CHAPTER W2

Applying Statistical Methods in Your Own Research Project

Chapter Outline

TIP FOR SUCCESS

Before reading this chapter, you should have read as many of the previous chapters as possible. You should also have tackled the Using SPSS sections that are found at the end of several chapters.

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 in your own research. 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, it will help to have some additional points to apply this smoothly to your own research. In particular, in this 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 on a personality test), 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: Selecting A 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 Web Chapter W1).

Before starting out on a new study, experienced researchers plan 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 available will be less than ideal for some reason. You also need to decide in advance what statistical test you will use in order to figure out how many participants you will need (sample size) and the power of your planned study.

What Test to Use in the Usual Situation of Equal-Interval Measurement

In the most common research situation, your scores are measured on an approximately equal-interval scale (see Chapter 1 for a discussion of "levels of measurement"). Figure W2–1 shows a decision tree for deciding on an appropriate statistical test in this standard situation. Answering the questions in the decision tree will guide you to the appropriate statistical test. (Figure W2–1 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 question in the decision tree is: Are you testing the *difference between means* or *the association among variables*? For example, consider a study comparing the effect of different colors of printing on reading time. This study focuses on the mean reading time for each color; thus, this is a study comparing means of variables. On the other hand, consider a study looking at the relation of mother's age to her oldest child's school grades (or, to put this another way, whether mother's age predicts her oldest child's school grades). This study is about associations among variables.

Suppose 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

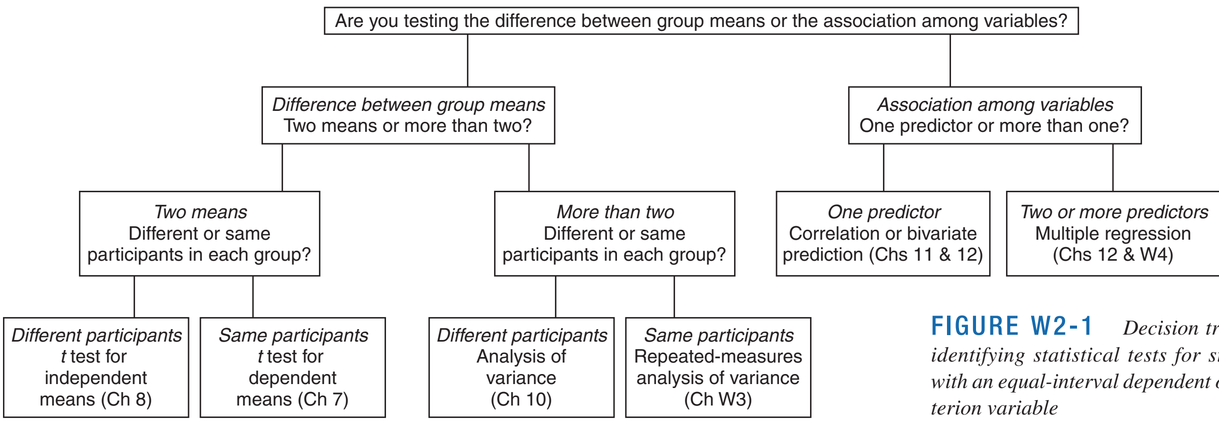

FIGURE W2-1 *Decision tree for identifying statistical tests for studies with an equal-interval dependent or criterion variable* printing vs. black printing) or whether there are three or more means (for example, if you were looking at differences among printing in red, black, green, and blue). A *t* test is the appropriate test when there are two means being compared and an analysis of variance is the appropriate test with three or more means. Once you have decided on a *t* test or analysis of variance, you have to make another decision. You next must decide whether this is a between-subjects design or a within-subjects design (also called a repeated-measures design). In a between-subjects design different people are in each group (for example, some participants will read a story in black printing and other participants will read the story in red printing). In a within-subjects design the same people will be in each group (for example, each participant will read one story in black printing and another story in red printing).

What if your study focuses on *associations or predictions*? In that case, the next question is whether there is one variable being correlated with or predicting the other (for example, mother's age predicting or associated with her oldest child's school grades) or more than one variable being correlated with or predicting the other (for example, mother's age and father's age predicting or being associated with oldest child's school grades). You use correlation or bivariate prediction when there is one predictor variable. You use multiple regression when there are two or more predictor variables. Notice that in all these situations we are correlating with or predicting about a single variable. In the language of prediction (regression) there is only one criterion variable (in the example it is oldest child's school grades). We will have more to say about this situation shortly.

The decision tree in Figure W2–1 leads you to the correct statistical test to use in each of these situations.

What Test To Use When Your Scores Are Categories

Suppose you are testing a hypothesis with a variable whose scores are categories (that is, it is a nominal variable). Examples of categorical variables are which of several candidates a person most favors, or which major a student is taking. The standard chi-square tests for goodness of fit and for independence (see Chapter 13) cover most such situations. However, as noted in Chapter 13, you can only use these tests when each person is in a single category on any one nominal variable. So, for example, it would not be appropriate to use a chi-square test to examine the change in distribution of men's and women's preferred brand of shampoo from before to after a new advertising campaign. 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 one or more nominal variables and any one person can be in more than one category on a particular nominal variable, or you have 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.

Incidentally, there is a trick you can use when your nominal variable has only two categories, such as male and female (or if you can combine categories so you end up with only two, as in whether various majors are arts or science). In this situation you can then give the two categories any two arbitrary numbers, such as 1 and 2, and then treat the variable as an ordinary equal-interval variable so you can use *t* tests, correlation, and so on. But this does not work for more than two categories; also you should not use this method if the split between the two categories is quite extreme (say more than 80% in one of the groups).

What Test to Use When You Have Rank-Order Scores

Suppose your are interested in carrying out hypothesis testing with scores for 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 14. (You will likely learn more about such tests in intermediate statistics courses.) Most of these are available in SPSS (by selecting *Analyze* and then *Nonparametric Tests*). You can use Figure W2–1 to find what test you would use if you had equal interval scores, and then find the appropriate equivalent rank-order test using Table 14–4. However, as you learned in Chapter 14, 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 Chapter Note 3 of Chapter 14.

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 1–14 (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. We focus on three potential solutions for handling such research situations: method of separate analyses; method of averaging measures; and multivariate tests.

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 criterion or dependent 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 the reading example, you would still have to reverse one of the measures before figuring 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 criterion or dependent variables together. Such procedures are called *multivariate statistical tests* and we describe them briefly in Chapter 15. After reading that chapter you should be able to understand what such tests do and make some sense of the results you may read in a research article. You can learn more about multivariate statistical tests in more advanced statistics courses. So, until you take such courses, it is best to use the procedures you know, either doing separate tests for each variable or combining the variables by averaging. (In fact, the general advice even to experienced researchers who are masters of advanced statistical methods, is to use the simplest methods available that are reasonably accurate. The idea is that with simpler procedures results will be more understandable to readers and also the researcher is less likely to make mistakes.)

Figuring Power and Needed Sample Size

As discussed in some detail in Chapter 6, a crucial issue when planning a study is deciding if there is sufficient statistical power. That is, power tells you, 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 decide what statistical test you will use. Then you have to make some estimate of the expected effect size. We discuss estimating effect size in Chapter 6—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, analysis of variance, chi-square tests) there are power tables in each chapter. (Table A-5 in the Appendix 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 or by using power statistical software or an Internet power calculator. If your planned study has power below about 80%, our Table 6–6 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 7–6 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 layout 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 experimental 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 questionnaires or data recording 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). Use a name that describes the particular variable—for example, "age" and "gender" are better than "variable1" or "question2"! Until a few years ago, SPSS limited variable names to a maximum of 8 characters. Recent versions of SPSS allow longer variable names, but we still recommend that you try not to use variable names that are longer than about 10–12 characters. (Otherwise it gets hard to read the names in the columns and the printouts of results are very lengthy.)

Sums and Codes

To avoid work and errors, enter the scores directly and let the program do any needed combining figuring. For example, if you have a 10-item questionnaire, it is better to enter the person's score on each item and let the program figure the total or average. Similarly, if the program 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 program 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. If possible, type in all the data at once. Save often. Keep notes of any irregularities (for example, a participant who answers 8 on a 1 to 7 scale) or any decisions you make (for example, that if a person marks between two scale points, you have decided you will always score it as the higher scale point). When you are done, be sure to save a copy of the data file in more than one location (for example, on the hard drive of your computer and on a CD or memory stick).

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 by selecting *Analyze*, *Descriptive Statistics*, *Descriptives*, then selecting all of your numeric variables and clicking *OK*.) 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 by selecting *Data*, *Split File*, checking "Organize output by groups," selecting the grouping variable and clicking *OK*.)

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 to 5 scale or number of siblings. (In SPSS, you can get such a listing by selecting *Analyze*, *Descriptive Statistics*, *Frequencies*, and then selecting the desired variables and clicking *OK*.) 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.* (When entering your data, it is usually best to leave the place blank where a missing value would go. If instead you put in some number like 999 or –1, it could end up being included in the figuring later!)

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 *File* and then *Print*). You can then see whether any particular participants have most of the 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 have two choices: You can just not do any analyses involving that variable, or you can 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—something SPSS does automatically most of the time—sometimes resulting in losing most of your participants from the analysis!) If the study has groups, you can substitute the value for the average 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 of the scale that he or she completed. When substituting missing values it is a good idea to create a new variable (with its own column on the SPSS spreadsheet) that includes the substituted values, keeping the original variable as it was. That way you don't lose track of exactly when you did and did not substitute.

Checking for Normal Distributions

Most statistical tests require that the variables be normally distributed in the population (see Chapter 14). 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 figure each variable's numerical skewness value; perfect normality is 0. (In SPSS, you can get skewness by selecting *Analyze*, *Descriptive Statistics*, *Descriptives*, *Options,* checking "Skewness", and then clicking *Continue* followed by *OK*.) 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 by selecting *Graphs, Legacy Dialogs, Histogram*.)

What To Do About Seriously Skewed Distributions

Presuming skewed distributions are not caused by one or a few very extreme cases (see below), then you can use one of the strategies spelled out in Chapter 14, such as transforming the scores or using a rank-order method.

Checking for Outliers (extreme scores)

Most statistical tests still give accurate results as long as skewness 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 by selecting *Analyze, Descriptive Statistics, Descriptives,* checking "Save standardized values as variables" and clicking *OK*. This gives for each variable a new column that has the *Z* scores for that variable. These are not exactly the same as the *Z* scores we figured in Chapter 3 because they use the $N - 1$ formula when figuring the standard deviation. But unless your sample is very small, they are very similar. And for purposes of data screening, they are just fine.) 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 aged 18-22.) In that situation you can just exclude the participant from the analysis. If the person really is part of what you consider the relevant 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 too many t tests (see Chapter 8); 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

When you come to write up the results of your research study, it will usually be in the form of a paper or report with the following main sections: Introduction, Methods, Results, Discussion. 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, and so on) and any reliability analyses on your measures (see Chapter 15). 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. You will learn more about writing up your results in research methods courses. We strongly recommend that you take one or more such courses.

1. Before conducting a research study, it is important to plan what statistical method(s) will be used to analyze the results.
2. When scores are measured on an apparently equal-interval scale, you must first decide whether you are testing the difference between means or the association among variables. If the focus is on differences between means, you have to determine whether you are comparing two means (in which case a *t* test is the appropriate test) or three or more means (for which an analysis of variance is the appropriate test), and you must also decide whether you have a between-subjects or a within-subjects (repeated-measures) design. If the focus is on the association among variables, you must decide whether there is one variable being correlated with or predicting another variable (in which case correlation or bivariate prediction are appropriate tests) or more than one variable correlated with or predicting another variable (for which multiple regression is the appropriate test).
3. Chi-square tests are used when your scores are categories, but they can only be used when each person is in a single category on any one nominal variable. Special rank-order tests (discussed briefly in Chapter 14) are used when your scores are measured on a rank-order scale.
4. When you have scores on one or more criterion or dependent variables, you can handle this situation using one of three approaches: (a) conduct separate analyses for each criterion or dependent variable; (b) create a single measure by averaging the scores across the criterion or dependent variables; (c) use a multivariate test that considers all the criterion or dependent variables together.

5. It is important to estimate the statistical power of a study before carrying out the actual study.
6. There are a number of factors you should consider when entering the scores from your study into the computer, including setting up the spreadsheet carefully, using appropriate variable names, letting the computer do any summing or variable coding, typing in the data accurately, and checking your work.
7. Before carrying out the main statistical analyses, you should screen your data in order to check for accuracy, to identify and handle missing values, and to check for normal distributions.
8. After carrying out your plan for the main statistical analyses, you may want to consider conducting some exploratory analyses.
9. When writing up the results of a research study, most of the results of the statistical analyses will go in the Results section of the paper or report, although some information about the study participants is usually given in the Methods section. The implications of the study results and conclusions are described in the Discussion section.