measures.

If you have any nominal variables, you can do a similar check with a listing for each such variable of how many participants fall into each category. This is also a good idea for numeric variables that have only a few possible whole number values, such as a 1-5 scale or number of siblings. (In SPSS, you can get such a listing with *Analyze / Descriptive Statistics / Frequencies* and then selecting the desired variables.) For each variable, check whether the numbers in the different categories seem reasonable.

MISSING VALUES

Often participants don't answer every question on a questionnaire, observers miss a particular behavior, recording devices fail, or a participant's results are lost. These kinds of situations create *missing values*.

You can tell how many missing values you have from the number of cases for each variable in the variable listing you make for your accuracy check (as described in the previous subsection). It is also a good idea to print out your data file (in SPSS, with the data on the screen, select *Print* from the *File* menu). You can then see whether any particular participants have most of missing values or the missing values are spread among different participants.

In general, what you do about missing values depends on whether you think they are missing for some systematic reason or just haphazardly. For example, if the question is about a sensitive topic, those who don't answer might be systematically different in relation to this topic than those who do. In such situations, you either have to drop any analyses involving the variable or go ahead with just those participants who answered on it, but keeping in mind that your results only apply to the kinds of people who are likely to answer such questions.

Suppose there are only a few missing values and the pattern seems more or less random? In this situation, a common procedure is to substitute for each missing value, the average of everyone else's score on that variable. (If you were just to leave them missing, it could mean that you would exclude any participant who had a missing value on any variable, sometimes resulting in losing most of your participants from the analysis!) If the study has groups, you can substitute the value for the avenge of that group. If a person has a missing value on an item that is part of a multi-item measure (such as one item not answered on a 12-item anxiety questionnaire), you might substitute that person's average on the items that were completed.

CHECKING FOR NORMAL DISTRIBUTIONS

Most statistical tests require that the variables be normally distributed in the population (see Chapter 15). There are also additional assumptions in many procedures, such as equal population variances, which should be checked before carrying out a particular analysis. But we focus here on normality because it applies to nearly all statistical tests.

Checking for skewness. Skewness means a distribution is not normal because it is lopsided with a long tail on one side (see Chapter 1). Thus, as part of data screening, you check each variable for its degree of skewness. First, you compute each variable's numerical skewness value; perfect normality is 0. (In SPSS, you can get skewness from *Analyze / Descriptive Statistics / Descriptives / Options* and checking "Skewness".) Then, for any variable that has a very high or low skewness value-say more extreme than ±1, you make a histogram and look at it visually to see if it looks seriously skewed. (In SPSS you make a histogram with *Graphs / Histogram* and selecting the variable.)

What to do about seriously skewed distributions. Presuming they are not caused by one or a few outliers (see below), then you can use one of the strategies spelled out in Chapter 15, such as transforming the scores or using a rank-order method.

Checking for outliers. Most statistical tests are robust to skewness that is not too extreme. The main cause of extreme skewness are outliers, which are problematic for almost all statistical tests. You check for outliers by looking for *very* high skewness values or for very long tails or separated scores in the histograms. Another method is to figure Z scores for your variables. (In SPSS, you can get Z scores from *Analyze / Descriptive Statistics / Descriptives* and checking "Save standardized values as variables"—this gives for each variable a new column that has the Z scores for that variable.) Then you can look down the columns of Z scores for any that are extreme—say more than ±3.

What to do about outliers. If you find an outlier, look at the raw questionnaire or data record to be sure it is not an error in typing in the score. Next consider whether there is something about the participant that might make him or her not part of the population to which you want to apply your results. (For example, in a college student sample, this person might be 45 years old when everyone else is 18-22.) In that situation you can just exclude the participant from the analysis. If the person really is part of the population, one option is to use a rank-order test. Another common option is to recode the extreme score so that it is still the most extreme score but just slightly higher than the next highest score. (For example, if this person's reading time was 300 seconds and the next highest was 141 seconds, you could recode the person as 142.) Some researchers simply make it a practice to exclude participants whose scores are very extreme. Whatever solution you adopt, it is very important to include in your research report a description of what you did and why. (One huge advantage of screening data before analyzing results is that you are in a better position to make such decisions without having to worry that you are unconsciously biasing the outcome in your favor.)

CARRYING OUT THE MAJOR ANALYSES

The most important advice here is, again, be sure you have thoroughly screened your data before you begin! Once you are ready, the next most important advice is to *look* at your data. Look at the means overall and by groups, at the histograms, at scatter diagrams, at patterns of correlations. Get to know your data.

Next, write out a systematic analysis plan—what analyses you are going to carry out and in what order—and follow it. Hopefully you have laid this out in advance when you designed the study. Be sure that your plan focuses first and foremost on the hypotheses or research questions with which you began the study. Only then conduct the analyses.

When you look at each output, before looking at the results part, be sure that the computer used the variables you intended, that it included all of the participants, and that it did the analysis you intended.

Once you have the major results, then it is a good idea to explore. But even here, it is wise to write out a list of the exploratory analyses you will do. Many of the most important discoveries in psychology came not from what was predicted in advance, but from unexpected findings in these explorations. Remember, however, that findings from exploration are more likely to be chance findings. It is like the problem of multiple comparisons (see chapters 11 and 12) or of too many t tests (see Chapter 10); with many tests a few will come out significant just by chance. Thus, any findings from exploratory analyses need to be labeled as such when you write them up in your report and need to be taken as very tentative until they are replicated in a new study.

WRITING UP YOUR RESULTS

Some of your results actually go in the methods section. These would include information on your participants (mean and standard deviations for age and any other relevant background variables, number of each gender, etc.) and any reliability analyses on your measures (see Chapter 17). The rest of your results go in the results section. Usually you begin with descriptive statistics-means and standard deviations of your major variables. Then you describe each analysis in a systematic fashion, starting with those testing your hypotheses and research questions and then turning to exploratory analyses. For each analysis you make clear what hypothesis or research question or exploratory issue it is designed to test; describe the analysis (what kind of analysis it is, such as a correlation or t test for independent means, and what variables are involved); and give the results, including means, standard deviations, significance results (with degrees of freedom and p values), and effect sizes. Wherever it would make results clearer, use tables and graphs. There are examples of how this is done in the "in research articles" sections of each chapter.

After the results section, there is usually a Discussion section where you summarize the key conclusions, describe how your results bear on the larger issues the study was designed to address (that is, how is what we know about what you are studying different now than it was before you did the study and how does this bear on previous research and theory), and note the limitations of your study and anomalous results. Here, it is important not to get bogged down in explaining failures to get significance more than briefly. And remember, a nonsignificant result means "inconclusive," not anything like showing "no difference" (unless you had very high power). Finally, consider the implications for future research. Above all, in the Discussion, it is important to keep to the big picture.