

# Testing Group Differences using T-tests, ANOVA, and Nonparametric Measures

Jamie DeCoster

Department of Psychology  
University of Alabama  
348 Gordon Palmer Hall  
Box 870348  
Tuscaloosa, AL 35487-0348

Phone: (205) 348-4431  
Fax: (205) 348-8648

January 11, 2006

I would like to thank Anne-Marie Leistico and Angie Maitner for comments made on an earlier version of these notes. If you wish to cite the contents of this document, the APA reference for them would be

DeCoster, J. (2006). *Testing Group Differences using T-tests, ANOVA, and Nonparametric Measures*. Retrieved (month, day, and year you downloaded the notes, without the parentheses) from <http://www.stat-help.com/notes.html>

For future versions of these notes or help with data analysis visit  
<http://www.stat-help.com>

ALL RIGHTS TO THIS DOCUMENT ARE RESERVED

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Testing One or two Means</b>	<b>5</b>
<b>3</b>	<b>Testing a Single Between-Subjects Factor</b>	<b>15</b>
<b>4</b>	<b>Testing Multiple Between-Subjects Factors</b>	<b>30</b>
<b>5</b>	<b>Testing a Single Within-Subjects Factor</b>	<b>47</b>
<b>6</b>	<b>Testing Multiple Within-Subjects Factors</b>	<b>54</b>

# Chapter 1

## Introduction

### 1.1 Data and Data Sets

- The information that you collect from an experiment, survey, or archival source is referred to as your *data*. Most generally, data can be defined as a list of numbers possessing meaningful relations.
- For analysts to do anything with a group of data they must first translate it into a *data set*. A data set is a representation of data, defining a set of “variables” that are measured on a set of “cases.”
  - A *variable* is simply a feature of an object that can be categorized or measured by a number. A variable takes on different *values* to reflect the particular nature of the object being observed. The values that a variable takes will vary when measurements are made on different objects at different times. A data set will typically contain measurements on several different variables.
  - Each time that we record information about an object we create a *case* in the data set. Like variables, a data set will typically contain multiple cases. The cases should all be derived from observations of the same type of object, with each case representing a different example of that type. Cases are also sometimes referred to as *observations*. The object “type” that defines your cases is called your *unit of analysis*. Sometimes the unit of analysis in a data set will be very small and specific, such as the individual responses on a questionnaire. Sometimes it will be very large, such as companies or nations. The most common unit of analysis in social science research is the participant or subject.
- If the value of a variable has actual numerical meaning (so that it measures the amount of something that a case has), it is called a *continuous* variable. An example of a continuous variable measured on a planet would be the amount of time it took to orbit the sun. If the value of a variable is used to indicate a group the case is in, it is called a *categorical* variable. An example of a categorical variable measured on a country would be the continent in which it is located.

In terms of the traditional categorizations given to scales, a continuous variable would have either an interval, or ratio scale, while a categorical variable would have a nominal scale. Ordinal scales sort of fall in between. Bentler and Chou (1987) argue that ordinal scales can be reasonably treated as continuous as long as they have four or more categories. However, if an ordinal variable only has one or two categories you are probably better off either treating it as categorical, or else using procedures specifically designed to handle ordinal data.

- When describing a data set, you should name your variables and state whether they are continuous or categorical. For each continuous variable you should state its units of measurement, while for each categorical variable you should state how the different values correspond to different groups. You should also report the unit of analysis of the data set. You typically would not list the specific cases, although you might describe where they came from.
- When conducting a statistical analysis, you often want to use values of one variable to predict or explain the values of another variable. In this case the variable you look at to make your prediction is

called the *independent variable* (IV) while the variable you are trying to predict is called the *dependent variable* (DV).

## 1.2 General information about statistical inference

- Most of the time that we collect data from a group of subjects, we are interested in making inferences about some larger group of which our subjects were a part. We typically refer to the group we want to generalize our results to as the *theoretical population*. The group of all the people that could potentially be recruited to be in the study is the *accessible population*, which ideally is a representative subset of the theoretical population. The group of people that actually participate in our study is our *sample*, which ideally is a representative subset of the accessible population.
- Very often we will calculate the value of an *estimate* in our sample and use that to draw conclusions about the value of a corresponding *parameter* in the underlying population. Populations are usually too large for us to measure their parameters directly, so we often calculate an estimate from a sample drawn from the population to give us information about the likely value of the parameter. The procedure of generalizing from data collected in a sample to the characteristics of a population is called *statistical inference*.

For example, let us say that you wanted to know the average height of 5th grade boys. What you might do is take a sample of 30 5th grade boys from a local school and measure their heights. You could use the average height of those 30 boys (a sample estimate) to make a guess about the average height of all 5th grade boys (a population parameter).

- We typically use Greek letters when referring to population parameters and normal letters when referring to sample statistics. For example, the symbol  $\mu$  is commonly used to represent mean value of a population, while the symbol  $\bar{X}$  is commonly used to represent the mean value of a sample.
- One type of statistical inference you can make is called a *hypothesis test*. A hypothesis test uses the data from a sample to decide between a *null hypothesis* and an *alternative hypothesis* concerning the value of a parameter in the population. The null hypothesis usually makes a specific claim about the parameters (like saying that the average height of 5th grade boys is 60 inches), while the alternative hypothesis says that the null hypothesis is false. Sometimes the alternative hypothesis simply says that the null hypothesis is wrong (like saying that the average height of 5th grade boys is *not* 60 inches). Other times the alternative hypothesis says the null hypothesis is wrong in a particular way (like saying that the average height of 5th grade boys is *less than* 60 inches).

For example, you might have a null hypothesis stating that the average height of 5th grade boys is equal to the average height of 5th grade girls, and an alternative hypothesis stating that the two heights are not equal. You could represent these hypotheses using the following notation.

$$H_0 : \mu\{\text{boys}\} = \mu\{\text{girls}\}$$

$$H_a : \mu\{\text{boys}\} \neq \mu\{\text{girls}\}$$

where  $H_0$  refers to the null hypothesis and  $H_a$  refers to the alternative hypothesis.

- To actually perform a hypothesis test, you must collect data from a sample drawn from the population of interest that will allow you to discriminate your hypotheses. For example, if your hypotheses involve the value of a population parameter, then your data should provide a sample estimate corresponding to that parameter. Next you calculate a *test statistic* that reflects how different the data in your sample are from what you would expect if the null hypothesis were true. You then calculate a *p-value* for the statistic, which is the probability that you would get a test statistic as extreme as you did due to random chance alone if the null hypothesis was actually true. To get the p-value we must know the distribution from which our test statistic was drawn. Typically this involves knowing the general form of the distribution (t, F, etc.) and the degrees of freedom associated with your test statistic. If the p-value is low enough, you reject the null hypothesis in favor of the alternative hypothesis. If the probability is high, you fail to reject the null hypothesis. The breakpoint at which you decide whether to accept or reject the null hypothesis is called the *significance level*, and is often indicated by the

symbol  $\alpha$ . You typically establish  $\alpha$  before you actually calculate the p-value of your statistic. Many fields use a standard  $\alpha = .05$ . This means that you would have a 5% chance of obtaining results as inconsistent with the null hypothesis as you did due to random chance alone.

- When reporting p-values you should use the following guidelines.
  - Use exact p-values (e.g.,  $p = .03$ ) instead of confidence levels (e.g.,  $p < .05$ ) whenever possible.
  - If the p-value is greater than or equal to .10, use two significant digits. For example,  $p = .12$ .
  - If the p-value is less than .10, use one significant digit. For example,  $p = .003$ .
  - Never report that your p-value is equal to 0. There is always some probability that the results are due to chance alone. Some software packages will say that your p-value is equal to 0 if it is below a specific threshold. In this case you should report that the p-value is less than the threshold. SPSS will report that your p-value is equal to .000 if it is less than .001. If you see this, then you should simply report that  $p < .001$ .
- To summarize the steps to a hypothesis test
  1. Determine your null and alternative hypotheses.
  2. Draw a sample from the population of interest.
  3. Collect data that can be used to discriminate the null and alternative hypotheses.
  4. Calculate a test statistic based on the data.
  5. Determine the p-value of your test statistic.
  6. State whether you reject or fail to reject the null hypothesis.

### 1.3 Working within SPSS

- Most users typically open up an SPSS data file in the data editor, and then select items from the menus to manipulate the data or to perform statistical analyses. This is referred to as the *interactive mode*, because your relationship with the program is very much like a personal interaction, with the program providing a response each time you make a selection. If you request a transformation, the data set is immediately updated. If you select an analysis, the results immediately appear in the output window.
- It is also possible to work with SPSS in *syntax mode*, where the user types code in a syntax window. Once the full program is written, it is then submitted to SPSS to get the results. Working with syntax is more difficult than working with the menus, because you must learn how to write the programming code to produce the data transformations and analyses that you want. However, certain procedures and analyses are only available through the use of syntax. For example, vectors and loops (described later) cannot be used in interactive mode. You can also save the programs you write in syntax. This can be very useful if you expect to perform the same or similar analyses multiple times, since you can just reload your old program and run it on your new data (or your old data if you want to recheck your old analyses). If you would like more general information about writing SPSS syntax, you should examine the *SPSS Base Syntax Reference Guide*.
- Whether you should work in interactive or syntax mode depends on several things. Interactive mode is best if you only need to perform a few simple transformations or analyses on your data. You should therefore probably work interactively unless you have a specific reason to use syntax. Some reasons to choose syntax would be:
  - You need to use options or procedures that are not available using interactive mode.
  - You expect that you will perform the same manipulations or analyses on several different data sets and want to save a copy of the program code so that it can easily be re-run.
  - You are performing a very complicated set of manipulations or analyses, such that it would be useful to document all of the steps leading to your results.

- Whenever you make selections in interactive mode, SPSS actually writes down syntax code reflecting the menu choices you made in a “journal file.” The name of this file can be found (or changed) by selecting **Edit** → **Options** and then selecting the **General** tab. If you ever want to see or use the code in the journal file, you can edit the journal file in a syntax window.

SPSS also provides an easy way to see the code corresponding to a particular menu function. Most selections include a **Paste** button that will open up a syntax window containing the code for the function, including the details required for any specific options that you have chosen in the menus.

You can have SPSS place the corresponding syntax in the output whenever it runs a statistical analysis. To enable this

- Choose **Edit** → **Options**.
- Select the **Viewer** tab.
- Check the box next to **Display commands in the log**.
- Click the **OK** button.

## Chapter 2

# Testing One or two Means

### 2.1 Comparing a mean to a constant

- One of the simplest hypothesis tests that we can perform is a comparison of an estimated mean to a constant value. We will therefore use this as a concrete example of how to perform a hypothesis test.
- You use this type of hypothesis test whenever you collect data from a sample and you want to use that data to determine whether the average value on some variable in the population is different from a specific constant. For example, you could use a hypothesis test to see if the height of 5th grade Alabama boys is the same as the national average of 60 inches.
- You would take the following steps to perform this type of hypothesis test. This is formally referred to as a *one-sample t-test for a mean*.

- *Determine your null and alternative hypotheses.* The point of statistical inference is to draw conclusions about the population based on data obtained in the sample, so these hypotheses will always be written in terms of the population parameters. When comparing a mean to a constant, the null hypothesis will always be that the corresponding population parameter is equal to the constant value. The alternate hypothesis can take one of two different forms. Sometimes the alternative hypothesis simply says that the null hypothesis is false (like saying that the average height of 5th grade boys in Alabama is *not* 60 inches. This is referred to as a *two-tailed hypothesis*, and could be presented using the following shorthand.

$$H_0 : \mu = 60$$

$$H_a : \mu \neq 60$$

On the other hand, someone might want the alternative hypothesis to say that the null hypothesis is wrong in a particular way (like saying that the average height of 5th grade boys in Alabama is *less than* 60 inches. This is referred to as a *one-tailed hypothesis*, and could be presented using the following shorthand.

$$H_0 : \mu = 60$$

$$H_a : \mu < 60$$

- *Draw a sample from the population of interest.* We must draw a sample from the population whose mean we want to test. Since our goal is to compare the average height of 5th grade boys from Alabama to the value of 60 inches, we must find a sample of Alabama boys on which we can make our measurements. This sample should be representative of our population of all of the 5th grade boys in Alabama.
- *Collect data that can be used to discriminate the null and alternative hypotheses.* You must next collect information from each subject on the variable whose mean you wish to test. In our example, we would need to measure the height of each of the boys in our sample.

- *Calculate a test statistic based on the data.* We need to compute a statistic that tells us how different our observed data is from the value under the null hypothesis. To determine whether a population mean is different from a constant value we use the test statistic

$$t = \frac{\bar{X} - X_0}{\frac{s_X}{\sqrt{n}}}, \quad (2.1)$$

where  $\bar{X}$  is the mean value of the variable you want to test in the sample,  $X_0$  is the value you want to compare the mean to (which is also the expected value under the null hypothesis),  $s_X$  is the standard deviation of the variable found in the sample, and  $n$  is the number of subjects in your sample. The fraction  $\frac{s_X}{\sqrt{n}}$  is called the *standard error of the mean*.

The mean  $\bar{X}$  can be computed using the formula

$$\bar{X} = \frac{\sum X_i}{n}, \quad (2.2)$$

while the standard deviation  $s_X$  can be computed using the formula

$$s_X = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}. \quad (2.3)$$

We divide by  $n - 1$  instead of  $n$  in our formula for the standard deviation because the standard deviation is supposed to be computed around the mean value in the population, not the value in the sample. The observations in your sample tend to be slightly closer to the sample mean than they would be to the population mean, so we divide by  $n - 1$  instead of  $n$  to increase our estimate back to the correct value. The square of the standard deviation  $s_X^2$  is called the *variance*.

- *Determine the probability that you would get the observed test statistic if the null hypothesis were true.* The statistic we just computed follows a  $t$  distribution with  $n - 1$  degrees of freedom. This means that if we know the value of the  $t$  statistic and its degrees of freedom, we can determine the probability that we would get data as extreme as we did (i.e., a sample mean at least as far away from the comparison value as we observed) due to random chance alone. The way we determine this probability is actually too complex to write out in a formula, so we typically either use computer programs or tables to get the p-value of the statistic.
- *State whether you reject or fail to reject the null hypothesis.* The last thing we would do is compare the p-value for the test statistic to our significance level. If the p-value is less than the significance level, we reject the null hypothesis and conclude that the mean of the population is significantly different from the comparison value. If the p-value is greater than the significance level, we fail to reject the null hypothesis and conclude that the mean of the population is not significantly different from the comparison value.
- The one-sample t test assumes that the variable you are testing follows a normal distribution. This means that the p-values you draw from the table will only be accurate if the population of your variable has a normal distribution. It's ok if your sample doesn't look completely normal, but you would want it to be generally bell-shaped and approximately symmetrical.
- Once you perform a one-sample t-test, you can report the results in the following way.

1. *State the hypotheses.*

The hypotheses for a two-tailed test would take the following form.

$$H_0 : \mu = X_0$$

$$H_a : \mu \neq X_0$$

Here,  $X_0$  represents the value you want to compare your mean to, also referred to as the "expected value under the null hypothesis." The hypotheses for a one-tailed test would be basically the same, except the alternative hypothesis would either be  $\mu < X_0$  or  $\mu > X_0$ .



2. Report the test statistic.

$$t = \frac{\bar{X} - X_0}{\frac{s_x}{\sqrt{n}}} \quad (2.4)$$

3. Report the degrees of freedom.

$$\text{df} = n - 1$$

4. Report the p-value.

The p-value for the test statistic is taken from the  $t$  distribution with  $n - 1$  degrees of freedom. If you perform a two-tailed test, the p-value should represent the probability of getting a test statistic either greater than  $|t|$  or less than  $-|t|$ . If you perform a one-tailed test, the p-value should represent the probability of getting a test statistic more extreme in the same direction as the alternate hypothesis. So, if your alternate hypothesis says that  $X < X_0$ , then your p-value should be the probability of getting a  $t$  statistic less than the observed value. A table containing the p-values associated with the  $t$  statistic can be found inside the back cover of Moore & McCabe (2003).

5. State the conclusion.

If the p-value is less than the significance level you conclude that the mean of the population is significantly different from the comparison value. If the p-value is greater than the significance level you conclude that the mean of the population is not significantly different from the comparison value.

- You can use the following Excel formula to get the p-value for a  $t$  statistic.

`=tdist(abs(STAT),DF,TAILS)`

In this formula you should substitute the value of your  $t$  statistic for **STAT**, the degrees of freedom for **DF**, and the number of tails in your test (either 1 or 2) for **TAILS**. The cell will then contain the p-value for your test.

- You can perform a one-sample  $t$ -test in SPSS by taking the following steps.
  - Choose **Analyze** → **Compare Means** → **One-sample T test**.
  - Move the variable whose mean you want to test to the box labelled **Test Variable(s)**.
  - Type the comparison value into the box labelled **Test Value**.
  - Click the **OK** button.
- The SPSS output will include the following sections.
  - **One-Sample Statistics**. This section provides descriptive statistics about the variable being tested, including the sample size, mean, standard deviation, and standard error for the mean.
  - **One-Sample Test**. This section provides the test statistic, degrees of freedom, and the p-value for a one-sample  $t$  test comparing the mean of the variable to the comparison value. It also provides you with the actual difference between the mean and the comparison value as well as a 95% confidence interval around that difference.
- An example of how to report the results of a  $t$  test would be the following. “The mean height for Alabama boys was significantly different from 60 ( $t[24] = 3.45$ ,  $p = .002$  using a two-tailed test).”
- While the one-sample  $t$ -test is the most commonly used statistic when you want to compare the population mean to a constant value, you can also use a *one-sample Z-test* if you know the population standard deviation and do not need to estimate it based on the sample data. For example, if we knew that the standard deviation of the heights of all 5th grade boys was 3, then we could use a one-sample  $Z$ -test to compare the average height of Alabama boys to 60 inches. The one-sample  $Z$ -test is more powerful than the one-sample  $t$ -test (meaning that it is more likely to detect differences between the estimated mean and the comparison value), but is not often used because people rarely have access to the population standard deviation. A one-sample  $Z$ -test is performed as a hypothesis test with the following characteristics.

- $H_0 : \mu = X_0$   
 $H_a : \mu \neq X_0$

Notice that these are the same hypotheses examined by a one-sample t-test. Just like the one-sample t-test, you can also test a one-tailed hypothesis if you change  $H_a$  to be either  $\mu < X_0$  or  $\mu > X_0$ .

- The test statistic is

$$Z = \frac{\bar{X} - X_0}{\frac{\sigma_X}{\sqrt{n}}}, \quad (2.5)$$

where  $\sigma_X$  is the population standard deviation of the variable you are measuring.

- The p-value for the test statistic  $Z$  can be taken from the standard normal distribution. Notice that there are no degrees of freedom associated with this test statistic. A table containing p-values for the  $Z$  statistic can be found inside the front cover of Moore & McCabe (2003).
- This test assumes that  $X$  follows a normal distribution.

An example of how to report the results of a  $Z$  test would be the following: “The mean height for Alabama boys was significantly different from 60 ( $Z = 2.6$ ,  $p = .009$  using a two-tailed test).”

- You can use the following Excel formula to get the p-value for a  $Z$  statistic.

`=1-normsdist(abs(STAT))`

In this formula you should substitute the value of your  $Z$  statistic for `STAT`. This will provide the p-value for a one-tailed test. If you want the p-value for a two-tailed test, you would multiply the result of this formula by two.

## 2.2 Within-subjects t test

- You can use the one-sample t test described above to compare the mean of any variable measured on your subjects to a constant. As long as you have a value on the variable for each of your subjects, you can use the one-sample t test to compare the mean of that variable to a constant.
- Let us say that you measure one variable on a subject under two different conditions. For example, you might measure patients’ fear of bats before and after they undergo treatment. The rating of fear before treatment would be stored as one variable in your data set while the rating of fear after treatment would be stored as a different variable.

You can imagine creating a new variable that is equal to the difference between the two fear measurements. For every subject you could calculate the value of this new variable using the formula

$$\text{difference} = (\text{fear before treatment}) - (\text{fear after treatment}).$$

Once you create this variable you can use a one-sample t test to see if the mean difference score is significantly different from a constant.

- You can use this logic to test whether the mean fear before treatment is significantly different from the mean fear after treatment. If the two means are equal, we would expect the difference between them to be zero. Therefore, we can test whether the two means are significantly different from each other by testing whether the mean of the difference variable is significantly different from zero.
- In addition to thinking about this as comparing the mean of two variables, you can also think of it as looking for a relation between a categorical IV representing whatever it is that distinguishes your two measurements and a continuous DV representing the thing you are measuring at both times. In our example, the IV would be the time at which the measurement was taken (before or after treatment) and the DV would be level of fear a subject has. If the means of the two groups are different, then we

would claim that the time the measurement was taken has an effect on the level of fear. If the means of the groups aren't different, then we would claim that time has no influence on the reported amount of fear.

- This procedure is called a *within-subjects t test* because the variable that distinguishes our two groups varies within subjects. That is, each subject is measured at both of the levels of the categorical IV. You would want to use a within-subjects t test anytime you wanted to compare the means of two groups and the same subjects were measured in both groups.
- A within-subjects t-test is performed as a hypothesis test with the following characteristics.
  - $H_0 : \mu_1 = \mu_2$   
 $H_a : \mu_1 \neq \mu_2$   
 You can also test a one-tailed hypothesis by changing  $H_a$  to be either  $\mu < \mu_2$  or  $\mu > \mu_2$ .
  - The test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\frac{s_d}{\sqrt{n}}}, \quad (2.6)$$

where  $\bar{x}_1$  is the mean of the first measurement,  $\bar{x}_2$  is the mean of the second measurement,  $s_d$  is the standard deviation of the difference score, and  $n$  is the sample size.

- The p-value for the test statistic  $t$  can be taken from the  $t$  distribution with  $n - 1$  degrees of freedom.
- The within-subjects  $t$  test assumes that your observations are independent except for the group effect and the effect of individual differences, that both of your variables follow a normal distribution, and that their variances are the same.
- You can perform a within-subjects  $t$  test in SPSS by taking the following steps.
  - The two measurements you want to compare should be represented as two different variables in your data set.
  - Choose **Analyze** → **Compare Means** → **Paired-Samples T Test**.
  - Click the variable representing the first measurement in the box on the left.
  - Click the variable representing the second measurement in the box on the left.
  - Click the arrow button in the middle of the box to move this pair of variables over to the right.
  - Click the OK button.
- The SPSS output will contain the following sections.
  - *Paired Samples Statistics*. This section contains descriptive statistics about the variables in your test, including the mean, sample size, standard deviation, and the standard error of the mean.
  - *Paired Samples Correlations*. This section reports the correlation between the two measurements.
  - *Paired Samples Test*. This section provides a within-subjects  $t$  test comparing the means of the two variables. It specifically reports the estimated difference between the two variables, the standard deviation and the standard error of the difference score variable, a confidence interval around the mean difference, as well as the  $t$  statistic, degrees of freedom, and two-tailed  $p$ -value testing whether the two means are equal to each other.
- An example of how to report the results of a within-subjects  $t$  test would be the following: “The ratings of the portrait were significantly different from the ratings of the landscape ( $t[41] = 2.1, p = .04$ ).” If you wanted to report a one-tailed test you would say either “significantly greater than” or “significantly less than” instead of “significantly different from.”

- If you know the standard deviation of the difference score in the population, you should perform a *within-subjects Z test*. A within-subjects Z test is performed as a hypothesis test with the following characteristics.
  - $H_0 : \mu_1 = \mu_2$   
 $H_a : \mu_1 \neq \mu_2$   
You can also test a one-tailed hypothesis if you change  $H_a$  to either  $\mu_1 < \mu_2$  or  $\mu_1 > \mu_2$ .
  - The test statistic is

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\frac{\sigma_d}{\sqrt{n}}}, \quad (2.7)$$

where  $\sigma_d$  is the population standard deviation of the difference score.

- The p-value for the test statistic Z can be taken from the standard normal distribution.
- This test assumes that both  $X_1$  and  $X_2$  follow normal distributions, and that the variances of these two variables are equal.

The within-subjects Z test is more powerful than the within-subjects t test, but is not commonly used because we rarely know the population standard deviation of the difference score.

## 2.3 Wilcoxon signed-rank test

- In the discussion of the within-subjects t test we mentioned the assumptions that these tests make about the distribution of your variables. If these assumptions are violated then you shouldn't use the within-subjects t test. As an alternative you can use the Wilcoxon signed-rank test, which is a *nonparametric test*. Nonparametric tests don't estimate your distribution using specific parameters (like the mean and standard deviation), and typically have less stringent assumptions compared to parametric tests. The Wilcoxon signed-rank test specifically does not assume that your variables have normal distributions, nor does it assume that the two variables have the same variance.
- The Wilcoxon signed-rank test determines whether the medians of two variables are the same. This is not exactly the same as the within-subjects t test, which determines whether the mean of two variables are equal. More broadly, however, both tests tell you whether the central tendency of two variables are the same. Most of the time the Wilcoxon signed-rank test can therefore be used as a substitute for the within-subjects t test. This procedure is particularly appropriate when the DV is originally measured as a set of ordered ranks.
- The only assumption that the Wilcoxon signed-rank test makes is that the distribution of the difference scores is symmetric about the median. It neither requires that the original variables have a normal distribution, nor that the variances of the original variables are the same.
- To perform a Wilcoxon signed-rank test by hand you would take the following steps.
  - The two measurements you want to compare should be represented as two different variables in your data set. In this example we will use the names **VAR1** and **VAR2** to refer to these variables.
  - Compute the difference between the two variables for each subject. In this example we will assume you compute the difference as **VAR1 - VAR2**.
  - Remove any subjects that have a difference score of 0 from the data set for the remainder of the analysis.
  - Determine the absolute value of each of the difference scores.
  - Rank the subjects based on the absolute values of the difference score so that the subject with the lowest score has a value of 1 and subjects with higher scores have higher ranks. If a set of subjects have the same score then you should give them all the average of the tied ranks. So, if the fourth, fifth, and sixth highest scores all were the same, all three of those subjects would be given a rank of five.

- Compute the sum of the ranks for all of the subjects that had positive difference scores (which we will call  $\sum R_+$ ) and the sum of the ranks for all of the subjects had had negative difference scores (which we will call  $\sum R_-$ ).
  - If  $\sum R_+$  is greater than  $\sum R_-$ , the median of **VAR1** is greater than the median of **VAR2**. If  $\sum R_+$  is less than  $\sum R_-$ , the median of **VAR1** is less than the median of **VAR2**. If  $\sum R_+$  is equal to  $\sum R_-$ , the two medians are equal.
  - If we want to know whether the medians are significantly different from each other we can compute the test statistic  $T$  and determine its p-value.  $T$  is equal to the lesser of  $\sum R_+$  and  $\sum R_-$ . This statistic is completely unrelated to the  $t$  distribution that we use for the within-subjects  $t$  test.
  - $T$  has a value of  $\frac{n(n+1)}{4}$  when there is no difference between the two variables. Smaller values of  $T$  indicate a greater difference between the two medians, and are correspondingly associated with smaller p-values. This  $T$  statistic is not the same as the one you obtain from a standard  $T$ -test, and so you must use a special table to determine its p-values. One such table can be found in Sheskin (2004) on page 1138. The  $n$  used in the table refers to the number of cases that have a nonzero difference score. This means that your p-values are unaffected by cases that have no differences between your groups, no matter how many of them you have.
- To perform a Wilcoxon signed-rank test in SPSS you would take the following steps.
    - The two measurements you want to compare should be represented as two different variables in your data set.
    - Choose **Analyze** → **Nonparametric Tests** → **2 Related Samples**.
    - Click the variable representing the first measurement in the box on the left.
    - Click the variable representing the second measurement in the box on the left.
    - Click the arrow button in the middle of the box to move this pair of variables over to the right.
    - Make sure the box next to **Wilcoxon** is checked.
    - Click the OK button.
  - The SPSS output will contain the following sections.
    - *Ranks*. The first part of this section reports how many subjects had positive, negative, and tied (zero) difference scores. The second tells you the mean ranks for subjects with positive and negative difference scores. The third tells you the sum of the ranks for subjects with positive and negative difference scores. The values in the last column correspond to  $\sum R_+$  and  $\sum R_-$ .
    - *Test Statistics*. This section reports a  $Z$  statistic and p-value testing whether the medians of the two variables are equal. This uses a different method to compute the p-values than the one described above but it is also valid.
  - An example of how to report the results of a Wilcoxon signed-rank test would be the following: “Music students ranked composers born in the 17th century as more exceptional than those born in the 18th century ( $Z = 2.4$ ,  $p = .01$  using the Wilcoxon signed-rank test).” We need to explicitly say what test we are using because people aren’t as familiar with this test as they are with the standard within-subjects  $t$  test.

## 2.4 Between-subjects $t$ test

- The within-subjects  $t$  test is very handy if you want to compare the means of two groups when the same people are measured in both of the groups. But what should you do when you want to compare two groups where you have different people in each group? For example, you might want to look at differences between men and women, differences between art and biology majors, or differences between those taught grammar using textbook A and those taught using textbook B. None of these comparisons can be performed using the same people in both groups, either because they involve mutually exclusive categories or else they involve manipulations that permanently change the subjects.

- You can use a *between-subjects t test* to compare the means of two groups on a variable when you have different people in your two groups. For example, you could test whether the average height of people from Australia is different from the average height of people from New Zealand, or whether people from rural areas like country music more than people from urban areas.
- The standard between-subjects t test can be performed as a hypothesis test with the following characteristics.
  - $H_0 : \mu_1 = \mu_2$   
 $H_a : \mu_1 \neq \mu_2$   
 You can also test a one-tailed hypothesis by changing  $H_a$  to either  $\mu_1 < \mu_2$  or  $\mu_1 > \mu_2$ .
  - The test statistic is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad (2.8)$$

where  $\bar{X}_1$  and  $\bar{X}_2$  are the means of your two groups,  $s_1^2$  and  $s_2^2$  are the variances of your two groups, and  $n_1$  and  $n_2$  are the sample sizes of your two groups.

- The p-value for this t statistic is unfortunately difficult to determine. It actually does not follow the t distribution, but instead follows something called the  $t(k)$  distribution. However, if the sample size within each group is at least 5, the statistic will approximately follow a t distribution with degrees of freedom computed by the formula

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1} \left(\frac{s_2^2}{n_2}\right)^2}. \quad (2.9)$$

- This test assumes that both  $X_1$  and  $X_2$  follow normal distributions and that the measurements made in one group are independent of the measurements made in the other group.
- Researchers often choose not to use the standard between-subjects t test because it is difficult to calculate the p-value for the test statistics it produces. As an alternative you can test the same hypotheses using a *pooled t test* which produces statistics that have the standard t distribution. The main difference between the between-subjects t test and the pooled t test is that the pooled t test assumes that the variances of the two groups are the same instead of dealing with them individually.
- Before you can perform a pooled t test, you must *pool* the variances of your two groups into a single estimate. You can combine your two individual group variances into a single estimate since you are assuming that your two groups have the same variance. The formula to pool two variances is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \quad (2.10)$$

You can see that this formula takes the weighted average of the two variances, where the variance for each group is weighted by its degrees of freedom. You can also pool variances across more than two groups using the more general formula

$$s_p^2 = \frac{\sum_{j=1}^k (n_j - 1)s_j^2}{\sum_{j=1}^k (n_j - 1)}, \quad (2.11)$$

where  $k$  is the number of groups whose variances you want to pool.

- Once you compute the pooled standard deviation you can perform a pooled t test as a hypothesis test with the following characteristics.

- $H_0 : \mu_1 = \mu_2$   
 $H_a : \mu_1 \neq \mu_2$

You can also test a one-tailed hypothesis by changing  $H_a$  to either  $\mu_1 < \mu_2$  or  $\mu_1 > \mu_2$ .

- The test statistic is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \quad (2.12)$$

- The p-value for the test statistic can be taken from the t distribution with  $n_1 + n_2 - 2$  degrees of freedom.
- The pooled t test assumes that your observations are independent except for group effects, that both of the variables have a normal distribution, that the variance of the variables is the same, and the the measurements made in one group are independent of the measurements made in the other group.
- SPSS will perform both a between-subjects t test and a pooled t test if you take the following steps.
  - Your data set should contain one variable the represents the group each person is in, and one variable that represents their value on the measured variable.
  - Choose **Analyze** → **Compare Means** → **Independent-Samples T Test**.
  - Move the variable representing the measured response to the **Test Variable(s)** box.
  - Move the variable representing the group to the **Grouping Variable** box.
  - Click the **Define Groups** button.
  - Enter the values for the two groups in the boxes labelled **Group 1** and **Group 2**.
  - Click the **OK** button.
- The SPSS output will contain the following sections.
  - *Group Statistics*. This section provides descriptive information about the distribution of the DV within the two groups, including the sample size, mean, standard deviation, and the standard error of the mean.
  - *Independent Samples Test*. This section provides the results of two t-tests comparing the means of your two groups. The first row reports the results of a pooled t test comparing your group means, while the second row reports the results of between-subjects t test comparing your means. The columns labelled **Levene's Test for Equality of Variances** report an F test comparing the variances of your two groups. If the F test is significant then you should use the test in the second row. If it is not significant then you should use the test in the first row. The last two columns provide the upper and lower bounds for a 95% confidence interval around the difference between your two groups.
- An example of how to report the results of a between-subjects t test would be the following: "Participants with a cognitive load were able to read fewer pages of the document compared to participants without a cognitive load ( $t[59] = 3.02, p = .002$ )."

## 2.5 Wilcoxon rank-sum test/Mann-Whitney U test

- Both the Wilcoxon rank-sum test and the Mann-Whitney U test are nonparametric equivalents of the between-subjects t test. These tests allow you to determine if the median of a variable for participants in one group is significantly different from the median of that variable for participants in a different group.

- These two nonparametric tests assume that the distribution of the response variable in your two groups is approximately the same, but they don't require that the distribution have any particular shape. However, the requirement that the distributions are the same implies that the variance must be the same in the two groups.
- Even though these two tests involve different procedures, they are mathematically equivalent so they produce the same p-values. Moore & McCabe (2003) explain how to perform the Wilcoxon rank-sum test in section 14.1. To perform a Mann-Whitney U test you would take the following steps.
  - Combining the data from both of the groups, rank the subjects based on the values of the response measure so that the subject with the lowest score has a value of 1 and the subject with the highest score has a value of  $n_1 + n_2$ . If a set of subjects have the same score then you should give them all the average of the tied ranks.
  - Compute the statistics

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \sum R_1 \quad (2.13)$$

and

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \sum R_2, \quad (2.14)$$

where  $\sum R_1$  is the sum of the ranks for group 1 and  $\sum R_2$  is the sum of the ranks for group 2.

- The test statistic  $U$  is equal to the smaller of  $U_1$  and  $U_2$ . Smaller values of  $U$  indicate a greater difference between the two medians. P-values for this test statistic can be found in Sheskin (2004) on pages 1151-1152.
- You take the following steps to perform a Wilcoxon rank-sum test/Mann-Whitney U test in SPSS.
  - Choose **Analyze** → **Nonparametric Tests** → **2 Independent Samples**.
  - Move the variable representing the measured response to the **Test Variable(s)** box.
  - Move the variable representing the group to the **Grouping Variable** box.
  - Click the **Define Groups** button.
  - Enter the values for the two groups in the boxes labelled **Group 1** and **Group 2**.
  - Click the **OK** button.
- The SPSS output will contain the following sections.
  - *Ranks*. This section reports the sample size, mean, and sum of the ranks for each of your groups.
  - *Test Statistics*. This section reports the test statistics for both the Mann-Whitney U and the Wilcoxon Rank-Sum test determining whether the median of one group is significantly different from the median of the other group. It also reports a Z statistic representing the normal approximation for these tests along with the two-tailed p-value for that Z.
- An example of how to report the results of a Mann-Whitney U test would be the following: "Across all American colleges, the median number of graduates for Art History is significantly different from the median number of graduates for Performing Art ( $Z = 4.9$ ,  $p < .001$  using the Mann-Whitney U test)." Just like for the Wilcoxon Signed-Rank test, we need to explicitly say what test we are using because many people aren't familiar with the Mann-Whitney U test.



## Chapter 3

# Testing a Single Between-Subjects Factor

### 3.1 Introduction

- In Section 2.4 we learned how we can use a between-subjects t test to determine whether the mean value on a DV for one group of subjects is significantly different from the mean value on that DV for a different group of subjects. In this chapter we will learn ways to compare the average value on a DV across any number of groups when you have different subjects in each group.
- This chapter will only consider how to compare groups in a between-subjects design. We will consider how to compare groups in a within-subjects design in Chapter 5.
- Another way to think about the tests in this section is that they examine the relation between a categorical IV and a continuous DV when you have different people in each level of your IV.

### 3.2 One-way between-subjects ANOVA

- The most common way to determine whether there are differences in the means of a continuous DV across a set of three or more groups is to perform an *analysis of variance* (ANOVA).
- There are many different types of ANOVAs. The specific type you should use depends on two things.
  - The first is whether or not your groups have the same subjects in them. If the groups have different subjects in them, then you must perform a *between-subjects ANOVA*. If the same subjects are in all of your groups then you must perform a *within-subjects ANOVA*. You should try to make sure that your study fits one of these two conditions if you want to perform a one-way ANOVA. It is difficult to use ANOVA to analyze designs when you have the same subjects in multiple groups but you don't have every subject in every group. For example, it would be difficult to analyze a study where you had four different conditions but only measured each subject in two of them.
  - The second thing that affects the type of ANOVA you will perform is the relation between the different groups in the study and the theoretical constructs you want to consider. In our case we will assume that the groups represent different levels of a single theoretical variable. This is called a *one-way ANOVA*. It is also possible that your groups might represent specific combinations of two or more theoretical variables. These analyses would be called *Multifactor ANOVAs*.

In this chapter we will discuss how to perform a one-way between-subjects ANOVA. This means that the groups represent different levels of a single theoretical variable, and that we have different subjects in each of our groups.

- The purpose of a one-way between-subjects ANOVA is to tell you if there are any differences among the means of two or more groups. If the ANOVA test is significant it indicates that at least two of the groups have means that are significantly different from each other. However, it does not tell you which of the groups are different.
- The IV in an ANOVA is referred to as a *factor*, and the different groups composing the IV are referred to as the *levels* of the factor.
- ANOVA compares the variability that you have within your groups to the variability that you have between your groups to determine if there are any differences among the means. While this may not be straightforward, it actually does do the job. We measure the variability between our groups by comparing the observed group means to the mean of all of the groups put together. As the differences between our groups get bigger we can expect that the average difference between the mean of the groups and the overall mean will also get bigger. So the between-group variability is actually a direct measure of how different our means are from each other. However, even if the true population means were equal, we would expect the observed means of our groups to differ somewhat due to random variability in our samples. The only thing that can cause cases in the same group to have different values on the DV is random variability, so we can estimate the amount of random variability we have by looking at the variability within each of our groups. Once we have that, we can see how big the differences are between the groups compared to the amount of random variability we have in our measurements to determine how likely it is that we would obtain differences as large as we observed due to random chance alone.
- A one-way between-subjects ANOVA is performed as a hypothesis test with the following characteristics.
  - $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$   
 $H_a$  : At least two of the means are different from each other  
 ANOVA doesn't consider the direction of the differences between the group means, so there is no equivalent to a one-tailed test.
  - The test statistic for a one-way between-subjects ANOVA is best calculated in several steps.
    1. Compute the *sum of squares between groups* using the formula

$$SS_{\text{bg}} = \sum_{i=1}^k (\bar{Y}_i - \bar{Y}_{..})^2, \quad (3.1)$$

where  $k$  is the total number of groups,  $\bar{Y}_i$  is the mean value of the DV for group  $i$ , and  $\bar{Y}_{..}$  is the mean of all of the observations on the DV, also called the *grand mean*. The sum of squares between groups is a measure of how different your means are from each other. The  $SS_{\text{bg}}$  reflects both variability caused by differences between the groups as well random variation in your observations. It is also sometimes referred to as the *sum of squares for groups*.

Throughout these notes we will use the “dot notation” when discussing group means. The response of each subject can be represented by the symbol  $Y_{ij}$ : the value of  $Y$  for the  $j$ th person in group  $i$ . If we want to average over all of the subjects in a group, we can replace  $j$  with a period and use the term  $\bar{Y}_{i.}$ . If we want to average over all the subjects in all of the groups, we can replace both  $i$  and  $j$  with periods and use the term  $\bar{Y}_{..}$ .

2. Compute the *sum of squares within groups* using the formula

$$SS_{\text{wg}} = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2. \quad (3.2)$$

The sum of squares within groups purely represents the amount of random variation you have in your observations. It is also sometimes called the *sum of squared error*.

3. Compute the *mean squares between groups* and the *mean squares within groups* by dividing each of the sum of squares by their degrees of freedom. The sum of squares between groups has  $k - 1$  degrees of freedom and the sum of squares within groups has  $N - k$  degrees of freedom, where  $N$  is the total sample size. This means that you can compute the mean squares using the formulas

$$MS_{bg} = \frac{SS_{bg}}{k - 1} \quad (3.3)$$

and

$$MS_{wg} = \frac{SS_{wg}}{N - k}. \quad (3.4)$$

The mean square between groups is also called the *mean square for groups*, and the mean square within groups is also called the *mean square error*. The mean square within groups is always equal to the variance pooled across all of your groups.

4. Compute the test statistic

$$F = \frac{MS_{bg}}{MS_{wg}}. \quad (3.5)$$

This statistic essentially tells us how big the differences are between our groups relative to how much random variability we have in our observations. Larger values of  $F$  mean that the difference between our means is large compared to the random variability in our measurements, so it is less likely that we got our observed results due to chance alone.

- The p-value for this test statistic can be taken from the  $F$  distribution with  $k - 1$  numerator degrees of freedom and  $N - k$  denominator degrees of freedom. You'll notice that this corresponds to the degrees of freedom associated with the  $MS_{bg}$  and the  $MS_{wg}$ , respectively.
  - This test assumes that your observations are independent except for the effect of the IV, that your DV has a normal distribution within each of your groups, and that the variance of the DV is the same in each of your groups.
- The results of an ANOVA are often summarized in an *ANOVA table* like Figure 3.1.

Figure 3.1: An example ANOVA table

Source	SS	df	MS	Test
Group	$SS_{bg}$	$k - 1$	$MS_{bg} = \frac{SS_{bg}}{k-1}$	$F = \frac{MS_{bg}}{MS_{wg}}$
Error	$SS_{wg}$	$N - k$	$MS_{wg} = \frac{SS_{wg}}{N-k}$	
Total	$SS_{total} = SS_{bg} + SS_{wg}$	$N - 1$		

Looking at Figure 3.1, you can see that the total SS (representing the total variability in your DV) is equal to the SS for groups plus the SS for error. Similarly, the total df is equal to the df for groups plus the df for error. You can therefore think about ANOVA as dividing the variance in your DV into the part that can be explained by differences in your group means and the part that is unrelated to your groups.

- While ANOVA will tell you whether there are any differences among your group means, you usually have to perform follow-up analyses to determine exactly which groups are different from each other. We will discuss this more in section 3.3.

Figure 3.2: Comparison of one-way between-subjects ANOVA and the pooled t test.

Test	Statistic	Numerator	Denominator
Pooled t test	$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	Difference between means	Pooled standard deviation
ANOVA	$F = \frac{MS_{bg}}{MS_{wg}}$	Variance due to group differences	Variance pooled across groups

- There are several ways a one-way between-subjects ANOVA is similar to a pooled t test. Looking at the formulas for their test statistics reveals several parallels. These are summarized in Figure 3.2.

Both the pooled t test and ANOVA are based on the ratio of an observed difference among the means of your groups divided by an estimate of group variability. In fact, if you perform an ANOVA using a categorical IV with two groups, the F statistic you get will be equal to the square of the t statistic you'd get if you performed a pooled t test comparing the means of the two groups. In this case the p-values of the F and the t statistics would be the same.

In both the pooled t test and in ANOVA, larger ratios are associated with smaller p-values. This makes perfect sense if we think about the factors that would affect our confidence in the results. If we observe larger differences between our means we can be more sure that they are due to real group differences instead of random chance. Larger group differences are therefore associated with larger test statistics. Conversely, if we observe more random variability in our data, we are going to be less sure in our results since there are more external factors that could happen to line up to give us group differences. Greater random variability is therefore associated with smaller test statistics.

- To perform a one-way between-subjects ANOVA in SPSS you would take the following steps.
  - Choose **Analyze** → **General Linear Model** → **Univariate**.
  - Move the DV to the **Dependent Variable** box.
  - Move the IV to the **Fixed Factor(s)** box.
  - Click the **OK** button.
- The SPSS output will contain the following sections.
  - *Between-Subjects Factors*. Lists how many subjects are in each level of your factor.
  - *Tests of Between-Subjects Effects*. The row next to the name of your factor reports a test of whether there is a significant relation between your IV and the DV. A significant F statistic means that at least two group means are different from each other, indicating the presence of a relation.

You can ask SPSS to provide you with the means within each level of your between-subjects factor by clicking the **Options** button in the variable selection window and moving your within-subjects variable to the **Display Means For** box. This will add a section to your output titled **Estimated Marginal Means** containing a table with a row for each level of your factor. The values within each row provide the mean, standard error of the mean, and the boundaries for a 95% confidence interval around the mean for observations within that cell.

### 3.3 Specific group comparisons

- After determining that there is a significant effect of your categorical IV on the DV using ANOVA, you must compare the means of your groups to determine which groups differ from each other. If your IV only has two levels then all you need to do is see which group has the higher mean. However, if

you have three or more groups you must use more complicated methods to determine the nature of the effect. You can do this using preplanned contrasts, post-hoc tests, or by nonstatistical methods.

### 3.3.1 Preplanned contrasts

- Preplanned contrasts allow you to test the equivalence of specific sets of means. In the simplest case you can test to see if one mean is significantly different from another mean. However, you can also use contrasts to determine if the average of one set of groups is significantly different from the average of another set of groups.
- By definition, you should determine what preplanned contrasts you want to examine before you look at your results. You should use post-hoc comparisons if you want to confirm the significance of a pattern you observe in the data.
- A preplanned contrast comparing the means of two groups will typically be more powerful than a between-subjects t test examining the same difference. This is because the contrast takes advantage of the assumption you make in ANOVA that the variance of the DV is the same in all of your groups. The contrast uses the pooled variance across all of your groups (not just the two involved in the comparison) to estimate the standard error for the comparison. This means that the degrees of freedom for a preplanned contrast will usually be substantially higher than the degrees of freedom you would have for a standard between-subjects t test comparing the means of the same groups.
- The first thing you need to do to perform a contrast is to write down what comparison you want to make. You do this by defining your contrast  $L$  as being the sum of the means of each of your groups multiplied by a *contrast weight*. The contrast weight for group  $j$  is referred to as  $c_j$ . You can define your weights however you want, but their sum must be equal to 0. You should also make sure that the contrast itself has theoretical meaning.

As an example, let us say that you conducted an ANOVA predicting how much popcorn that people ate while watching four different types of movies: action (group 1), comedy (group 2), horror (group 3), and romance (group 4). One valid contrast would use the contrast weights (1 -1 0 0). The value of this contrast would be equal to the mean popcorn eaten by people watching action movies minus the mean popcorn eaten by people watching comedy movies. A test of this contrast would tell you whether people eat more popcorn when they watch action movies or comedy movies.

- Most commonly your contrast will be equal to the mean of one set of groups minus the mean of another set of groups. It is not necessary that the two sets have the same number of groups, nor is it necessary that your contrast involve all of the groups in your IV. You should choose one of your sets to be the “positive set” and one to be the “negative set.” This is an arbitrary decision and only affects the sign of the resulting contrast. If the estimated value of the contrast is positive, then the mean of the positive set is greater than the mean of the negative set. If the estimated value of the contrast is negative, then the mean of the negative set is greater than the mean of the positive set.

Once you define your sets you can determine the values of your contrast weights in the following way.

- For groups in the positive set, the value of the contrast weight will be equal to  $\frac{1}{g_p}$ , where  $g_p$  is the number of groups in the positive set.
- For groups in the negative set, the value of the contrast weight will be equal to  $-\frac{1}{g_n}$ , where  $g_n$  is the number of groups in the negative set.
- For groups that are not in either set, the value of the contrast weight will be equal to 0.

Continuing our example above, let us say that you wanted to test whether the amount of popcorn eaten while watching horror films (positive set) is significantly different from the amount of popcorn eaten while watching action, comedy, or romance films (negative set). In this case our contrast weight vector would be equal to  $(-\frac{1}{3} -\frac{1}{3} 1 -\frac{1}{3})$ .

- After you determine your contrast weights, you are ready to test whether your contrast is significantly different from zero. You can do this as a hypothesis test with the following characteristics.

- $H_0 : L = 0$   
 $H_a : L \neq 0$
- To compute the test statistic you must first determine the value of your contrast using the formula

$$L = \sum_{j=1}^k c_j \bar{X}_j, \quad (3.6)$$

where  $k$  is the number of groups. You then must determine the standard error of your contrast using the formula

$$s_L = \sqrt{\text{MSE} \sum_{j=1}^k \left( \frac{c_j^2}{n_j} \right)}, \quad (3.7)$$

where MSE is the mean squared error from the ANOVA model, and  $n_j$  is the sample size for group  $j$ . Finally, you can compute your test statistic using the formula

$$t = \frac{L}{s_L}. \quad (3.8)$$

- The p-value for this test statistic can be taken from the  $t$  distribution with the same degrees of freedom as the MSE in the overall ANOVA. In a one-way between-subjects ANOVA this will be equal to  $N - k$ , where  $N$  is the total sample size.
- Just like ANOVA, the test for a contrast assumes that your DV is normally distributed within each of your groups and that the variance is approximately equal across all of the groups.
- You can test multiple preplanned contrasts based on the same ANOVA, although you shouldn't try to test too many at once. Testing multiple contrasts will increase the probability of you getting one significant just due to chance. See section 3.3.2 for more information about this.
- To test preplanned contrasts in SPSS you would take the following steps.
  - Contrasts are performed within the context for an ANOVA, so you must first perform all the steps for an ANOVA as described in section 3.2. However, stop before clicking the **OK** button.
  - Click the **Paste** button. SPSS will copy the syntax you need to run the one-way between-subjects ANOVA you defined into a new window. Pre-planned contrasts cannot be performed using interactive mode, so you will need to edit the syntax to get the results you want.
  - Insert a line after the **/CRITERIA = ALPHA(.05)** statement.
  - On the blank line, type **/LMATRIX**, followed first by the name you want to give your contrast in quotes, then by the name of your IV, and finally by a list of your contrast weights. The name of your contrast can be up to 255 characters long. The individual weights in the vector should be separated by spaces, and can be written as whole numbers, decimals, or fractions. You must list the weights in alphabetical order, based on the values used to represent the groups. So, for example, if your categorical IV has the values "White," "Black," and "Asian," you must list the contrast weight for Asians first, followed by the weight for Blacks, and then followed by the weight for Whites. If your IV uses numbers to distinguish your groups, you must list the weights in ascending order based on their level number.
  - Select the entire block of syntax.
  - Choose **Run** → **Selection**. You can alternatively click the arrow button beneath the main menu.
- The code you run should look something like the following.

## UNIANOVA

```
shine BY brand
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/LMATRIX = "2 and 3 versus 1, 4, and 5" brand -1/3 1/2 1/2 -1/3 -1/3
/DESIGN = brand .
```

- Including the statement to run your contrast will add the following sections to your SPSS output.
  - *Contrast Results (K Matrix)*. This section provides you with descriptive statistics on your contrast. Since your contrast represents a specific mathematical combination of the group means, you can actually compute the result of that combination.
  - *Test Results*. This section tests whether your contrast is significantly different from zero. This is the same thing as testing whether your two sets are equal. Notice that the MSE for your contrast is the same as the MSE from the ANOVA model.
- SPSS will allow you to test multiple contrasts from the same ANOVA model. If you want to examine each of the contrasts independently, all you need to do is include multiple **/LMATRIX** commands in your syntax. For example, the following syntax would compare group 1 to group 2 and then independently compare group 4 to group 5.

## UNIANOVA

```
strain BY machine
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/LMATRIX = "1 versus 2" machine 1 -1 0 0 0
/LMATRIX = "4 versus 5" machine 0 0 0 1 -1
/DESIGN = machine .
```

- You also have the option of jointly testing a set of contrasts. In this case the resulting test statistic would tell you whether any of the contrasts are significant, in the same way that ANOVA tests whether there are any significant differences among a group of means. To do this in SPSS you would first go through all of the steps listed above to test the first contrast in your set. You would then add a semicolon (;) to the end of your **/LMATRIX** statement, list your IV, and then write out the contrast weight vector for your second contrast. You would continue doing this until you listed out all of the contrasts in your set. As an example, the code below would jointly test whether group 1 is different from 2, whether 2 is different from 3, and whether 3 is different from 4.

## UNIANOVA

```
strain BY machine
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/LMATRIX = "Joint contrast" machine 1 -1 0 0 0;
machine 0 1 -1 0 0; machine 0 0 1 -1 0
/DESIGN = machine .
```

In this case your resulting test statistic would be significant if any of the contrasts were significantly different from zero. However, note that the results that you get from the joint test statistic will not be exactly the same as performing each of the contrasts independently and then checking to see if any of the individual test statistics are significant. This is because the joint test takes into account how many different contrasts you're examining while the independent contrasts do not.

You should also note that you cannot include completely redundant contrasts in your joint test. Each contrast must be testing something different than the other contrasts in the set. One implication of this is that you can only jointly test a total of  $k - 1$  contrasts for a categorical IV with  $k$  groups. It turns out that any time you write out  $k$  contrasts based on an IV with  $k$  groups, there is always some way to write an equation that would give you the result of one of your contrasts as a mathematical function of your other contrasts.

### 3.3.2 Post-hoc comparisons

- Sometimes there isn't a specific comparison that you know you want to make before you conduct your study. In this case what you would do is perform your ANOVA first and then look at the pattern of means afterwards to determine the nature of the effect. This is referred to as performing *post-hoc comparisons*.
- When you wait until after you look at the results before deciding what groups you want to compare, you must be concerned with *inflating your chance of type I error*. A type I error is when you claim that there is a significant difference between two groups when they really are the same. The probability of you making a type I error is set by the significance level  $\alpha$  that you choose. If you are willing to say that there is a significant difference between two groups whenever the probability of getting differences as large as you got is 5% or less, then you can expect that 5% of the time you are going to say that two groups are significantly different when they are not.

When deciding which groups to test after looking at the pattern of means, you will naturally want to test those that look like they have the biggest differences. When you do that, it is conceptually the same thing as testing the differences between all of the means and then only looking at the ones that turned out to be significant. The problem with this is that when you perform a bunch of tests, the probability that at least one of the tests is significant due to random chance alone is going to be bigger than  $\alpha$ . In fact, if you consider all the independent tests that you would make when comparing the means of the different levels of a categorical IV, the actual probability of making a type 1 error if you just compared the means to each other would be the following.

- For 2 groups: .05 (there is only one comparison to be made between two groups, so our probability isn't inflated)
- For 3 groups: .0975
- For 4 groups: .1426
- For 5 groups: .1855

In general, the probability of making at least one type I error when you perform multiple independent tests is

$$p = (1 - .95^m), \quad (3.9)$$

where  $m$  is equal to the number of tests you perform.

- The probability of having at least one type I error among your entire set of tests is called the *familywise error rate*. When performing post-hoc comparisons you need to make sure that the familywise error rate is less than your  $\alpha$ , not just the p-values for the individual comparisons.
- The simplest way to control the familywise error rate is to divide your  $\alpha$  by the number of tests you are performing. This is known as applying a *Bonferroni correction* to your  $\alpha$ . This method is also the most conservative approach, meaning that it is the one that is least likely to claim that an observed difference is significant. A Bonferroni correction can be applied any time you perform multiple tests - not just when performing post-hoc comparisons in ANOVA.



- Sidak (1967) provides a variant of the Bonferroni test that doesn't reduce your  $\alpha$  as much. In the Sidak method you use the value

$$\alpha_c = 1 - (1 - \alpha)^{\frac{1}{j}} \quad (3.10)$$

as the significance level for your individual comparisons, where  $j$  is the number of different tests you are making. Just like Bonferroni, the Sidak method guarantees that the overall experimental error rate stays below your established  $\alpha$ .

- There are several other post-hoc tests that are specifically designed to compare the group means in ANOVA. Although the details of how they work vary, all of these tests compare each group mean to all of the other group means in a way that keeps the familywise error rate under your established  $\alpha$ . The most commonly used post-hoc methods, ordered from most conservative to least conservative are

1. Scheffé
2. Tukey
3. SNK (Student-Newman-Keuls)

All of these methods are less conservative than performing a Bonferroni correction to your  $\alpha$ .

- Post-hoc comparisons will typically give you two different types of output. Sometimes a method will give you a series of individual tests comparing the mean of each group to the mean of every other group. These tests will be less powerful than a standard t-test comparing the two group means because the post-hoc tests will adjust for the total number of comparisons you make. Other times you will receive a table of *homogenous subsets* like those presented in Figure 3.3.

Figure 3.3: Examples of homogenous subset tables.

Lightbulb color	n	Mean score
Red	16	2.57 <sub>a</sub>
Orange	14	2.82 <sub>a</sub>
Yellow	16	4.58 <sub>ab</sub>
Green	16	6.02 <sub>b</sub>
Blue	15	5.33 <sub>b</sub>

Lightbulb color	n	Subset 1	Subset 2
Red	16	2.57	
Orange	14	2.82	
Yellow	16	4.58	4.58
Green	16		6.02
Blue	15		5.33

The idea behind homogenous subsets is to divide your groups into sets such that groups that are in the same set are not different from each other, while groups that don't have any sets in common are different from each other. The table on the left in Figure 3.3 uses subscripts to indicate the sets that the groups belong to. The general pattern is that performance is better in groups with lower wavelengths. We can see that there are two sets, indicated by the subscripts **a** and **b**. The red and orange groups are only in set **a**, so the mean scores in both of these are significantly different from green and blue groups, which are only in set **b**. However, the yellow group is in both sets, indicating that no group is significantly different from yellow. The table on the right illustrates the same sets, but it does so by putting groups that are not different from each other in the same column instead of using subscripts.

- Some statisticians have suggested using the F statistic from ANOVA as *omnibus test* to control the familywise error rate as an alternative to performing special post-hoc comparisons. The logic is that once you have obtained a significant result from ANOVA, you are justified in performing whatever analyses are needed to explain the nature of your result. The ANOVA itself takes the number of different groups you are comparing into account when you look up its p-value, so there is no reason to apply additional corrections when comparing the means of specific groups. In this case you would use uncorrected t tests to determine whether any given pair of means are different. Using uncorrected

t tests to look for differences between group means is also called *Fisher's least significant difference method*.

- The idea of an omnibus test providing justification for performing multiple followup tests has also been applied to other areas of statistics. For example, researchers will use a significant overall model test in regression to justify looking at tests of individual parameters, and they will use a significant multivariate test to justify looking at the underlying univariate models (Rencher & Scott, 1990).
- Simulations performed by Carmer & Swanson (1973) have demonstrated that your familywise error rate is right around your established  $\alpha$  if you use uncorrected t tests to explore differences between your groups after obtaining a significant result from ANOVA. However, this requires that you *only* perform the group comparisons after obtaining a significant ANOVA. If you perform the t tests even when the ANOVA is not significant, then your familywise error rate will be substantially greater than your established  $\alpha$ .

The same article showed that other post-hoc methods typically have familywise error rates that are substantially below your  $\alpha$ . While this is fine if you obtain a significant result, it makes it harder to obtain significance.

- To obtain post-hoc analyses in SPSS you can take the following steps.
  - Repeat the steps to perform a one-way ANOVA presented in section 3.2, but do not press the **OK** button at the end.
  - Click the **Post-Hoc** button.
  - Move the IV to the **Post-Hoc Tests for** box.
  - Check the boxes next to the post-hoc tests you want to perform.
  - Click the **Continue** button.
  - Click the **OK** button.
- Requesting a post-hoc test will add one or both of the following sections to your ANOVA output.
  - *Multiple Comparisons*. This section is produced by LSD, Tukey, and Bonferroni tests. It reports the difference between every possible pair of factor levels and tests whether each is significant. It also includes the boundaries for a 95% confidence interval around the size of each difference.
  - *Homogenous Subsets*. This section is produced by SNK and Tukey tests. It reports a number of different subsets of your different factor levels. The mean values for the factor levels within each subset are not significantly different from each other. This means that there is a significant difference between the mean of two factor levels only if they do not appear in any of the same subsets.

### 3.3.3 Nonstatistical methods of comparing means

- Both preplanned contrasts and post-hoc comparisons represent ways to explore the nature of a significant ANOVA using statistical tests. However, it is also possible to interpret an ANOVA without performing any follow-up analyses.
- Although ANOVA is designed to determine whether there are any significant differences among the group means, we have already discussed how it is not the same thing as looking at t tests comparing all of the means. In most ways this is a positive feature, since performing multiple t tests would lead to a familywise error rate that would be larger than your significance level  $\alpha$ . However, it also means that the results from the ANOVA will not always be consistent with the results of post-hoc comparisons. You will typically find that the ANOVA will be significant whenever you have a pair of means that are significantly different from each other. However, it is possible to have a significant ANOVA without having any pairs of means that are significantly different. It is also possible to have a non-significant ANOVA at the same time that a post-hoc test shows that a pair of means is significantly different.

- Whenever the results your post-hoc analyses mismatch the results of the ANOVA, you should base your conclusion on the results of the ANOVA. If the ANOVA is not significant but post-hoc tests say that there is a significant difference between a pair of groups, then you simply ignore the post-hocs and conclude that there are no differences among the group means. If the ANOVA is significant but post-hoc tests fail to show any differences among the group means, however, it can be difficult to explain how the categorical IV relates to the DV.
- Partly because of this issue, some statisticians have proposed that we should interpret significant ANOVAs simply by looking at a table or graph of the group means. They highlight the fact that the post-hoc analyses are not meant to be interpreted directly, but instead are tools used to help us explain a significant ANOVA. If those tools fail to do the job, we should simply find other ways to describe the relation between the IV and the DV. The significant ANOVA tells us that there are real differences among the group means, so we don't actually have to use any other statistical tests to back that up.
- To obtain a graph of the group means in SPSS you would take the following steps.
  - Repeat the steps to perform a one-way ANOVA presented in section 3.2, but do not press the **OK** button at the end.
  - Click the **Plots** button.
  - Move the IV to the **Horizontal Axis** box.
  - Click the **Add** button.
  - Click the **Continue** button.
  - Click the **Ok** button.
- This will add the following section to your ANOVA output.
  - *Profile Plots*. This section will contain a line graph of the group means.
- To obtain a table of the group means in SPSS you would take the following steps.
  - Repeat the steps to perform a one-way ANOVA presented in section 3.2, but do not press the **OK** button at the end.
  - Click the **Options** button.
  - Move the IV to the **Display Means for** box.
  - Click the **Continue** button.
  - Click the **OK** button.
- This will add the following section to your ANOVA output.
  - *Estimated Marginal Means*. This section provides a table containing the mean, standard error, and a 95% confidence interval around the mean of each of your groups.

### 3.4 Kruskal-Wallis test

- Recall that a one-way between-subjects ANOVA makes two assumptions.
  1. The DV is normally distributed within each of your groups.
  2. The variances in your groups is the same.
- ANOVA is robust to violations of these assumptions just like the t test. The p-values it produces will therefore still be reasonably accurate as long as you do not violate the assumptions severely.
- The *Kruskal-Wallis test* is a nonparametric equivalent of the one-way between-subjects ANOVA that does not require that your DV has a normal distribution. However, it does assume that the distribution of the DV has approximately the same shape in each of your groups, which implies that the variance is approximately equal across your groups.

- The Kruskal-Wallis test will tell you if there are any significant differences among the medians of two or more groups. It is an extension of the Mann-Whitney U test, and will give you the same results as the Mann-Whitney U test if you just compare two groups.
- To perform a Kruskal-Wallis test by hand you would take the following steps.
  - Combine the data from all of the groups and rank the subjects based on the values of the response measure. The subject with the lowest score has a value of 1 and the subject with the highest score has a value of  $n_1 + n_2$ . If a set of subjects have the same score then you should give them all the average of the tied ranks.
  - Compute the sum of the ranks for each of your groups.
  - Compute the test statistic

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \left( \frac{R_j^2}{n_j} \right) - 3(N+1), \quad (3.11)$$

where  $N$  is the total sample size (summing over all the groups),  $k$  is the total number of groups,  $R_j$  is the sum of the ranks for group  $j$ , and  $n_j$  is the sample size for group  $j$ .

- P-values for  $H$  can be drawn from the chi-square distribution with  $k - 1$  degrees of freedom. A significant test statistic indicates that at least two medians are significantly different from each other. Like an ANOVA, however, the Kruskal-Wallis test will not tell you exactly which groups differ.
- After obtaining a significant result from the Kruskal-Wallis test, you will want to further examine the medians of the groups so that you can understand the nature of the relation between the categorical IV and the DV. Researchers have developed some specific post-hoc methods for the Kruskal-Wallis test, but we can also use two of the methods presented in these notes to follow up significant ANOVAs.
  1. Using a Bonferroni or Sidak correction to the significance level  $\alpha$ , we can perform multiple Mann-Whitney U tests to see which of the individual groups are significantly different from each other. Just like for ANOVA, you would want to divide your  $\alpha$  by  $k - 1$ , since there are only  $k - 1$  independent comparisons among  $k$  groups.
  2. You can also treat the Kruskal-Wallis as an omnibus test. In this case you would use a significant result on the Kruskal-Wallis test as justification to perform a set of uncorrected Mann-Whitney U tests to see which of the individual groups are significantly different from each other.
- To perform a Kruskal-Wallis test in SPSS you would take the following steps.
  - Choose **Analyze** → **Nonparametric Tests** → **K Independent Samples**.
  - Move the DV to the **Test Variable List** box.
  - Move the IV to the **Grouping Variable** box.
  - Click the **Define Range** box.
  - Enter the minimum and maximum values for your IV in the corresponding boxes. In order to perform a Kruskal-Wallis test in SPSS, your IV must use numbers to represent different groups.
  - Click the **Continue** button.
  - Click the **OK** button.
- The SPSS output will contain the following sections.
  - *Ranks*. Reports the sample size and mean rank for each of your groups.
  - *Test Statistics*. Reports the value of  $H$  (in the row labeled **Chi-square**, its degrees of freedom, and its p-value (in the row labeled **Asymp. Sig.**).

## 3.5 Chi-square test of independence

- All of the tests we have discussed in this chapter so far test the relation between a categorical IV and a continuous DV. The *chi-square test of independence*, on the other hand, allows you to test the relation between two categorical variables.
- Suppose a researcher brought male and female participants into the lab and asked them whether they prefer the color blue or the color green. If there is a relation between gender and color preference, we would expect that the proportion of men who prefer blue would be different than the proportion of women who prefer blue. In general, you have a relation between two categorical variables when the proportion of subjects found under each level of the first variable changes across the different levels of the second variable.
- To perform a chi-square test by hand you would take the following steps.
  - Create a two-way table where each row of the table represents a level of the first variable and each column of the table represents a level of the second variable. The value of a given cell should be equal to the number of subjects that have the value corresponding to the row on the first variable paired with the value corresponding to the column on the second variable. The values in this table are called the *observed frequencies*.
  - Next you must determine what frequencies you would expect if there were actually no relation between your two variables. To compute these you must first determine the *marginal proportions* for your two variables. These are the proportions of the observations that appear in each level of one variable averaging over the different levels of the other variable. You can find the *expected proportion* of observations that would be found in each cell of a two-way table by multiplying the marginal proportion for the row by the marginal proportion for the column. The *expected frequencies* can then be calculated by multiplying the expected proportion of each cell by the total sample size.
  - Once you have the observed and expected frequencies for each cell in your design, you can compute the test statistic

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \left[ \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right], \quad (3.12)$$

where  $r$  is the number of rows,  $c$  is the number of columns,  $O_{ij}$  is the observed frequency for cell  $ij$ , and  $E_{ij}$  is the expected frequency for cell  $ij$ .

- The test statistic  $\chi^2$  follows a chi-square distribution with  $(r - 1)(c - 1)$  degrees of freedom. A significant statistic indicates that the proportions of subjects in the levels of one of your variables changes across different levels of the other variable.
- To perform a chi-square test of independence in SPSS you would take the following steps.
  - Choose **Analyze** → **Descriptive Statistics** → **Crosstabs**.
  - Move one of the variables to the **Row(s)** box.
  - Move the other variable to the **Column(s)** box.
  - Click the **Statistics** button.
  - Check the box next to **Chi-square**.
  - Click the **Continue** button.
  - Click the **OK** button.
- The SPSS output will contain the following sections.
  - *Case Processing Summary*. Provides information about missing values in your two variables.

- *Crosstabulation*. Provides you with the observed counts within each combination of your two variables.
- *Chi-Square Tests*. The first row of this table will give you the chi-square value, its degrees of freedom and the p-value associated with the test. Note that the p-values produced by a chi-square test are inappropriate if the expected count is less than 5 in 20% of the cells or more. If you are in this situation, you should either redefine your coding scheme (combining the categories with low cell counts with other categories) or exclude categories with low cell counts from your analysis.

### 3.6 Equating experimental groups

- All the analyses discussed in the last two chapters (except for the one-sample t and Z tests) are designed to determine the relation between a categorical IV with a measured DV. In each case, we conclude that there is a significant relation between the IV and the DV when we observed differences in the value of the DV vary across the different levels of the IV.
- The validity of this conclusion rests on the assumption that your factor represents the only substantial differences among your groups. Other differences among your groups can act as *alternative explanations* for the observed differences on the DV.
- It is therefore important that when you are assigning subjects to conditions that you do so in a way that will create groups that are more or less equivalent. There are two primary methods that researchers use to do this.
  1. *Matching*, where you make sure that the people in your groups are equivalent on characteristics that you think could influence the DV.
  2. *Randomization*, where each subject has an equal probability of being in any of your groups.
- When attempting to match your groups, you can choose to perform the matching at the individual level or at the group level. If you match at the individual level, then for each person in one group you have a different person with similar features in each of the other groups. If you match at the group level, you simply try to make sure that the average values of the characteristics for each of the groups is the same. These methods are equally good at removing alternative explanations, but require different analytic methods. If you choose to match at the individual level you will need to analyze your data using a within-subjects analysis, where your unit of analysis is a match. If you choose to match at the group level you will need to analyze your data using a between-subjects analysis.
- While randomization does not guarantee that your groups will be the same on all of the important characteristics, it still is a good way of equating your groups. If you randomly assign people to your groups, you would expect that the average value in one of your groups would be the same as the average value in your other groups. In reality, your groups will never be exactly the same, and it is possible that there might be some characteristic that differs substantially across groups. However, the nice thing is that this is purely random variability, which is explicitly accounted for by our statistics.
- The groups formed by matching typically are more similar to each other than the groups formed by random assignment. However, if you had to choose a single way to equate your groups, it is actually better to randomize them than it is to match them. There are two primary reasons for this.
  1. Matching only equates your groups on the specific set of characteristics you chose to match on. If there are characteristics that affect the DV that you don't know about, you could conclude that there is an effect of your factor when there really isn't one. Randomization, on the other hand, equates your groups on everything, whether you know about it or not.
  2. Any differences that may appear between your groups after randomizing is going to be due to random error. The p-values that we compute take into account the likelihood that such random influences might have stacked up in such a way as to give us the observed test statistic. When

matching, on the other hand, it is possible that you develop systematic biases in the way that you assign subjects to groups. Our p-values cannot take this into account, so the conclusions that you draw from your analyses could be invalid, and you would have no way of knowing how likely this is.

- It is possible to use both randomization and matching at the same time, which solves many problems. This is best done when you are matching at the individual level. In this case you would first identify a set of subjects that are all matched on the characteristics you believe could influence the DV. You would then randomly assign one person in each match to a different group. The initial match reduces the differences between your groups, and then the randomization makes sure that there isn't any systematic biases built into the assignment of subjects to groups.

It is also possible to include randomization when you match at the group level. In this case you would first establish which sets of subjects you want to be in the same group so that the different groups are equivalent. You then randomly determine which set is actually assigned to each experimental condition.

- You should note that it is not always possible to randomly assign subjects to groups. If your factor represents a measured variable such as race or gender, you have no choice as to what subjects are in what group. The only possible way to equate your groups in these circumstances is through matching. In this case you should be aware of the fact that there is a possibility that there are variables that you failed to match on that can act as alternative explanations for group differences.

## Chapter 4

# Testing Multiple Between-Subjects Factors

### 4.1 Introduction

- Sometimes you want to conduct a study where you measure or manipulate two or more factors at the same time. In this case you could analyze your data using several independent one-way ANOVAs. However, you also have the option of testing all of your factors in the same analysis using a *multifactor ANOVA*.
- There are several reasons that performing a multifactor ANOVA is better than performing several one-way ANOVAs.
  - Performing several one-way ANOVAs increases the probability that you get a significant result due to chance alone. A multifactor ANOVA, on the other hand, takes into account how many different factors you are testing, so the probability of getting a type I error is not inflated.
  - A multifactor ANOVA allows you to test for *interactions* among the factors. You have an interaction between two factors when the effect of one factor on the DV is different for different levels of the second factor.
  - Variability associated with other factors in the design is removed from the error term in multifactor ANOVA, making the tests more powerful.
  - The tests of individual factors in a multifactor ANOVA automatically control for any other variables you include in your design. This way you can determine the extent to which each factor can *uniquely* explain variability in the DV.
- In this chapter we will first introduce the idea of performing a multifactor ANOVA by considering an ANOVA with just two factors. This will make it easier to see the way that the analysis works. After introducing the basics, we will then consider how to extend those ideas to test models that have three or more factors. Finally we will discuss methods for performing specific group comparisons in multifactor ANOVA.

### 4.2 Two-way between-subjects ANOVA

- A two-way between-subjects ANOVA is used to determine the joint ability of two categorical IVs to predict the value of a continuous DV.
- In this section, we will refer to the two categorical IVs as *Factor A* and *Factor B*. We will use the symbol  $a$  to refer to the number of levels within Factor A, and the symbol  $b$  to refer to the number of levels within Factor B.



- To perform a two-way between-subjects ANOVA, each of your subjects must be in one level of Factor A and in one level of Factor B. Each specific combination of one level from Factor A with one level of Factor B is referred to as a *cell* of your design. Cell  $ij$  refers to the combination of the  $i$ th level of Factor A with the  $j$ th level of Factor B.
- We will assume that you conduct your study using a *factorial design*. This means that you have at least one observation at every possible combination of the two IVs. You will have a total of  $ab$  cells when you create a factorial combination of Factor A and Factor B. It is more difficult to analyze data from studies in which you do not have a factorial combination of all of your IVs, so throughout these notes we will always assume that your data comes from a factorial design.
- While ANOVA does not require that you have the same number of subjects in each cell of the design, we will make that assumption in this section to simplify our equations. Having the same number of subjects in each cell is called having a *balanced design*. In section 4.3 we will provide more general equations that do not assume that you have a balanced design. We will use the symbol  $n$  to refer to the number of subjects in each cell.
- We will use the following symbols to refer to the values of the DV.
  - $Y_{ijk}$  = the value of the DV for the  $k$ th subject in cell  $ij$
  - $\bar{Y}_{ij}$  = the mean of all of the subjects in cell  $ij$
  - $\bar{Y}_{i..}$  = the *marginal mean* for the  $i$ th level of Factor A, which is the mean of all of the subjects in the  $i$ th level of Factor A, no matter what level they have for Factor B
  - $\bar{Y}_{.j}$  = the marginal mean for the  $j$ th level of Factor B, which is the mean of all of the subjects in the  $j$ th level of Factor B, no matter what level they have for Factor A
  - $\bar{Y}_{...}$  = the *grand mean*, which is the mean of all of the subjects in all the cells
- The purpose of an ANOVA is to use your categorical IVs to explain the variability in your DV. In section 3.2 we saw how a one-way between-subjects ANOVA partitions the variability in the DV into the part that can be explained by the categorical IV (the between-groups variance) and the part that is not explained by the categorical IV (the within-groups variance). A two-way between-subjects ANOVA does basically the same thing, except that the presence of the second IV gives you additional ways to explain variance in the DV. One of the best ways to understand how a two-way between-subjects ANOVA works is to see how it partitions the variability in the DV.
  - The total amount of variability in the DV is measured by the *total sum of squares*, which can be computed using the formula

$$SS_{\text{total}} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{...})^2. \quad (4.1)$$

From this we can see that the total variability is based on how different each of the observations is from the grand mean.

- The variability in the DV that can be explained by what particular cell each observation falls in is called the *model sum of squares*, which can be computed using the formula

$$SS_{\text{model}} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij} - \bar{Y}_{...})^2. \quad (4.2)$$

From this we can see that the amount of variability that can be explained by the IVs is determined by how different each of the cell means is from the grand mean. The model sum of squares tells you how much of the variability in the DV can be explained by your two IVs. It is conceptually similar to the between-groups sum of squares we computed for a one-way between-subjects ANOVA. In fact, if you created a single factor that had a different level for each of the combinations of Factor A and Factor B (so that it would have a total of  $ab$  levels), the between groups sum of squares from a one-way ANOVA would be exactly equal to the model sum of squares.

- The variability that cannot be explained by your IVs is called the *sum of squared error*, and can be computed using the formula.

$$\text{SSE} = \text{SS}_{\text{total}} - \text{SS}_{\text{model}} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij.})^2. \quad (4.3)$$

From this we can see that the sum of squared error is based on the difference between the value of each observation and the mean value for those in its cell. The sum of squared error is conceptually the same as the within-groups sum of squares we computed for a one-way between-subjects ANOVA.

- The variability explained by the IVs (the model sum of squares) can be further divided into three parts.
  - The *sum of squares for Factor A* represents the extent to which there is an independent influence of Factor A on the DV, and can be computed using the formula

$$\text{SS}_A = nb \sum_{i=1}^a (\bar{Y}_{i..} - \bar{Y}_{...})^2. \quad (4.4)$$

From this we can see that the sum of squares for Factor A is based on how different the marginal means for Factor A are from the grand mean. The independent influence of Factor A on the DV is also referred to as the *main effect of Factor A*.

- The *sum of squares for Factor B* represents the extent to which there is an independent influence of Factor B on the DV, and can be computed using the formula

$$\text{SS}_B = na \sum_{j=1}^b (\bar{Y}_{.j.} - \bar{Y}_{...})^2. \quad (4.5)$$

From this we can see that the sum of squares for Factor B is based on how different the marginal means for Factor B are from the grand mean. The independent influence of Factor B on the DV is also referred to as the *main effect of Factor B*.

- The main effects of our variables each measure the extent to which the IVs have an independent influence on the DV. However, it is also possible that a particular combination of your two IVs might have an effect on the DV that can't be explained by either variable alone. The variability that isn't explained by either factor alone but which can be explained by the particular combination of the factors is referred to as the *interaction sum of squares*, and can be computed by the formula

$$\text{SS}_{AB} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2. \quad (4.6)$$

The interaction sum of squares is based on how much of the variability in the cell means cannot be explained by the marginal means of your two factors. Your interaction sum of squares increases whenever the specific combination of the factor levels is important.

- A two-way between-subjects ANOVA will provide you with a test of the overall model (the combined ability of your IVs to predict the DV), tests of the main effects for each of your IVs, as well as a test of the interaction between your IVs. The ANOVA table presented in Figure 4.1 below summarizes the computations for these tests.
- The test of the overall model tells you whether the combination of your IVs can account for a significant amount of variability in the DV. It tells you whether the main effects and interaction together allow you to explain a significant amount of the variability in the DV. The test of the overall model is often used as an omnibus test, such that people only examine the tests of the individual effects if the test of the overall model is significant.

Figure 4.1: ANOVA table for a balanced two-way between-subjects design.

Source	SS	df	MS	Test
Main effect of A	$SS_A$	$a - 1$	$MS_A = \frac{SS_A}{a-1}$	$F = \frac{MS_A}{MSE}$
Main effect of B	$SS_B$	$b - 1$	$MS_B = \frac{SS_B}{b-1}$	$F = \frac{MS_B}{MSE}$
A×B interaction	$SS_{AB}$	$(a - 1)(b - 1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F = \frac{MS_{AB}}{MSE}$
Overall model	$SS_{\text{model}} = SS_A + SS_B + SS_{AB}$	$ab - 1$	$MS_{\text{model}} = \frac{SS_{\text{model}}}{ab-1}$	$F = \frac{MS_{\text{model}}}{MSE}$
Error	SSE	$ab(n - 1)$	$MSE = \frac{SSE}{ab(n-1)}$	
Total	$SS_{\text{total}} = SS_{\text{model}} + SSE$	$abn - 1$		

- The test of the main effect of Factor A tells you whether there is a significant effect of Factor A on the DV, averaging over the different levels of Factor B. A significant test indicates that at least two of the marginal means for Factor A are significantly different from each other. Similarly, the test of the main effect of Factor B tells you whether there is a significant effect of Factor B on the DV, averaging over the different levels of Factor A.
- The test of the A×B interaction (pronounced “A by B interaction”) tells you how the relation between one factor and the DV is influenced by the levels of the other factor. A significant interaction effect indicates that the effect of one factor on the DV changes across the different levels of the other factor. A nonsignificant interaction effect indicates that the effect of each of your factors on the DV is same at all levels of the other factor.

Interaction effects are symmetric. If the effect of Factor A on your DV varies across the different levels of Factor B, then it is also the case that the effect of Factor B on your DV varies across the different levels of Factor A.

ANOVA only produces an accurate test of the A × B interaction when you also include the main effect for A and the main effect for B in the model. If you try to test the A × B interaction by itself, ANOVA will often report a significant interaction when there really isn’t one.

- You will notice that the test statistic for each effect is equal to the mean square for that effect divided by the mean square error. The numerator of this fraction represents the extent to which the effect can explain variability in the DV, and the denominator of this fraction represents how much random variability you have in your data. The p-values for the test statistics are drawn from the F distribution with numerator degrees of freedom equal to the df for the effect, and denominator degrees of freedom equal to the error df.
- The tests that we get from a two-way between-subjects ANOVA will tell us whether our IVs have main effects or interaction effects on the DV, but they do not tell us the nature of these effects. Section 4.5 discusses how to use mean comparisons to determine exactly how your different cells relate to each other.
- Sometimes people will use a graph to illustrate the cell means from a two-way between-subjects ANOVA. In this case, the DV is usually placed on the Y-axis and Factor A is usually placed on the X-axis. The means for each cell are then plotted on the graph, and lines are used to connect the values from cells with the same level of Factor B.

If there is a significant effect of Factor A, then you would expect that there would be large differences among these column averages. If there is a significant effect of Factor B, then you would expect that

there would be large differences among the average values for each of the lines. If there is a significant effect of the interaction, then you would expect that the different lines would not be parallel to each other.

- Sometimes people will use a table to illustrate the cell means from a two-way between-subjects ANOVA. In this case the rows of the table usually correspond to the different levels of Factor A and the columns of the table usually correspond to the different levels of Factor B. The value inside the table is equal to the mean of the observations that are in the level of Factor A corresponding to the row and the level of Factor B corresponding to the column.

You can determine whether you have a main effect of either of your Factors by looking at the marginal means. If there are large differences among the row means, then you probably have a main effect of Factor A. If there are large differences among the column means, then you probably have a main effect of Factor B. To determine if you have an interaction between your factors, you should look to see if the differences between the levels of one of your factors are approximately the same across the levels of the other factor.

### 4.3 Multifactor between-subjects ANOVA

- The main difference between a two-way ANOVA and a multifactor subjects ANOVA is that the latter allows you to include any number of categorical IVs in your predictive model. This not only increases the number of main effects and two-way interactions in the model (the type of interaction you have in a two-way ANOVA), but also provides you with the opportunity to test for higher-order interactions.
- The first thing that you need to do when performing a multifactor ANOVA is determine the *design* of your analysis. This is essentially a description of what factors you are testing, what levels you have in each factor, whether each factor varies between or within subjects, and how those factors are combined with each other.
  - In this chapter we will assume that all of your factors vary between subjects. We will consider multifactor ANOVAs that include within-subjects factors in Chapter 6.
  - We will assume that you have a factorial combination of all of your factors when performing a multifactor ANOVA. This means that you must have observations at every possible combination of your factors. You can compute the total number of cells that this requires by multiplying together the number of levels you have in each of your factors. For example, if you had one factor with 3 levels, one factor with 2 levels, and one factor with 4 levels, a factorial design would have a total of  $3 \times 2 \times 4 = 24$  different cells.
  - An example of the way you would report the design for a multifactor between-subjects ANOVA would be the following. “The study was conducted using a 3 (List type)  $\times$  2 (Study condition)  $\times$  2 (Test condition) between-subjects factorial design. The list type factor referred to whether participants were asked to memorize a list of words, a list of pronounceable nonwords, or a list of unpronounceable nonwords. The study condition factor referred to whether participants studied their list in a noisy or quiet environment. The test condition factor referred to whether participants tried to recall items from the list in either a noisy or quiet environment. The dependent variable was the number of words participants were able to recall from the list.” More detailed information about how to report the design of a study is provided in Section 6.5.
- The next thing you must do is determine what effects you want your ANOVA to test. This is referred to as *building your statistical model*. Your statistical model is the collection of all of the effects that you want to examine.
  - You can test the main effects of each factor in your design. The test of a main effect of a factor tells you whether there are any significant differences among the marginal means for that factor. It essentially tells you whether there is an influence of the factor on the DV averaging over the levels of the other factors in the design.

- You can test for any or all two-way interactions between pairs of your factors. The interpretation of a two-way interaction in a multifactor ANOVA is exactly the same as the interpretation of the interaction effect in a two-way between subjects ANOVA. It tells you whether the influence of the first factor in the interaction on the DV depends on the level of the second factor in the interaction, averaging over any other factors in the model. You must include tests of the main effects for both of the factors composing a two-way interaction for the test of the interaction effect to be valid.

Two-way interactions are symmetric, so if the effect of the first factor on the DV depends on the level of the second factor, then it is necessarily the case that the effect of the second factor on the DV depends on the level of the first factor.

- You can also test for higher-order interactions among your variables. Some examples would be three-way, four-way, or five-way interactions. Higher-order interactions indicate that the nature of a lower-order interaction depends on the level of another factor. For example, the presence of a three-way interaction indicates that the nature of the two-way interaction between the first two factors changes across different levels of the third factor. Similarly, the presence of a four-way interaction indicates that the nature of the three-way interaction between the first three factors changes across different levels of the fourth factor.

In order for the test of a higher-order interaction to be valid, your statistical model must also include the main effects for each factor in the interaction, as well as all possible lower-order interactions that can be formed among the factors in the interaction. For example, let us say you wanted to test a three-way interaction between Factor A, Factor B, and Factor C (referred to as the  $A \times B \times C$  interaction). For the test of this interaction to be valid, your statistical model must also include the main effects of all three factors, as well as the  $A \times B$ ,  $B \times C$ , and  $A \times C$  interaction effects.

Higher-order interactions are symmetric. For example, let us say that you have a three-way interaction between Factor A, Factor B, and Factor C. One way to interpret this interaction would be to say that the  $A \times B$  interaction changes across the different levels of Factor C. However, you could also say that the  $B \times C$  interaction changes across the different levels of Factor A, or that the  $A \times C$  interaction changes across the different levels of Factor B.

- You should consciously decide exactly what effects you want to include in your statistical model. Each effect you include in your model will pull one or more degrees of freedom away from the error term, making it slightly more difficult to obtain significant results. It is especially important to think about the necessity of including higher-order interaction terms. People can usually understand a two-way interaction, and with a little work you can usually come up with a conceptual interpretation of a three-way interaction. Explaining four-way or five-way interactions is a very difficult task, so people in the social sciences will commonly not bother to include them in their statistical models. When the purpose of your analysis is to develop a theoretical explanation of the variability in the DV, there is little to be gained by testing uninterpretable effects.
- After you have decided upon your statistical model, you must then compute the sum of squares for each effect. The sum of squares represents the amount of variability in the DV that can be explained by each effect. The formulas for the sums of squares will vary depending on the number of factors in your model. These computations will usually be performed by a computer, so we will not provide specific equations in these notes. More information on calculating the sums of squares in multifactor ANOVA can be found in Chapter 23 of Neter, Kutner, Nachtsheim, and Wasserman (1996).
- The sum of squares for your overall model represents how much variability in the DV can be explained by your entire statistical model. In a balanced design it can be computed by adding up the sums of squares for all of the effects in your model.
- You then divide the sum of squares for each effect by its degrees of freedom to compute the mean square for each of your effects. The degrees of freedom for each of your main effects will be equal to  $(k - 1)$ , where  $k$  is the number of groups in the corresponding factor. The degrees of freedom for an interaction can be determined by multiplying together the degrees of freedom from the main effects for

each of the factors involved in the interaction. For example, let us say you are testing the interaction between Factor A with  $a$  levels, Factor B with  $b$  levels, and Factor C with  $c$  levels. The degrees of freedom for the  $A \times B \times C$  interaction would be equal to  $(a - 1)(b - 1)(c - 1)$ .

You can also compute the mean squares for your overall model by dividing the sum of squares for your model by the total degrees of freedom in your model. You can compute the degrees of freedom for your model by adding up the degrees of freedom for each of the effects in your model.

- Next you determine the sum of squared error, which is the variability in the DV that is not explained by any of the effects in your statistical model. You then divide the sum of squared error by the degrees of freedom for error to obtain the mean square error. The degrees of freedom for error is equal to  $N - g - 1$ , where  $N$  is the total number of subjects and  $g$  is the degrees of freedom in your overall model.
- The test statistic for each of your effects can be computed by dividing the mean square for that effect by the mean square error. The statistic will follow a F distribution with numerator df equal to the df for the effect and denominator df equal to the df for error. This holds for main effects, interactions, and the test of the overall model.

Researchers will sometimes use the test of the overall model as an omnibus test, only looking at the tests of the individual effects if the test of the overall model is significant. However, people do regularly publish articles where they look at the tests of the individual effects without reporting the test of the overall model first.

- Multifactor between-subjects ANOVA assumes that your observations are independent except for the effects included in your statistical model, that the variance of the DV within each cell of your design is approximately the same, and that your DV is normally distributed within each cell of your design.
- You can use tables or graphs to help you interpret the significant effects that you find. These can become somewhat complex if you want to look at higher-order interactions. For these you will typically need to use multiple tables or graphs to display the complete set of cell means.
- To perform a multifactor between-subjects ANOVA in SPSS you would take the following steps.
  - Choose **Analyze** → **General Linear Model** → **Univariate**.
  - Move the DV to the **Dependent Variable** box.
  - Move all of your IVs to the **Fixed Factor(s)** box.
  - By default SPSS will include all possible interactions among your categorical IVs. If this is not the model you want then you will need to define it by hand by taking the following steps.
    - \* Click the **Model** button.
    - \* Click the radio button next to **Custom**.
    - \* Add all of your main effects to the model by clicking all of the IVs in the box labelled **Factors and covariates**, setting the pull-down menu to **Main effects**, and clicking the arrow button.
    - \* Add each of the interaction terms to your model. You can do this one at a time by selecting the variables included in the interaction in the box labelled **Factors and covariates**, setting the pull-down menu to **Interaction**, and clicking the arrow button for each of your interactions. You can also use the setting on the pull-down menu to tell SPSS to add all possible 2-way, 3-way, 4-way, or 5-way interactions that can be made between the selected variables to your model.
    - \* Click the **Continue** button.
  - To obtain tables containing marginal or cell means you should take the following steps.
    - \* Click the **Options** button.
    - \* Move any main effects for which you want to obtain the marginal means to the **Display Means for** box.

- \* Move any interaction effects for which you want to obtain the means within the different cells of that effect to the **Display Means for** box.
- \* Click the **Continue** button once you have selected all of the tables you want to see.
- o To obtain graphs of the marginal or cell means you should take the following steps.
  - \* Click the **Plots** button.
  - \* To plot a main effect, move the factor to the **Horizontal Axis** box and click the **Add** button.
  - \* To plot a two-way interaction, move the first factor to the **Horizontal Axis** box, move the second factor to the **Separate Lines** box, and click the **Add** button.
  - \* To plot a three-way interaction, move the first factor to the **Horizontal Axis** box, move the second factor to the **Separate Lines** box, move the third factor to the **Separate Plots** box, and click the **Add** button.
  - \* Click the **Continue** button once you have added all of the plots you want to see.
- o Click the **OK** button.
- The output of this analysis will contain the following sections.
  - o *Between-Subjects Factors*. Lists how many subjects are in each level of each of your factors.
  - o *Tests of Between-Subjects Effects*. The row next to the name of each factor or interaction reports a test of whether there is a significant relation between that effect and the DV.
  - o *Estimated Marginal Means*. Contains a table for each effect for which you requested means. The rows in the tables will correspond to a cell within the effect. The different columns provide the mean, standard error of the mean, and the boundaries for a 95% confidence interval around the mean for observations within that cell.
  - o *Profile Plots*. Displays any graphs of the marginal or cell means that you requested.

## 4.4 Working with unbalanced designs

- You do not need to have the same number of subjects in each cell of your design, although you do have to have at least some observations in each cell in order to have a factorial design. When you have unequal numbers of subjects in your cells you are said to have an *unbalanced design*.
- ANOVA does not assume that you have equal cell sizes, so working with an unbalanced design does not violate any of the assumptions of ANOVA. However, it is important to note that ANOVA is less robust to violations of the assumption of equal variances when the sample sizes in your cells differ by a large amount.
- The power of the tests provided by ANOVA decreases as the differences among the sample sizes of your cells increases.
- Probably the most important effect of having an unbalanced design is that it creates correlations among your IVs. As an example, let us assume that you conducted a study to examine the effect of a treatment on depression across two different sites. Let's further assume that you ended up with an unbalanced design, as illustrated in Figure 4.2 below.

Figure 4.2: Correlated IVs from an unbalanced design.

	Treatment	Control
Texas site	n = 10	n = 40
New Jersey site	n = 40	n = 10

Notice that this design creates a relation between your two IVs: knowing what site a participant came from gives me some information about what condition they are probably in. Participants from the Texas site are more likely to be in the treatment condition, while participants from the New Jersey site are more likely to be in the control condition. The presence of correlations among your IVs is referred to as *multicollinearity*.

- Multicollinearity influences the way you interpret the effects in your model. Imagine that we analyzed the above data and found that the mean depression score for subjects in the treatment condition was significantly lower than the mean depression score for subjects in the control condition. Ideally we would like to claim that our treatment was successful in reducing depression. The correlation between treatment and site, however, prevents us from being able to draw this conclusion. Since we have more people from Texas in the treatment condition and more people from New Jersey in the control condition, the observed difference between treatment and control could just reflect the fact that people from Texas tend to be less depressed than those from New Jersey.
- One way to think about the influence of multicollinearity is to consider the ability of each of the IVs to predict the DV. When your factors are uncorrelated, each of the effects explains a different part of the variability in the DV. When your factors are correlated, however, the variability that is explained by one effect may also be explained by other effects in the model. In this case you need decide what to do with the variability that is explained by more than one effect when computing the sums of squares. Statisticians have proposed four different ways of computing the sums of squares for your effects in an unbalanced design (Goodnight, 1980).
  - To compute *Type I sums of squares* you must first generate an ordered list of the effects in your model. Any variability that is explained by more than one of your effects is credited to whichever effect occurs earliest in the list. The test of each effect in your list examines the extent to which that effect can explain variability in the DV above and beyond the variability that is explained by effects earlier in the list. Another way of saying this is that the test of each effect *controls* for all of the effects that occur earlier in the list.

In this list your main effects must come before your interaction effects, and lower-order interactions must come before higher-order interactions. As mentioned previously, you cannot get a valid test of a higher-order interaction unless the main effects and all lower-order interactions are already present in the model. By including the lower-order effects earlier in your list you ensure that the tests of all of your higher-order effects are valid.

Within the sets of main effects, the set of two-way interactions, the set of three-way interactions, and so forth, you have complete discretion as to the order of the effects in the list. Usually people will put more effects that are more important or have higher causal priority earlier in the list. Note that different ways of listing out your effects will result in different sums of squares for your effects.
  - *Type II sums of squares* represent the variability that each effect can explain in the DV above and beyond the variability explained by effects that are of the same or of lower order. This means that the tests of your main effects control for the other main effects in the model, the tests of each of your two-way interactions control for both the main effects and the other two-way interactions, and the tests of each of your three-way interactions control for the main effects, the two-way interactions, and the other three way interactions.
  - *Type III sums of squares* represent the variability that each effect can explain in the DV above and beyond that which is explained by any of the other effects in the model. The test of each effect in your model controls for all of the other effects in the model. This means that the test for given effect represents the unique ability of that effect to explain variability in the DV.
  - *Type IV sums of squares* were originally devised to test “balanced” hypotheses when you had missing cells in your design (i.e., you did not have a factorial design). The computations for Type IV sums of squares are somewhat complicated, so we will not go into them here. It turns out that tests provided by Type IV sums of squares are difficult to interpret, even in the situations where they are supposed to be used. In designs with no missing cells, the Type IV sums of squares are always equal to the Type III sums of squares.



- Researchers most commonly use Type III sums of squares, where the test of each effect represents its unique ability to explain variability in the DV. However, the Type III sums of squares (along with Type II and Type IV sums of squares) do not allow you to partition the variance in the DV. If you sum up the sums of squares from the individual effects you will usually come up with a sum that is actually less than the amount of variability that can be predicted by your statistical model. The Type I sums of squares, on the other hand, assigns credit to all of the variance you can predict in the DV to specific IVs. However, the tests that you get when you use Type I sums of squares are more complicated to interpret.
- Since the tests of your effects in an unbalanced design are affected by the correlations among your predictors, they are no longer solely reflect the raw differences among the cell means. Instead, what they represent are the cell means you would expect to have if your factors were not correlated with each other. The raw means are therefore not useful when you are trying to interpret your significant effects. However, you can obtain *least squares means* that represent the actual differences that your ANOVA considers when testing your effects. The values of your least squares means represent the values you would expect if you actually had a balanced design. They are equal to the raw means if you have a balanced design.
- By default, the results produced by SPSS will be based on the Type III sums of squares. However, you can change this when using the **Analyze** → **General Linear Model** → **Univariate** procedure by taking the following steps.
  - Set up your multifactor ANOVA as described in section 4.3, but do not press **OK** at the end.
  - Click the **Model** button.
  - Set the *Sum of Squares* pull-down menu to the type of sums of squares that you want SPSS to use.
  - Click the **Continue** button.
  - Click the **OK** button.
- If you want to use Type I sums of squares, you should make sure that the order of the variables in your model matches the order of the variables in your list. You can change the order of the variables in your model by clicking the **Model** button and then adding the variables to the model in the order in which you want them to be tested.
- Your output will contain the same sections no matter what type of sums of squares you decide to use. The only thing that will change will be the specific values in the tables. Your tests will be based on the type of sums of squares that you selected, and any tables or graphs of your means will be based on the least squares means instead of the raw means. SPSS refers to the least squares means as the **Estimated marginal means**.

## 4.5 Specific group comparisons

- The tests of our main and interaction effects in multifactor ANOVA will not tell us which specific groups within the effect are different from each other. After obtaining a significant effect we must therefore compare the means of our groups so that we can understand the nature of the effect. Just as in one-way ANOVA, we can do this using preplanned contrasts, post-hoc comparisons, or by nonstatistical methods.

### 4.5.1 Preplanned contrasts

- You use preplanned contrasts when there is a comparison among your group means that is of specific theoretical interest. You should specifically determine any preplanned contrasts you want to perform before you actually look at the pattern of cell means. You should also only examine a small number of preplanned contrasts based on a given effect.

- The basic purpose of and method for performing preplanned contrasts is the same in multifactor ANOVA as it is in one-way ANOVA. The information contained in this section will concentrate on the aspects of performing preplanned contrasts that are unique to multifactor ANOVA. You should therefore be familiar with the material on performing preplanned contrasts in single-factor designs presented in section 3.3.1 before trying to understand the material in this section.
- There are two basic ways to test a preplanned contrast in a multifactor between-subjects design. The first (and simplest) method transforms your IVs so that you can use the procedure for testing contrasts in a one-way between-subjects ANOVA. The second method requires you to define multiple contrast vectors, one for each effect in the model.
- The first method rests on the assumption that the comparison defined by a contrast is the same whether the individual groups represent levels of a single factor or whether they represent levels of a multifactor design, as long as the groups themselves are the same. The test of a contrast is determined by three things: the group means, the sample sizes, and the mean square error from the corresponding ANOVA. As long as we are using the same groups, we know that the means and sample sizes have to be the same. We can also deduce that the mean square error will be the same. Recall that the sum of squares for your overall model in a multifactor design is exactly the same as that computed for a single factor that has a different level for each of the groups in your design. This means that the sum of squared error and the mean square error from those two models must also be the same. We therefore can conclude that it is acceptable to test our contrast in a one-way ANOVA model that contains a single factor differentiating all of the groups in the design.
- While you could manually recode the levels of your IVs into a new factor, there is actually a trick that allows you to create your new factor using a single formula. In order to use this trick, your categorical IVs must all use numbers to represent the different groups. As an example, let's assume that you wanted to use a single factor to code the groups created by a factorial combination of three categorical IVs. We will say that the first IV (*A*) has two levels coded by the values 1 and 2, the second IV (*B*) has two levels again coded by the values 1 and 2, and that the third IV (*C*) has three levels coded by the values 1, 2, and 3. In this case, you can calculate the value of your new factor using the formula

$$X = (A \times 100) + (B \times 10) + C, \quad (4.7)$$

where  $X$  is the value of the new factor you want to create, and  $A$ ,  $B$ , and  $C$ , are the values of the three original IVs. In this example, the value of  $X$  would always be a three-digit number. The first digit would be equal to the level on variable  $A$ , the second digit would be equal to the level on variable  $B$ , while the third digit would be equal to the level on variable  $C$ . You can therefore directly see the correspondence between the levels of the original factors and the level of the composite variable just by looking at the digits of the composite variable.

If you ever want to create a composite of a different number of factors (besides 3 factors, like in this example), you follow the same general principle, basically multiplying each factor by decreasing powers of 10, such as the equation

$$X = (A \times 10) + B, \quad (4.8)$$

which recodes the combination of two IVs, and the equation

$$X = (A \times 1000) + (B \times 100) + (C \times 10) + D, \quad (4.9)$$

which recodes the combination of four IVs. Once you have created your new composite factor, you can simply perform your preplanned contrast using the guidelines presented in Section 3.3.1 by choosing which means you want to compare.

- The second way to test a preplanned contrast in a multifactor between-subjects design is to define your contrast directly in terms of the different effects in your design. The first thing you must do is decide exactly what comparison among your cell means you want to make. Most commonly you will want to compare the mean of one set of cells to the mean of a second set of cells. In this case you will want to label one as the “positive” set and the other as the “negative” set. This decision is arbitrary, but will affect how you interpret the value of the contrast.
- After you have decided what comparison you want to test, you must define each of the groups involved in the comparison in terms of your effects. Each level of your main effects and each cell within an interaction can be thought of as providing a specific contribution to your cell means. For example, Figure 4.3 below illustrates the various effects present in a two-factor design. You will

Figure 4.3: Effects on cell  $ij$  within a two-factor design.

Effect	Contribution to cell mean
Grand mean	$\bar{Y}_{...}$
Factor A level $i$	$A_i = \bar{Y}_{i..} - \bar{Y}_{...}$
Factor B level $j$	$B_j = \bar{Y}_{.j.} - \bar{Y}_{...}$
Interaction for cell $ij$	$AB_{ij} = \bar{Y}_{ij.} - \bar{Y}_{...} - A_i - B_j$ $= \bar{Y}_{ij.} - \bar{Y}_{...} - (\bar{Y}_{i..} - \bar{Y}_{...}) - (\bar{Y}_{.j.} - \bar{Y}_{...})$ $= \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$

find that the sum of the grand mean, the effect of Factor A level  $i$ , the effect of Factor B level  $j$ , and the effect of the A×B interaction is equal to the mean of cell  $ij$ ,  $\bar{Y}_{ij.}$ . When you define a contrast in a multifactor ANOVA you have to do so based on these types of effects instead of simply defining it in terms of the cells of the design.

- You can use the following steps to determine how to represent a particular cell in terms of the effects in a multifactor design.
  1. Write down the name of a main effect in your design.
  2. For each level in the main effect, write down a 1 if your cell is in that level, and write down a 0 if it is not. This should give you a number of values equal to the number of levels in your main effect.
  3. Go back to Step 1 until you run out of main effects.
  4. Write down the name of an interaction in your design.
  5. List out the different combinations that can be made among the IVs involved in the interaction. This list should start with the lowest levels of all your IVs, and should be organized such that the variable last in the name of the interaction changes fastest, while the one earliest in the list changes slowest. For example, let us say you wanted to list the combinations of the A×B×C interaction, where each of the IVs had two levels. You should obtain the following list.
    - A1 B1 C1
    - A1 B1 C2
    - A1 B2 C1
    - A1 B2 C2
    - A2 B1 C1
    - A2 B1 C2
    - A2 B2 C1
    - A2 B2 C2

6. For each combination in the list you just created, you should write down a 1 if your cell falls within that combination, and write down a 0 if it does not.
7. Go back to Step 4 until you run out of interaction effects.

The list of numbers following each of the effects in your design are referred to as the *effect vectors* for your cell.

- After you determine the effect vectors for all of the cells involved in your comparison, you compute a contrast weight vector for each of your effects. For each effect, you would take the following steps to compute its contrast weight vector.
  1. Add together all of the effect vectors corresponding to the effect of interest from the cells in the positive set. Divide each number in this result by the total number of cells in the positive set to obtain the positive weight vector.
  2. Add together all of the effect vectors corresponding to the effect of interest from the cells in the negative set. Divide each number in this result by the total number of cells in the negative set to obtain the negative weight vector.
  3. Subtract the negative weight vector from the positive weight vector to obtain the contrast weight vector for that effect.
- The collection of contrast weight vectors can then be used to test the value of your contrast. The formulas to do this involve matrix algebra so we will not present them here. Most statistical software packages will be able to test your contrast if you have the contrast weight vectors.
- To use SPSS to test a preplanned contrast within a multifactor between-subjects ANOVA, you would take the following steps.
  - Use the steps described in Section 4.3 to set up your multifactor ANOVA, but do not press **OK** at the end.
  - Click the **Paste** button. SPSS will copy the syntax you need to run the multifactor ANOVA you defined into a syntax window.
  - Insert a blank line after the **/CRITERIA = ALPHA(.05)** statement.
  - Type **/LMATRIX =** on the blank line, followed by the name you want to give this contrast in quotes.
  - On the next several lines you will need to provide SPSS with your contrast weight vectors. Take the following steps for each effect in your model that has a contrast weight vector.
    - \* On a new line, write the name of the effect.
    - \* Following each effect, write out the contrast weight vector for that effect. The order in which the conditions are defined in interactions is based on the order of the variables following the **BY** statement. Your vector must be ordered such that IVs listed to the left change slower than those that are to the right. It doesn't matter how you write the name of your interaction in either the **/LMATRIX** or the **/DESIGN** statement.
    - \* You do not have to do either of these steps if the contrast weight vector for an effect consists entirely of zeroes. The effect will not influence the contrast, so you don't even have to list it in the **/LMATRIX** statement.
  - Select the entire block of syntax.
  - Choose **Run** → **Selection**, or click the button with the arrow on it.
- The code you run should look something like the following.

```
UNIANOVA
  income BY race sex
  /METHOD = SSTYPE(3)
```

```

/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/LMATRIX = "Race 1 Sex 1 versus Race 3 Sex 1"
  race 1 0 -1
  sex 0 0
  race*sex 1 0 0 0 -1 0
/DESIGN = race sex race*sex .

```

You could have omitted the line containing the sex contrast weight vector, since it only has zeroes in it.

- Notice that the collection of contrast weight vectors defines a single contrast. If you want to independently test multiple contrasts from the same model you can simply include multiple `/LMATRIX` statements.

You also have the option of jointly testing multiple contrasts. To add a second contrast to your test you would simply need to put a semicolon after the last contrast weight vector from the first contrast, and then follow it with the collection of contrast weight vectors for the second contrast. The test statistic will be significant if any of the contrasts are significantly different from zero, although the test does adjust for the total number of contrasts you are examining.

#### 4.5.2 Post-hoc comparisons

- You use post-hoc comparisons when your ANOVA reports a significant main effect or interaction and you want to explore the means to determine the nature of the effect. They serve the same purpose in multifactor ANOVA that they did in one-way ANOVA.
- To understand the nature of an effect in multifactor ANOVA you will typically want to make comparisons among the least squares means rather than the raw means. As discussed in Section 4.4, the least squares means represent the values you would expect if you had no correlations among your IVs. Since the tests in ANOVA represent each effect's independent ability to explain variability in the DV, they are based on differences among the least squares means (which remove the effect of correlations) rather than differences among the raw means.
- To perform post-hoc analyses for a main effect, you want to consider differences among the estimated marginal means for that factor. To perform post-hoc analyses for an interaction, researchers will typically look at differences among the least squares means for one factor at different combinations of the other factors in the interaction. For example, let us say that you wanted to explore the nature of a three-way interaction between Factor A (with 2 levels), Factor B (with 2 levels) and Factor C (with 3 levels). In this situation you could use post-hoc analyses to test whether the difference between the two levels of Factor A is significant within each of the 6 combinations of Factors B and C.
- It is important to note that the standard post-hoc options available in SPSS (those you get by clicking the **Post Hoc** button) are all based on the raw means, making them inappropriate for exploring results in a multifactor ANOVA unless there are no correlations among your IVs.
- To perform a post-hoc analysis on the least squares means for a main effect in SPSS you would take the following steps.
  - Use the steps described in Section 4.3 to set up your multifactor ANOVA, but do not press **OK** at the end.
  - Click the **Options** button.
  - Move the main effect of interest to the **Display Means for** box.
  - Check the box next to **Compare main effects**.

- Set the **Confidence interval adjustment** pulldown menu to the type of post-hoc analysis you want to perform. Unfortunately the options for comparing least squares means are much more limited than the options for comparing raw means.
- Click the **Continue** button.
- Click the **OK** button.
- To perform post-hoc analyses for interactions in SPSS you should take the following steps.
  - Use the steps described in Section 4.3 to set up your multifactor ANOVA, but do not press **OK** at the end.
  - Click the **Paste** button. SPSS will copy the syntax you need to run the multifactor ANOVA you defined into a syntax window.
  - Insert a blank line after the **/CRITERIA = ALPHA(.05)** statement.
  - Type **/EMMEANS =** on the blank line.
  - On the same line, type **TABLES** followed by the interaction you want to explore in parentheses.
  - On the same line, type **COMPARE** followed by the name of one of your IVs in parentheses. SPSS will compare the least squares means of this variable within each combination of the levels of the other IVs in the interaction.
  - On the same line, type **ADJ** followed by the name of the adjustment method for  $\alpha$  that you want to use. Acceptable values are **LSD** (which does not adjust your  $\alpha$ ), **BONFERRONI**, or **SIDAK**.
  - Select the entire block of syntax.
  - Choose **Run** → **Selection**, or click the button with the arrow on it.
- The code you run should look something like the following.

**UNIANOVA**

```

income BY race sex
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/EMMEANS = TABLES(race*sex) COMPARE(race) ADJ(SIDAK)
/DESIGN = race sex race*sex .

```

- Using either of these post-hoc methods will add the following sections to your ANOVA output.
  - *Estimates.* Reports the least squares means within each cell of your interaction. It also provides the standard error of the mean, as well as a 95% confidence interval around the mean. Recall that SPSS refers to the least squares means as “estimated marginal means.”
  - *Pairwise Comparisons.* Reports the post-hoc comparisons you requested. For a main effect, SPSS tests the difference between each pair of factor levels. For an interaction, SPSS tests the difference between each pair of levels for one factor within each combination of the other factors in the design.

### 4.5.3 Nonstatistical methods

- Just like with a one-way ANOVA, researchers will sometimes choose to rely solely on a table or graph of the means to interpret a significant effect in a multifactor ANOVA. However, when working in a multifactor design you should provide the least squares means rather than the raw means since these form the basis for your tests.
- To obtain a table of the least squares means in SPSS you would take the following steps.

- Use the steps described in Section 4.3 to set up your multifactor ANOVA, but do not press **OK** at the end.
- Click the **Options** button.
- Move the effects for which you want to obtain the least squares means to the **Display Means for** box.
- Click the **Continue** button.
- Click the **OK** button.

SPSS will report the least squares means in a section labelled *Estimates*.

- To obtain a graph of the estimated marginal means in SPSS you would take the following steps.
  - Use the steps described in Section 4.3 to set up your multifactor ANOVA, but do not press **OK** at the end.
  - Click the **Plots** button.
  - You define the way your plot will look in the next screen.
    - \* If you just want to plot a main effect, you should move the appropriate variable to the **Horizontal Axis** box and then click the **Add** button.
    - \* If you want to plot a two-way interaction, you should move one of the variables in the interaction to the **Horizontal Axis** box, move the other variable in the interaction to the **Separate Lines** box, and then click the **Add** button. It doesn't matter where you put each variable, but the graph is usually easier to interpret when you put the factor with the most levels on the horizontal axis.
    - \* If you want to plot a three-way interaction, you should move one of the variables in the interaction to the **Horizontal Axis** box, a second variable in the interaction to the **Separate Lines** box, the last variable to the **Separate Plots** box, and then click the **Add** button. You can assign the variables wherever you like, but the plots are usually easiest to interpret when you put the variable with the most levels on the horizontal axis and the variable with the fewest levels defines the separate plots.

SPSS will not graph interactions involving more than three factors. The graphs you select will be displayed in a section labelled *Profile Plots*.

## 4.6 Blocking factors and ANCOVA

- Recall that the denominator for all of our tests in ANOVA is the mean square error from the overall model. This means that the strength of our tests is based in part on how much unexplained variability that we have in our DV. If we can reduce the amount of unexplained variability, we will improve the tests of our effects.
- One way that you can reduce the unexplained variability in your DV is to find a variable that can explain it. If you then add that variable to your statistical model, the variability associated with it will be pulled away from the sum of squared error and put into the model sum of squares. This will then reduce the denominator for all of your F statistics, making those statistics larger.
- If you use a categorical variable in this fashion it is referred to as a *blocking factor*. If you use a continuous variable in this fashion it is referred to as a *covariate*. Theoretically there is no difference between the functions of a blocking factor and a covariate. Performing an ANOVA while including a covariate is called an *analysis of covariance*, often abbreviated as ANCOVA.
- You should be careful about what variables you choose to add as blocking factors or covariates. If your new variable is correlated with the other IVs, including the new variable in your statistical model will actually hurt your ability to test your IVs. Also, including the blocking factor or covariate in your model will take degrees of freedom away from the error term. If the new variable does not account for

a substantial amount of unique variability in the DV, then the loss of error degrees of freedom could hurt your tests more than the reduction of the unexplained variability helps them.

Sometimes you can actually use this to your advantage. When you include a blocking factor or covariate in your design, the tests of any other factors in your model tell you whether those IVs can explain a significant amount of variability above and beyond that explained by the blocking factor or covariate. For example, let us say that you wanted to examine the effect of marital status on the ratings of how much love couples feel in their relationship. Marital status is typically confounded with how long two people have been involved, so if you just did a t-test examining the relation between marital status and love ratings you couldn't be sure what is really driving the relation. If you performed an ANCOVA including relationship age as a covariate, however, the test of marital status would tell you whether it can explain a significant amount of variability in the ratings independent of that explained by relationship age.

- You typically will not interact blocking factors or covariates with any of the other factors in your design. You are only including these variables in your design to “soak up” variance from the error term. You are not typically interested in interpreting them theoretically, nor do you expect that they will influence the other effects in your design.
- To add a blocking factor or covariate to your model when analyzing your data in SPSS you would take the following steps.
  - Use the steps described in Section 4.3 to set up your multifactor ANOVA, but do not press **OK** at the end.
  - You can add a blocking factor to your model by moving the variable to the **Fixed Factor(s)** box. However, by default SPSS automatically interacts all of your categorical IVs with each other. This means that it will interact your blocking factor with your other IVs, which is not what you want. You will therefore need to explicitly define your statistical model so that you only test interactions among your theoretical factors. See section 4.3 for more information on how to define your statistical model.
  - You can add a covariate to your model by moving the variable to the **Covariate(s)** box. By default SPSS will not interact your covariates with your theoretical factors, so you will not need to manually adjust your statistical model.
- Including a blocking factor or covariate in your design will not change the sections of output you receive from ANOVA. The section labelled *Tests of Between-Subjects Effects* will include a test of the blocking factor or covariate's ability to uniquely explain variability in the DV. While this is usually not of theoretical interest, if this test is not significant then there is typically no reason to keep the new variable in your model since it is not removing a significant amount of variability from the error term.



## Chapter 5

# Testing a Single Within-Subjects Factor

### 5.1 Introduction

- In Section 2.2 we learned how we can use a within-subjects t test to determine whether the mean value on a DV in one condition is significantly different from the mean value on that DV for a different condition when both conditions contain the same subjects. In this chapter we will learn ways to compare the average value on a DV across any number of groups when you have the same or matched subjects in all of your groups.
- You can alternatively think about the procedures in this chapter as testing whether there are differences among the average values of variables in a data set that uses subject as the unit of analysis.

### 5.2 One-way within-subjects ANOVA

- Recall that one of the advantages of using a within-subjects t test over using a between-subjects t test is that the within-subjects t test allows you to control for individual differences. A within-subjects ANOVA also lets you control for individual differences because you have measurements from each subject in each of your groups.
- The primary difference between the computations for between-subjects and within-subjects ANOVA is the way the two methods define error. Variability in responses associated with subject-to-subject differences is included in the error term for between-subjects ANOVA while it is excluded from the error term for within-subjects ANOVA.
- There are actually several different ways to analyze the influence of within-subjects factors in ANOVA. The most common method is to use *repeated measures analysis*, which is what we will primarily discuss in these notes.
- A one-way repeated-measures ANOVA is performed as a hypothesis test with the following characteristics.
  - $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$   
 $H_a$  : At least two of the means are different from each other  
ANOVA doesn't consider the direction of the differences between the group means, so there is no equivalent to a one-tailed test.
  - You must take the following steps to compute the test statistic for a one-way repeated-measures ANOVA.

1. Determine the sum of squares between groups in the same way that you would for a one-way between-subjects ANOVA.
2. Determine the *sum of squares between subjects*, which measures the amount of variability in the DV that can be explained by subject-to-subject differences.
3. Determine the *within-subjects sum of squared error*, which is the within-subject variability that cannot be explained by differences among your groups. You can compute this by subtracting the sum of squares between groups and the sum of squares between subjects from the total sum of squares.  
The within-subjects sum of squared error can also be thought of as the extent to which the differences between your groups vary from subject to subject.
4. Compute the mean squares by dividing each of the sums of squares by their degrees of freedom. The sum of squares between groups has  $k - 1$  df, the sum of squares between subjects has  $N - 1$  df, and the within-subjects sum of squared error has  $(k - 1)(N - 1)$  df, where  $k$  is the number of groups and  $N$  is the total sample size.
5. Compute the test statistic

$$F = \frac{MS_{\text{between groups}}}{MS_{\text{within-subjects error}}}. \quad (5.1)$$

Larger values of  $F$  mean that that the difference between our means is large compared to the random variability in our measurements, so it is less likely that we got our observed results due to chance alone.

- The p-value for this test statistic can be taken from the  $F$  distribution with  $k - 1$  numerator degrees of freedom and  $(k - 1)(N - 1)$  denominator degrees of freedom. You'll notice that this corresponds to the degrees of freedom associated with the  $MS_{\text{between groups}}$  and the  $MS_{\text{within-subjects error}}$ , respectively.
- The one-way within-subjects ANOVA assumes that your observations are independent (except for the effect of your IV and individual differences between subjects), that your DV has a normal distribution within each of your groups, and that your data meet the assumptions of *sphericity*. To have sphericity, the variances of the difference scores between pairs of your groups all have to be equal. For example, let us say that you conduct a study where you collect measurements at three different time points for each subject. You can imagine constructing new variables for each subject that are equal to (time 1 score - time 2 score), (time 2 score - time 3 score), and (time 1 score - time 3 score). The assumption of sphericity would require that the variances of all three of these new variables be approximately equal. If your within-subjects factor only has two levels you will always have perfect sphericity, since there is only a single difference score between your groups.
- Sphericity can be somewhat difficult to check directly, so researchers often check for *compound symmetry* instead. To have compound symmetry, the variances of all of your groups have to be equal and the covariances among all of your groups have to be equal. The variances do not have to be equal to the covariances, however. If your data have compound symmetry you know that your data also have sphericity, but there are some ways that your data can have sphericity without also having compound symmetry. For practical purposes, you know that you have met the assumptions of repeated measures ANOVA if you can demonstrate compound symmetry.
- If your data do not meet the assumption of sphericity you have two options. First, you can correct the degrees of freedom and p-value of your test to take the violation of sphericity into account. Both SPSS and SAS will automatically provide both corrected statistics when you perform a repeated-measures ANOVA. Second, you can use a different method to analyze the effect of your within-subjects factors. Some alternatives would be to use multivariate analysis of variance (MANOVA), mixed linear models, or hierarchical linear models (HLM).
- To perform a within-subjects ANOVA in SPSS you would take the following steps.

- You must have your data set organized so that the subject is the unit of analysis and you have different variables containing the value of the DV at each level of your within-subjects factor.
  - Choose **Analyze** → **General linear model** → **Repeated measures**.
  - Type the name of the factor in the **Within-Subjects Factor Name** box. This just affects how the within-subjects factor is labelled in the output.
  - Type the number of levels the factor has into the **Number of Levels** box.
  - Click the **Add** button.
  - Click the **Define** button.
  - Move the variables representing the different levels of the within-subjects factor to the **Within-Subjects Variables** box.
  - Click the **OK** button.
- The output of this analysis will contain the following sections.
    - *Within-Subjects Factors*. Tells you what variables represent the different levels of your factor
    - *Multivariate Tests*. Provides the results for a test of your within-subjects factor using MANOVA. Multivariate analysis actually provides a matrix of results, which would naturally be very difficult to interpret on its own. Statisticians have therefore developed four common methods of converting the results of a multivariate test to an F-test. The most commonly used and accepted statistic is Wilks' Lambda. More recently statisticians have used the Pillai-Bartlett Trace, since research has indicated that this statistic is somewhat more robust to violations of the model assumptions than Wilks' Lambda.
    - *Mauchly's Test of Sphericity*. Provides a test of the assumption of sphericity. A significant test means that sphericity has been violated, indicating that you should not use the uncorrected results of a repeated-measures analysis.
    - *Tests of Within-Subjects Effects*. Contains the results of a repeated-measures test of your within-subjects factor. If the assumption of sphericity is satisfied, you examine the test provided in the row labelled **Sphericity Assumed**. If the sphericity assumption is violated, you should examine the tests provided in either the row titled **Greenhouse-Geisser** or **Huynh-Feldt**, which provide tests corrected for your assumption violations.
    - *Tests of Within-Subjects Contrasts*. Provides the results of polynomial contrasts among your within-subject conditions. These can provide you with some information about the specific nature of the relation between your factor and the DV. However, the results will be meaningless unless your groups have some type of ordinal relation with each other.
    - *Tests of Between-Subjects Effects*. This section is not typically examined when performing a one-way within-subjects ANOVA.
  - An example of how you would report the results of a repeated measures analysis that does not use a correction for violation of sphericity would be "There was a significant relation between the testing environment and performance ( $F[2,68] = 5.32, p = .007$ )."

An example of how you would report the results of a repeated measures analysis that uses a correction for violation of sphericity would be "There was a significant relation between the time of measurement and depression ( $F[2.55,63.8] = 3.48, p = .04$  using a Geisser-Greenhouse correction). The corrections will usually give you non-integer degrees of freedom, so don't worry if you see this.

An example of how you would report the results of a within-subjects test using a multivariate test would be "There was a significant relation between the body part being tested and sensitivity ( $F[3,330] = 4.3, p = .005$  using Wilks' Lambda).

## 5.3 Specific group comparisons

- If you find that your within-subjects factor has a significant influence on your DV, you will need to compare the means of your groups to determine the exact nature of the effect. As with other types of ANOVA you can do this using preplanned contrasts, post-hoc comparisons, or by using nonstatistical methods.

### 5.3.1 Preplanned contrasts

- You should perform preplanned contrasts when you are interested in testing one or more specific comparisons among your group means.
- To perform a preplanned contrast based on a one-way repeated measures ANOVA in SPSS you would take the following steps.
  - Define the contrast weight vector for the comparison of interest. Information about how to do this is provided in Section 3.3.1.
  - SPSS requires that you simultaneously test a number of contrasts equal to the degrees of freedom in your IV. You must therefore come up with an additional number of contrasts to fulfill this requirement. The contrasts must all be valid (so the sum of the weights must equal zero) and they must all test different comparisons among your means.
  - Contrasts in repeated measures ANOVA cannot be performed using interactive mode. You must therefore open a syntax window to enter the appropriate code.
  - Type **MANOVA** followed by a list of the variables that represent the different groups of your within-subjects IV.
  - On the next line type **/TRANSFORM** followed by a list of the variables that represent the different groups of your within-subjects IV in parentheses.
  - On the same line, type **=SPECIAL(**
  - Type the number 1 followed by a space a number of times equal to the number of groups in your within-subjects IV. Put a comma at the end. It is sort of like you are defining a contrast weight vector that has a weight of 1 for each group. Your contrasts won't run if you don't include this vector of 1s in your **TRANSFORM** statement.
  - Type out all of your contrast weight vectors, with each weight separated by a space and each vector separated by a comma. Put a right parenthesis at the end of your last vector instead of a comma.
  - On the next line type **/RENAME=** followed by the names that you want to give to each of your contrasts in quotes, each separated by a space. The results corresponding to the set of 1s are not of theoretical interest so type "skip" as the name of the first contrast.
  - On the next line type **/NOPRINT=SIGNIF(MULTIV)** . Don't forget the period at the end.
  - Select the entire block of syntax.
  - Choose **Run** → **Selection**, or click the button with the arrow on it.
- The syntax you run should look something like the following.

```
MANOVA time1 time2 time3
/TRANSFORM(time1 time2 time3)=SPECIAL(1 1 1, 1 -2 1, 0 1 -1)
/RENAME="skip" "2 versus 1,3" "2 versus 3"
/NOPRINT=SIGNIF(MULTIV).
```

- After running the syntax, SPSS will provide information about the sum of squares associated with each contrast as well as an F statistic and p-value testing whether the value of each contrast is significantly different from zero. The degrees of freedom for the contrasts will all be the same, and are listed at the top of the output. There is no theoretical interpretation of the test of the contrast labelled **skip**.

### 5.3.2 Post-hoc comparisons

- You should use post-hoc comparisons when you want to test whether there are differences between pairs of group means in order to understand the nature of a significant ANOVA.
- To perform post-hoc comparisons among the groups in a repeated-measures ANOVA in SPSS you would take the following steps.
  - Perform a repeated-measures ANOVA using the steps described in Section 5.2, but do not press the **OK** button at the end.
  - Click the **Options** button.
  - Move the main effect of interest to the **Display Means for** box.
  - Check the box next to **Compare main effects**.
  - Set the **Confidence interval adjustment** pulldown menu to the type of post-hoc analysis you want to perform.
  - Click the **Continue** button.
  - Click the **OK** button.
- Requesting post-hoc analyses will add the following sections to your ANOVA output.
  - *Estimates*. Reports the means for each of your groups. It also provides the standard error of the mean, as well as a 95% confidence interval around the mean.
  - *Pairwise Comparisons*. Reports post-hoc comparisons testing the difference between each pair of factor levels. It also provides a confidence interval around the difference score.

### 5.3.3 Nonstatistical methods

- You might choose to simply display tables or graphs of the group means instead of testing for statistical differences.
- You can obtain means or graphs of the cell means from a one-way repeated-measures ANOVA in SPSS using the same procedures that you would for a one-way between-subjects ANOVA. See Section 3.3.3 for specific details.

## 5.4 Friedman test

- The *Friedman test* is a nonparametric alternative to a one-way within-subjects ANOVA that does not require that your DV be normally distributed within each group and does not require that you have sphericity.
- We typically use the Friedman test when we have evidence that the DV does not have a normal distribution. People don't typically use it for violations of sphericity since there are established ways to simply correct the results of a repeated measures ANOVA for these violations.
- The Friedman test will tell you if there are any significant differences among the medians of two or more groups. It is an extension of the "Binomial sign test," a nonparametric test for comparing two within-subject groups that we do not cover in these notes.
- To perform a Friedman test by hand you would take the following steps.
  - For each subject, rank order the groups of the IV. It doesn't matter whether high values on the original IV correspond to high or low ranks - the results will be the same whichever way you do it. If two or more groups are tied, assign all of tied groups the average of the tied ranks.
  - Record these ranks in a new set of variables, one for each group in the original within-subject IV.

- Compute the test statistic

$$\chi^2 = \frac{12}{nk(k+1)} \left[ \sum_{i=1}^k \left( \sum_{j=1}^n R_{ij} \right)^2 \right] - 3n(k+1), \quad (5.2)$$

where  $k$  is the number of groups in your IV,  $n$  is the number of subjects, and  $R_{ij}$  is the rank given to group  $i$  for subject  $j$ .

- P-values for the  $\chi^2$  statistic can be drawn from the chi-square distribution with  $k - 1$  degrees of freedom. A significant test statistic means that at least two of the medians are significantly different from each other.
- After you obtain a significant result from the Friedman test, you can perform a series of Wilcoxon Signed-Rank tests as post-hoc comparisons to determine the exact nature of the relation. You may want to apply a Bonferroni or Sidak correction to your  $\alpha$  to prevent inflation of the experimentwide error rate. Alternatively, you can treat the Friedman test as an omnibus test. In this case you would not apply any corrections to your  $\alpha$ , but you would only perform the individual group comparisons if you get a significant result from the Friedman test.
- To perform a Friedman test in SPSS you should take the following steps.
  - Choose **Analyze** → **Nonparametric Tests** → **K Related Samples**.
  - Move the variables corresponding to the different levels of your within-subjects IV to the **Test Variables** box.
  - Click the **OK** button.
- The resulting output will contain the following sections.
  - *Ranks*. Reports the mean rank for each level of your within-subjects IV.
  - *Test Statistics*. Reports the sample size as well as the chi-square, degrees of freedom, and p-value associated with the Friedman test.

## 5.5 Cochran's Q test

- You can use *Cochran's Q test* to determine if there is a relation between a within-subjects factor and a dichotomous DV. It is similar to the chi-square test of independence, except that it is used in within-subjects designs instead of between-subjects designs, and Cochran's Q requires that your DV only has two levels.
- As an example of when you might use Cochran's Q test, you might want to see how people feel about different ways of handling pain. You could give people a list of different ways of dealing with pain (medicine, meditation, and acupuncture) and ask them whether they believed each item on the list could soothe pain. You could then use Cochran's Q test to see if there were any differences between their evaluations of the different methods. In this case the categorical IV is method being evaluated, and the categorical DV is the decision of whether they thought the method would soothe pain or not. This is clearly a within-subjects design since the same people are evaluating each of the three methods.
- To perform Cochran's Q test by hand you would take the following steps.
  - Your data set should be organized such that the subject is the unit of analysis and the response within each level of the IV is stored in a separate variable. These variables should have a value of 0 when the response falls in one category of the DV and a value of 1 when the response falls in the opposite category. It doesn't matter which category is 0 and which is 1, since the results will be the same either way you do it.

- Determine how many times category 1 appeared within each subject. This is the total number of 1s in each row, so we use  $R_i$  to refer to the total for subject  $i$ .
- Determine how many times category 1 appeared within each level of the IV. This is the total number of 1s in each column, so we will use  $C_j$  to refer to the total for the  $j$ th level of the IV.
- Determine how many times category 1 appeared in the entire study. This is the total number of 1s in the data set, so we will refer to this as  $T$ .
- Compute the test statistic

$$Q = \frac{(k - 1) \left[ k \left( \sum_{j=1}^k C_j \right) - T^2 \right]}{kT - \sum_{i=1}^n R_i}, \quad (5.3)$$

where  $k$  is the total number of groups in the IV,  $n$  is the total number of subjects, and  $R_i$ ,  $C_j$ , and  $T$  are defined above.

- The p-value for this statistic can be taken from the chi-square distribution with  $k - 1$  degrees of freedom. A significant statistic indicates that at least two levels of the IV have different proportions in each category.
- To perform Cochran's Q test in SPSS you should take the following steps.
  - Choose **Analyze** → **Nonparametric Tests** → **K Related Samples**.
  - Move the variables representing the responses for the different levels of the IV to the **Test Variables** box.
  - Uncheck the box next to **Friedman**.
  - Check the box next to **Cochran's Q**.
  - Click the **OK** button.
- This analysis will produce the following sections of output.
  - *Frequencies*. Provides the number of times each response was observed in each of the levels of the DV.
  - *Test Statistics*. Provides the total sample size as well as the test statistic, degrees of freedom, and p-value of Cochran's Q test.

## Chapter 6

# Testing Multiple Within-Subjects Factors

### 6.1 Introduction

- The procedures described in this chapter allow you to test the joint effect of two or more categorical IVs on a continuous DV when at least one of the IVs varies within subjects.
- We will start with a discussion of two-way within-subjects ANOVA, which is used when you want to predict a continuous DV from two within-subjects IVs and their interaction. We will then extend the basic points developed in that section and describe how you can use ANOVA to test the joint effects of any number of within-subjects IVs on a DV.
- We will end with a discussion of how you can test models that include both between-subjects and within-subjects IVs. In such models you can look at the main effects of all of the IVs, interactions among between-subjects factors, interactions among within-subjects factors, and interactions involving both both between-subjects and within-subjects factors.

### 6.2 Two-way within-subjects ANOVA

- A two-way within-subjects ANOVA is used to examine the joint effect of two within subjects factors on a continuous DV. It allows you to detect main effects of each of the factors as well as their interaction.
- In this section, we will refer to the two categorical IVs as *Factor A* and *Factor B*. We will use the symbol  $a$  to refer to the number of levels within Factor A, and the symbol  $b$  to refer to the number of levels within Factor B. We will also assume that you have a factorial combination of your IVs. This means that each subject is measured under every possible combination of the levels of your two factors.
- The theoretical interpretations of the tests in a two-way within-subjects ANOVA are the same as the interpretations of the tests in between-subjects ANOVA.
  - The test of each main effect represents the unique influence of the corresponding factor on the DV.
  - The test of the interaction represents the degree to which the influence of one factor on the DV depends on the level of the second factor.
- Just as in a one-way within-subjects design, there are actually multiple ways to test your effects in a multifactor within-subjects design. We will again focus on using repeated-measures ANOVA since it is the most commonly used method and has strong parallels to the within-subjects t test.



- The first thing that we need to do is determine what effects we want to include in our statistical model. We will of course want to include the main effect of Factor A, the main effect of Factor B, as well as the  $A \times B$  interaction, since these are the effects we are interested in theoretically. In addition to these, we must consider the influence of subject differences on the DV. We did not need to do this in between-subjects ANOVA because each subject was only observed under a single combination of the factors. Since we have several observations from each subject in a within-subjects ANOVA, the assumption of independence requires that our statistical model account for any variability due to subject effects. This means that we need to include the main effect of subject as well as all of the possible interactions between subject and the other effects in the model.

We must therefore include the following effects in our statistical model.

- A (differences between the levels of Factor A)
  - B (differences between the levels of Factor B)
  - S (individual differences among the subjects)
  - $A \times B$  (the extent to which the effect of Factor A varies across the levels of Factor B)
  - $A \times S$  (the extent to which the effect of Factor A varies across subjects)
  - $B \times S$  (the extent to which the effect of Factor B varies across subjects)
  - $A \times B \times S$  (the extent to which the  $A \times B$  interaction effect varies across subjects)
- The next thing that we need to do is compute the mean squares associated with each of the effects in the statistical model. You can compute these using the same equations as you would for between-subjects ANOVA if you organize your data set so that the observation (and not the subject) is the unit of analysis.
  - Once you have obtained the mean squares, you can then determine the test statistics for your theoretically interesting effects (those not involving the subjects factor). The way you do this, however, is different than the way you would in a between-subjects design. Recall that the general form of a test statistic in ANOVA is

$$F = \frac{MS_{\text{effect}}}{MS_{\text{error}}}. \quad (6.1)$$

We use this same form for our tests in a multifactor between-subjects ANOVA, except that we define error in a different way. In a between-subjects ANOVA, the error term represents variability in the scores that is not explained by the statistical model. Since each subject only provides one observation, the error actually represents individual differences that are unrelated to your factors. In fact, if we included subject as a factor in a between-subjects model, you would find that its mean square would be exactly equal to the MSE from a model that does not include subject as a factor.

- In a within-subjects ANOVA, however, it doesn't really make sense to use the effect of subject as our error. Recall that one of the main reasons to use a within-subjects design is that variability due to individual differences doesn't effect our results because we have the same subjects in each of our cells. It is still possible to see random error in our results, but only if the individual differences have different effects in each of our cells. The extent to which subject differences vary by the level of a particular effect is captured by the interaction between subject and that effect, so it is this interaction that is actually what we should use as the error when computing our test statistic.
- Since the way that individual differences vary across the levels of one effect is not necessarily the same as the way individual differences vary across the levels of a second effect, we cannot expect to use the same error for each of our tests. Instead, the appropriate error term for a particular effect is always the interaction between subjects and that effect.
- Figure 6.1 summarizes the computations for the tests in a two-way repeated-measures ANOVA. We typically do not compute test statistics for the main effect of subjects or any interactions involving the subjects factor, since they represent individual variability and are not of theoretical interest.

Figure 6.1: ANOVA table for a two-way repeated-measures design.

Source	SS	df	MS	Test
S	$SS_S$	$n - 1$	$MS_S = \frac{SS_S}{n-1}$	$F = \frac{MS_A}{MS_{SA}}$
A	$SS_A$	$a - 1$	$MS_A = \frac{SS_A}{a-1}$	
S×A	$SS_{SA}$	$(n - 1)(a - 1)$	$MS_{SA} = \frac{SS_{SA}}{(n-1)(a-1)}$	$F = \frac{MS_B}{MS_{SB}}$
B	$SS_B$	$b - 1$	$MS_B = \frac{SS_B}{b-1}$	
S×B	$SS_{SB}$	$(n - 1)(b - 1)$	$MS_{SB} = \frac{SS_{SB}}{(n-1)(b-1)}$	$F = \frac{MS_{AB}}{MS_{SAB}}$
A×B	$SS_{AB}$	$(a - 1)(b - 1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	
S×A×B	$SS_{SAB}$	$(n - 1)(a - 1)(b - 1)$	$MS_{SAB} = \frac{SS_{SAB}}{(n-1)(a-1)(b-1)}$	

- The p-values for test statistics computed in a repeated-measures ANOVA can be drawn from the F distribution with numerator degrees of freedom equal to the df associated with the effect of interest and denominator degrees of freedom equal to the df associated with the interaction of subject with that effect.
- An interesting way of thinking about the tests in repeated-measures ANOVA is that the denominator for a given effect represents the extent to which that effect varies from subject to subject.

### 6.3 Multifactor within-subjects ANOVA

- Multifactor within-subjects ANOVA is conceptually the same as two-way within-subjects ANOVA except that you can examine any number of within-subjects factors. In this section we will assume that you are using repeated-measures ANOVA to analyze the effect of your within-subjects factors.
- You can choose exactly which main effects and interactions among your IVs that you want to include in the statistical model, with the restriction that you must include any lower-order effects that are embedded within any higher-order effects present in your model. You will also need to include the main effect of subject, as well as the interaction between the subject factor and any of the other effects you decided to include in your model.
- The test statistics for your effects are computed the same way that they were in a two-way repeated-measures ANOVA. Specifically, the numerator for the F statistic should be the mean square for the effect of interest while the denominator should be the mean square of the interaction between the subject factor and the effect of interest. The p-values can be drawn from the F distribution with numerator degrees of freedom equal to the df associated with the effect of interest and denominator degrees of freedom equal to the df associated with the interaction of subject with that effect.
- You don't typically report a test of the overall model in a multifactor repeated-measures ANOVA since each effect is tested against a different error term. Each effect is conceptually explaining a different type of variability.
- Multifactor repeated-measures ANOVA assumes that your observations are independent except for effects included in your statistical model, that your DV is normally distributed within each cell of your design, and that your variance-covariance matrix has sphericity. If your data violates the assumption

of sphericity you can either apply corrections to your degrees of freedom and p-values as discussed in Section 5.2, or analyze your data using a different procedure such as MANOVA or HLM.

- Although our theoretical discussions of multifactor repeated-measures ANOVA have so far assumed that your data set was organized such that the observation is the unit of analysis, most statistical programs require that your data set be organized such that subject is the unit of analysis. In this case, you would want to use different variables to hold the value of the DV within each combination of your within-subjects factors. For example, let us say that you wanted to perform a repeated-measures ANOVA on a study that examined the effect of the type of stimulus (either a picture or a word) and the valence of the stimulus (either positive, neutral, or negative) on the reaction time to the stimulus. In this case your data set should have six different variables, each one recording the reaction time of the subject to a particular combination of valence and type.
- To perform a multifactor repeated-measures ANOVA in SPSS you would take the following steps.
  - Your data set should be organized such that subject is the unit of analysis and the values of the DV at different combinations of your within-subjects factors are stored in different variables.
  - Choose **Analyze** → **General linear model** → **Repeated measures**.
  - Next you define the within-subject factor. For each within-subjects factor in your design you should take the following steps.
    - \* Enter the name of the factor in the **Within-Subject Factor Name** box.
    - \* Enter the number of levels the factor has in the **Number of Levels** box.
    - \* Click the **Add** button.
    - \* Click the **Define** button.
  - The next thing you will need to do is identify which variables correspond to the particular combinations of your within-subject factors. The Within-Subject Variables box in the next window will contain something that looks like Figure 6.2.

Figure 6.2: The Within-Subject Variables Box in SPSS

.\_ (1,1)  
.\_ (1,2)  
.\_ (2,1)  
.\_ (2,2)  
.\_ (3,1)  
.\_ (3,2)

The numbers on the right-hand side correspond to the levels of your factors. The first number corresponds to the level of your first factor, while the second number corresponds to the level of your second factor. The order of your factors will be listed above the Within-Subjects Variables box. For each combination represented in this box you should select the corresponding variable in the box on the left-hand side and then press the arrow button next to the Within-Subjects Variables box. It doesn't matter what level of a variable you decide to associate with a number on this screen, but you must be sure that you are consistent. For example, the variable you put in the (3,1) slot should have the same level of the first factor as the variable in the (3,2) slot, and the variable you put in the (1,2) slot should have the same level of the second factor as the variable you put in the (2,2) slot.

- To obtain tables containing marginal or cell means you should take the following steps.
  - \* Click the **Options** button.
  - \* Move any main effects for which you want to obtain the marginal means to the **Display Means for** box.

- \* Move any interaction effects for which you want to obtain the means within the different cells of that effect to the **Display Means for** box.
  - \* Click the **Continue** button once you have selected all of the tables you want to see.
- To obtain graphs of the marginal or cell means you should take the following steps.
  - \* Click the **Plots** button.
  - \* To plot a main effect, move the factor to the **Horizontal Axis** box and click the **Add** button.
  - \* To plot a two-way interaction, move the first factor to the **Horizontal Axis** box, move the second factor to the **Separate Lines** box, and click the **Add** button.
  - \* To plot a three-way interaction, move the first factor to the **Horizontal Axis** box, move the second factor to the **Separate Lines** box, move the third factor to the **Separate Plots** box, and click the **Add** button.
  - \* Click the **Continue** button once you have added all of the plots you want to see.
- Click the OK button.
- The output of this analysis will include the following sections.
  - *Within-Subjects Factors.* Tells you what variables represent each combination of your factors.
  - *Multivariate Tests.* Contains the first test of your main effects and interactions, making use of multivariate analysis. The test statistics provided here can be interpreted in the same way as described in Section 5.2.
  - *Mauchly's Test of Sphericity.* Provides a test of the sphericity assumption for each of your within-subject terms (including both main effects and interactions).
  - *Tests of Within-Subjects Effects.* Contains the repeated-measures tests of your within-subjects terms. If the assumption of sphericity is satisfied, you examine the test provided in the row labelled **Sphericity Assumed**. If the sphericity assumption is violated, you should examine the tests provided in either the row titled **Greenhouse-Geisser** or **Huynh-Feldt**, which provide tests corrected for your assumption violations.
  - *Tests of Within-Subjects Contrasts.* Provides the results of polynomial contrasts among your within-subject conditions. These will have no meaning unless your levels have an ordinal relation.
  - *Tests of Between-Subjects Effects.* This section is not typically examined when performing a multifactor within-subjects ANOVA.
  - *Estimated Marginal Means.* Contains a table for each effect for which you requested means. The rows in the tables will correspond to a cell within the effect. The different columns provide the mean, standard error of the mean, and the boundaries for a 95% confidence interval around the mean for observations within that cell.
  - *Profile Plots.* Displays any graphs of the marginal or cell means that you requested.

## 6.4 Specific group comparisons in multifactor within-subjects ANOVA

- If your ANOVA indicates a significant influence of one or more of your effects, you will need to compare the means of your groups to determine the exact nature of the effects. As with other types of ANOVA you can do this using preplanned contrasts, post-hoc comparisons, or by using nonstatistical methods.

### 6.4.1 Preplanned contrasts

- You should perform preplanned contrasts when you are interested in testing one or more specific comparisons among your group means.
- To use SPSS to test contrasts in a multifactor within-subjects ANOVA you would take the following steps.

- Define the contrast weight vector for the comparison of interest. Basic information about how to do this is provided in Section 3.3.1. However, the way you do this for a within-subjects design is slightly more complex. Specifically, you must define your contrast in terms of the entire collection of variables that you use to define your within subjects conditions. For example, let us say you conducted a study using the design presented in Figure 6.4.1.

Figure 6.3: Example within-subjects design

		Factor B		
		Level 1	Level 2	Level 3
Factor A	Level 1	A1B1	A1B2	A1B3
	Level 2	A2B1	A2B2	A2B3

In this case, you would need six different variables to hold the results, one for each cell of the design. Preplanned contrasts for this design would need to be defined in terms of these six variables.

- You must simultaneously test a number of contrasts equal to one less than the number of within-subject conditions in your study. This will also be equal to the number of variables in your **MANOVA** statement - 1. You can choose whatever contrasts you want, as long as none of them are completely collinear with the others.
  - Open a syntax window to enter the appropriate code.
  - Type **MANOVA** followed by a list of the entire collection of variables representing the within-subjects conditions.
  - On the next line type **/TRANSFORM** followed by the same list of variables in parentheses.
  - On the same line, type **=SPECIAL(**
  - Type the number 1 a number of times equal to the number of variables in your **MANOVA** statement. Follow each 1 with a space and put a comma at the end. It is sort of like you are defining a contrast weight vector that has a weight of 1 for each group. Your contrasts won't run if you don't include this vector of 1s in your **TRANSFORM** statement.
  - Type out all of your contrast weight vectors, with each weight separated by a space and each vector separated by a comma. Type a right parenthesis after the last contrast vector instead of a comma. You may have a large number of vectors in a multifactor within-subjects design, so you might decide to put each vector on a separate line so that your program is easier to read.
  - On the next line type **/RENAME=** followed by the names that you want to give to each of your contrasts in quotes, each separated by a space. The results corresponding to the set of 1s are not of theoretical interest so type "skip" as the name of the first contrast.
  - On the next line type **/NOPRINT=SIGNIF(MULTIV) .** Don't forget the period at the end.
  - Select the entire block of syntax.
  - Choose **Run** → **Selection**, or click the button with the arrow on it.
- The syntax you run should look something like the following.

```
MANOVA a1b1 a1b2 a1b3 a2b1 a2b2 a2b3
/TRANSFORM(a1b1 a1b2 a1b3 a2b1 a2b2 a2b3)=SPECIAL(
  1 1 1 1 1 1,
  1 1 1 -1 -1 -1,
  1 0 -1 1 0 -1,
  0 1 -1 0 1 -1,
  1 0 -1 -1 0 1,
```

```

0 1 -1 0 -1 1)
/RENAME="skip" "A1-A2" "B1-B3" "B2-B3" "B1-B3 x A" "B2-B3 x A"
/NOPRINT=SIGNIF(MULTIV).

```

- After running the syntax, SPSS will provide information about the sum of squares associated with each contrast as well as an F statistic and p-value testing whether the value of each contrast is significantly different from zero. The degrees of freedom for the contrasts will all be the same, and are listed at the top of the output. There is no theoretical interpretation of the test of the contrast labelled **skip**.

## 6.4.2 Post-hoc comparisons

- You use post-hoc comparisons when your ANOVA reports a significant main effect or interaction and you want to explore the means to determine the nature of the effect.
- Just as in multifactor between-subjects ANOVA, we typically want to examine the estimated marginal means instead of the raw means whenever we perform our post-hoc analyses. These means control for the other factors in the model, which corresponds to the way the main effects and interactions are tested.
- To use SPSS to perform post-hoc comparisons for the main effects in a multifactor within-subjects ANOVA you would take the following steps.
  - Perform the appropriate repeated-measures ANOVA using the steps described in Section 6.3, but do not press the **OK** button at the end.
  - Click the **Options** button.
  - Move the main effects of interest to the **Display Means for** box.
  - Check the box next to **Compare main effects**.
  - Set the **Confidence interval adjustment** pulldown menu to the type of post-hoc analysis you want to perform.
  - Click the **Continue** button.
  - Click the **OK** button.
- Requesting post-hoc analyses will add the following sections to your ANOVA output.
  - *Estimates*. Reports the group means for each of the main effects you wanted to examine. It also provides the standard error of the mean, as well as a 95% confidence interval around the mean.
  - *Pairwise Comparisons*. Reports post-hoc comparisons testing the difference between each pair of factor levels for each main effect you wanted to examine. It also provides a confidence interval around the difference score.
- To perform post-hoc analyses for interactions in SPSS you should take the following steps.
  - Use the steps described in Section 6.3 to set up your multifactor ANOVA, but do not press **OK** at the end.
  - Click the **Paste** button. SPSS will copy the syntax you need to run the multifactor ANOVA you defined into a syntax window.
  - Insert a blank line after the **/CRITERIA = ALPHA(.05)** statement.
  - Type **/EMMEANS =** on the blank line.
  - On the same line, type **TABLES** followed by the interaction you want to explore in parentheses.
  - On the same line, type **COMPARE** followed by the name of one of your IVs in parentheses. SPSS will compare the least squares means of this variable within each combination of the levels of the other IVs in the interaction.

- On the same line, type **ADJ** followed by the name of the adjustment method for  $\alpha$  that you want to use. Acceptable values are **LSD** (which does not adjust your  $\alpha$ ), **BONFERRONI**, or **SIDAK**.
  - Select the entire block of syntax.
  - Choose **Run** → **Selection**, or click the button with the arrow on it.
- The code you run should look something like the following.

**GLM**

```

a1b1 a1b2 a1b3 a2b1 a2b2 a2b3
/WSFACTOR = A 2 Polynomial B 3 Polynomial
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/EMMEANS = TABLES(A*B) COMPARE(A) ADJ(SIDAK)
/DESIGN = A B A*B .

```

- Using either of these post-hoc methods will add the following sections to your ANOVA output.
  - *Estimates.* Reports the least squares means within each cell of your interaction. It also provides the standard error of the mean, as well as a 95% confidence interval around the mean.
  - *Pairwise Comparisons.* Reports the post-hoc comparisons you requested. For a main effect, SPSS tests the difference between each pair of factor levels. For an interaction, SPSS tests the difference between each pair of levels for one factor within each combination of the other factors in the design.

### 6.4.3 Nonstatistical methods

- You might choose to simply display tables or graphs of the group means instead of testing for statistical differences.
- You can obtain means or graphs of the cell means from a multifactor repeated-measures ANOVA in SPSS using the same procedures that you would for a multifactor between-subjects ANOVA. See Section 4.5.3 for specific details.

## 6.5 ANOVA models with both between-subjects and within-subjects factors

- It is possible to conduct an ANOVA including both between- and within-subject factors in the same analysis. In this case you can include any or all of the following types of effects in your statistical model.
  - Main effects of between-subject factors.
  - Main effects of within-subject factors.
  - Interactions solely involving between-subject factors.
  - Interactions solely involving within-subject factors.
  - Interactions involving both between- and within-subject factors.
- A study that includes both between- and within-subject factors is sometimes referred to as a *mixed design*. Similarly, an ANOVA that includes both between- and within-subject effects is sometimes referred to as a *mixed ANOVA*.

- Before we discuss how to compute tests in a mixed ANOVA we must first discuss the notation used to refer to different effects in a mixed design. To understand the notation you must learn about two ways of combining the levels of your variables when collecting data.
  - If you choose to *cross* a pair of factors, you make observations at each possible combination of the levels from the two factors. We can see an example of two crossed factors in Figure 6.4.

Figure 6.4: Example of a crossed design.

		Factor A		
		$a_1$	$a_2$	$a_3$
Factor B	$b_1$	x	x	x
	$b_2$	x	x	x
	$b_3$	x	x	x
	$b_4$	x	x	x
	$b_5$	x	x	x
	$b_6$	x	x	x

The different columns correspond to different levels of Factor A, and the different rows correspond to different levels of Factor B. To cross Factor A and Factor B, you must make observations in every cell where you see an “x”. As you can see, each replication of this design takes a total of 18 observations (equal to the number of levels in Factor A times the number of levels in Factor B), each one pairing a single level of Factor A with a single level of Factor B. In a crossed design you can test for main effects of both factors as well as the interaction between the two factors. None of these effects are confounded with each other.

Researchers typically use the notation  $\mathbf{A} \times \mathbf{B}$  to refer to the crossing of Factor A and Factor B. Notice that crossed designs are symmetric: If Factor A is crossed with Factor B, then Factor B must necessarily be crossed with Factor A.

- If you choose to *nest* one of the factors underneath another, each level of one factor is paired with a single level of the other factor. We can see an example of nested factors in Figure 6.5

Figure 6.5: Example of a nested design.

		Factor A		
		$a_1$	$a_2$	$a_3$
Factor B	$b_1$	x		
	$b_2$	x		
	$b_3$		x	
	$b_4$		x	
	$b_5$			x
	$b_6$			x

Again, the different columns correspond to different levels of Factor A, and the different rows correspond to different levels of Factor B. This time you can see that every level of Factor B is paired with only a single level of Factor A. We would therefore say that Factor B is *nested under* Factor A (or equivalently, that Factor B is *nested within* Factor A). Each replication of this design would take a total of 6 observations, which would be equal the number of levels of Factor B. In a nested design you can test for main effects of either factor, although these tests are confounded with each other. However, the interpretation becomes clearer if you make the assumption that Factor B is a *random effect*. This means that the specific levels of this factor observed in the



study were randomly chosen from an underlying distribution of possible levels. This means that the test of Factor B will not be theoretically interesting, but it also allows you to directly interpret the main effect for Factor A. This is because we would expect the average influence of a random variable to be the same across the different conditions of Factor A, so we can ignore the relation between Factor A and Factor B. There is no way to test for an interaction between two nested factors.

Researchers typically use the notation  $\mathbf{B(A)}$  to refer to Factor B being nested under Factor A. As a rule, you never refer to a factor that is nested under something else without reporting the nesting. So you would never just talk about the effect of “B” in a model where Factor B was nested under Factor A - you would always say “B(A).”

Notice that while each level of Factor B is associated with a single level of Factor A, each level of Factor A is associated with multiple levels of Factor B. Nested designs are **not** symmetric: When Factor B is nested under Factor A, Factor A will typically not be nested under Factor B.

- You can use these terms to describe the relations among the different variables in a mixed design.
  - Assuming that you have a factorial design, each of your IVs is crossed with every other IV in the study.
  - The subject factor is nested under each of your between-subject IVs. If you have more than one between-subject IV in your design, you would actually say that the subject factor is nested under the interaction of all of your between-subject IVs, since each subject is only in a single combination of the levels of your between-subject IVs.
  - The subject factor is crossed with each of your within-subject IVs. If your within-subject IVs are themselves all crossed with each other (which would be the case in a factorial design) you could also say that the subject factor is crossed with the interaction of all your within-subject IVs.
- With this in mind, we use the following notation to refer to the various effects in a mixed ANOVA.
  - The main effects of your IVs are simply represented by the factor name, whether the IV is manipulated between-subjects or within-subjects. So the main effect of a within-subject time factor is simply called “time,” while the main effect of a between-subject gender factor is simply called “gender.” This is the same notation that we’ve been using up until now.
  - Interactions among your IVs are represented by a listing of the factors involved in the interaction separated by “ $\times$ ” symbols. It doesn’t matter whether the IVs are manipulated between-subjects or within-subjects, and the order of the variables doesn’t matter. So the interaction between a within-subject time factor and the between-subject gender factor could be called either “time  $\times$  gender” or “gender  $\times$  time.”
  - The effect of subject is nested under the interaction of all of your between-subject IVs, so you represent this as “S(*interaction*),” where *interaction* is the interaction of all of your between-subject IVs. So the subject effect in a design that included the between-subject factors race and gender would be “S(race  $\times$  gender).”
  - Interactions between the subject factor and your within-subject IVs can be treated just like normal interactions, except that you need to indicate that the subject factor is nested under your between-subject IVs. So the interaction between subject and the within-subject factor time in a model that also includes the between-subject factors gender and race would be represented by “S(gender  $\times$  race)  $\times$  time.”
- Now that we have a way of referring to the different effects in a mixed ANOVA, we can discuss how we can use these to compute test statistics. We can compute a test statistic for each of the theoretically interesting effects (those that don’t involve the subject factor). The test statistic will always take the form of

$$F = \frac{MS_{\text{effect}}}{MS_{\text{error}}}. \quad (6.2)$$

$MS_{\text{effect}}$  represents the variability in the DV that can be explained by the effect, while  $MS_{\text{error}}$  represents the random variability within that effect. What you should use for  $MS_{\text{error}}$  depends on the type of effect that you are looking at.

- If you are testing either the main effect of a between-subjects IV or an interaction that only involves between-subjects IVs, then  $MS_{\text{error}}$  is the mean square for the subject factor.
  - If you are testing either the main effect of a within-subjects IV or an interaction that only involves within-subjects IVs, then  $MS_{\text{error}}$  is the mean square for the interaction between the subject factor and the effect you are testing.
  - If you are testing an interaction that involves both between- and within-subject IVs, then  $MS_{\text{error}}$  is the mean square for the interaction between the subject factor and all of the within-subject IVs involved in the interaction effect.
- As an example, Figure 6.6 below reports how you would compute the tests in a mixed ANOVA that has two between-subjects variables (race and gender) and two within-subject variables (time and measure).
  - The interpretation of the tests in a mixed ANOVA is exactly the same as in a between-subjects ANOVA. A significant main effect indicates that the mean values of at least two of your factor levels are significantly different. A significant two-way interaction means that the difference between the levels of one of your factors varies depending on the level of a second factor. A significant three-way interaction indicates that the nature of a two-way interaction depends on the level of a third factor.
  - The way you compute the degrees of freedom and mean squares for your theoretically interesting effects is exactly the same as you would if you were running a between-subjects ANOVA. However, the computation of the degrees of freedom and mean squares for the effects involving the subject factor are somewhat more complicated. A thorough discussion of this is beyond the scope of these notes, but you can see Chapter 29 of Neter, et al. (1996) for more information.
  - To perform a mixed ANOVA in SPSS you would take the following steps.
    - Your data set needs to be organized such that subject is the unit of analysis. This means that you will have one variable representing the level of each between-subject IV, and a number of different variables representing the value of the DV for each possible combination of the levels of your within-subjects IVs.
    - Choose **Analyze** → **General linear model** → **Repeated measures**.
    - Next you define the within-subject factor(s). For each within-subject factor you would take the following steps.
      - \* Enter the name of the factor in the **Within-Subject Factor Name** box.
      - \* Enter the number of levels the factor has in the **Number of Levels** box.
      - \* Click the **Add** button.
    - Click the **Define** button.
    - Identify which variables are associated with each combination of your within-subjects conditions as described in Section 6.3.
    - Move any between-subjects IVs to the **Between-subjects factor(s)** box.
    - To obtain tables containing marginal or cell means you should take the following steps.
      - \* Click the **Options** button.
      - \* Move any main effects for which you want to obtain the marginal means to the **Display Means for** box.
      - \* Move any interaction effects for which you want to obtain the means within the different cells of that effect to the **Display Means for** box.
      - \* Click the **Continue** button once you have selected all of the tables you want to see.
    - To obtain graphs of the marginal or cell means you should take the following steps.

Figure 6.6: Tests in an example mixed design.

Effect	Test
race	$F = \frac{MS_{\text{race}}}{MS_{S(\text{race} \times \text{gender})}}$
gender	$F = \frac{MS_{\text{gender}}}{MS_{S(\text{race} \times \text{gender})}}$
time	$F = \frac{MS_{\text{time}}}{MS_{\text{time} \times S(\text{race} \times \text{gender})}}$
measure	$F = \frac{MS_{\text{measure}}}{MS_{\text{measure} \times S(\text{race} \times \text{gender})}}$
race × gender	$F = \frac{MS_{\text{race} \times \text{gender}}}{MS_{S(\text{race} \times \text{gender})}}$
race × time	$F = \frac{MS_{\text{race} \times \text{time}}}{MS_{\text{time} \times S(\text{race} \times \text{gender})}}$
race × measure	$F = \frac{MS_{\text{race} \times \text{measure}}}{MS_{\text{measure} \times S(\text{race} \times \text{gender})}}$
gender × time	$F = \frac{MS_{\text{gender} \times \text{time}}}{MS_{\text{time} \times S(\text{race} \times \text{gender})}}$
gender × measure	$F = \frac{MS_{\text{gender} \times \text{measure}}}{MS_{\text{measure} \times S(\text{race} \times \text{gender})}}$
time × measure	$F = \frac{MS_{\text{time} \times \text{measure}}}{MS_{\text{time} \times \text{measure} \times S(\text{race} \times \text{gender})}}$
race × gender × time	$F = \frac{MS_{\text{race} \times \text{gender} \times \text{time}}}{MS_{\text{time} \times S(\text{race} \times \text{gender})}}$
race × gender × measure	$F = \frac{MS_{\text{race} \times \text{gender} \times \text{measure}}}{MS_{\text{measure} \times S(\text{race} \times \text{gender})}}$
race × time × measure	$F = \frac{MS_{\text{race} \times \text{time} \times \text{measure}}}{MS_{\text{time} \times \text{measure} \times S(\text{race} \times \text{gender})}}$
gender × time × measure	$F = \frac{MS_{\text{gender} \times \text{time} \times \text{measure}}}{MS_{\text{time} \times \text{measure} \times S(\text{race} \times \text{gender})}}$
race × gender × time × measure	$F = \frac{MS_{\text{race} \times \text{gender} \times \text{time} \times \text{measure}}}{MS_{\text{time} \times \text{measure} \times S(\text{race} \times \text{gender})}}$

- \* Click the **Plots** button.
- \* To plot a main effect, move the factor to the **Horizontal Axis** box and click the **Add** button.
- \* To plot a two-way interaction, move the first factor to the **Horizontal Axis** box, move the second factor to the **Separate Lines** box, and click the **Add** button. It doesn't matter whether the factors you choose are between-subject or within-subject factors.
- \* To plot a three-way interaction, move the first factor to the **Horizontal Axis** box, move the second factor to the **Separate Lines** box, move the third factor to the **Separate Plots** box,

and click the **Add** button.

\* Click the **Continue** button once you have added all of the plots you want to see.

- Click the **OK** button.
- The output from this analysis will contain the following sections.
  - *Within-Subjects Factors*. Tells you what variables represent each combination of your within-subjects factors.
  - *Between-Subjects Factors*. Lists how many subjects are in the combination of each of your between-subjects factors.
  - *Multivariate Tests*. Contains multivariate tests of the main effects of your within-subjects factors as well as interactions that contain at least one within-subjects factor. The test statistics provided here can be interpreted in the same way as described in Section 5.2.
  - *Mauchly's Test of Sphericity*. Provides a test of the sphericity assumption for each of your within-subject terms (including the main effects of your within-subject factors as well as interactions involving at least one within-subject factor).
  - *Tests of Within-Subjects Effects*. Contains repeated-measures tests of the main effects of your within-subjects factors as well as interactions that contain at least one within-subjects factor. If the assumption of sphericity is satisfied, you examine the test provided in the row labelled **Sphericity Assumed**. If the sphericity assumption is violated, you should examine the tests provided in either the row titled **Greenhouse-Geisser** or **Huynh-Feldt**, which provide tests corrected for your assumption violations.
  - *Tests of Within-Subjects Contrasts*. Provides the results of polynomial contrasts among your within-subject main effects and interactions. These will have no meaning unless your levels have an ordinal relation.
  - *Tests of Between-Subjects Effects*. Contains tests of the main effects of your between-subjects factors as well as tests of any interactions that only involve between-subjects factors.
  - *Estimated Marginal Means*. Contains a table for each effect for which you requested means. The rows in the tables will correspond to a cell within the effect. The different columns provide the mean, standard error of the mean, and the boundaries for a 95% confidence interval around the mean for observations within that cell.
  - *Profile Plots*. Displays any graphs of the marginal or cell means that you requested.

## 6.6 Specific group comparisons in ANOVA models with both between-subjects and within-subjects factors

- If your ANOVA indicates a significant influence of one or more of your effects, you will need to compare the means of your groups to determine the exact nature of the effects. As with other types of ANOVA you can do this using preplanned contrasts, post-hoc comparisons, or by using nonstatistical methods.

### 6.6.1 Preplanned contrasts

- You should perform preplanned contrasts when you are interested in testing one or more specific comparisons among your group means.
- If you want to use SPSS to compute a preplanned contrast that only involves comparisons among your between-subjects conditions you should take the following steps.
  - Compute a new variable that is equal to the average of all of the variables representing the responses in the different within-subjects conditions.
  - Perform an ANOVA predicting this new variable from just your between-subjects factors.

- Use the techniques to perform contrasts in a multifactor between-subjects ANOVA presented in section 4.5.1.
- If you want to use SPSS to compute a preplanned contrast that only involves comparisons among your within-subjects conditions, you can use the same procedures as you would to test comparisons in a multifactor within-subjects ANOVA. These are presented in section 6.4.1.
- It can be somewhat difficult to test contrasts comparing specific cells within a crossing of a between-subjects factor and a within-subjects factor. The basic procedure to do this in SPSS is to define a within-subjects contrast, and then perform tests of the extent to which the value of this contrast varies across the level of a between-subjects factor.

For example let us say that you had the design illustrated in Figure 6.6.1.

Figure 6.7: Example mixed design

		Within Factor		
		Level 1	Level 2	Level 3
Between Factor	Level 1	<i>B1W1</i>	<i>B1W2</i>	<i>B1W3</i>
	Level 2	<i>B2W1</i>	<i>B2W2</i>	<i>B2W3</i>
	Level 3	<i>B3W1</i>	<i>B3W2</i>	<i>B3W3</i>

In this case, let's assume that we defined our within-subjects contrast to be  $(1 - \frac{1}{2} - \frac{1}{2})$ , which compares the mean of the first level of our within-subjects factor to the average of the other two levels. Let's then say that we wanted to see how this contrast varied across the levels of the between-subjects factor. The final contrast would then test the extent to which the mean of the within-subjects contrast changes across the different levels of the between-subjects factor.

To test these types of contrasts in SPSS you would take the following steps.

- Define the contrast weight vectors for your within-subjects conditions. Basic information about how to do this is provided in Section 3.3.1. You must define your within-subjects contrast in terms of the entire collection of variables that you use to define your within subjects conditions, just as was described in section 6.4.1 for multifactor within-subjects ANOVA.
 

Unlike the within-subjects contrast, your between-subjects contrast should be defined in terms of the levels of a single between-subjects variable. If you want to look at the value of the within-subjects contrast at different levels of the interaction between two or more between-subjects factors, you should combine those factors into a single variable as discussed in the first part of section 4.5.1. While it is possible to run contrasts in a model containing multiple between-subjects variables (as described in chapter 9 of Page, Braver, & MacKinnon, 2003), it will be much easier to interpret the results if you combine all of your factors into a single variable.
- You must simultaneously test a number of within-subjects contrasts equal to one less than the number of within-subject conditions in your study. You can choose whatever contrasts you want, as long as none of them are completely collinear with the others.
- Open a syntax window to enter the appropriate code.
- Type **MANOVA** followed by a list of the entire collection of variables representing the within-subjects conditions.
- On the same line, type **BY**
- For each between-subjects variable in the model, perform the following steps (all on the same line).
  - \* Type the name of the between-subjects variable followed by a left parenthesis.
  - \* Type the minimum value of the variable followed by a comma.

- \* Type the maximum value of the variable followed by a right parenthesis.
  - On the next line type **/TRANSFORM** followed by the list of the variables representing your within-subjects conditions in parentheses.
  - On the same line, type **=SPECIAL(**
  - Type the number 1 followed by a space a number of times equal to the number of within-subjects conditions you have. Type a comma at the end.
  - Type out all of your within-subjects contrast weight vectors, with each weight separated by a space and each vector separated by a comma. Type a right parenthesis at the end of the first contrast vector instead of a comma.
  - On the next line type **/RENAME=** followed by the names that you want to give to each of your contrasts in quotes, each separated by a space. The results corresponding to the set of 1s are not of theoretical interest so type “skip” as the name of the first contrast.
  - On the next line type **/CONTRAST** followed by the name of the between-subjects variable in parentheses.
  - On the next line type **/NOPRINT=SIGNIF(MULTIV)**
  - On the next line type **/DESIGN =** followed by the name of the between-subjects variable. Be sure to put a period at the end of this statement.
  - Select the entire block of syntax.
  - Choose **Run** → **Selection**, or click the button with the arrow on it.
- The syntax you run should look something like the following.

```
MANOVA a1b1 a1b2 a1b3 a2b1 a2b2 a2b3 = cond(1,4)
/TRANSFORM(a1b1 a1b2 a1b3 a2b1 a2b2 a2b3)=SPECIAL(
  1 1 1 1 1 1,
  1 1 1 -1 -1 -1,
  1 0 -1 1 0 -1,
  0 1 -1 0 1 -1,
  1 0 -1 -1 0 1,
  0 1 -1 0 -1 1)
/RENAME="skip" "A1-A2" "B1-B3" "B2-B3" "B1-B3 x A" "B2-B3 x A"
/NOPRINT=SIGNIF(MULTIV)
/DESIGN = cond.
```

- The above analysis will produce a text box containing the output from your contrasts. It will start with overall F statistics testing the extent to which the mean values of the within-subjects contrasts vary across each of your between-subjects factors. It will then be followed by tests specifically examining how the value of the within-subjects contrast differs between individual groups of your between-subjects variable. There is no meaningful interpretation of any of the tests involving the contrast labelled **skip**.

## 6.6.2 Post-hoc comparisons

- You use post-hoc comparisons when your ANOVA reports a significant main effect or interaction and you want to explore the means to determine the nature of the effect.
- As in any ANOVA involving multiple factors, we typically want to examine the estimated marginal means whenever we perform our post-hoc analyses. These means control for the other factors in the model, which corresponds to the way the main effects and interactions are tested.
- To use SPSS to perform post-hoc comparisons for the main effects in a mixed ANOVA (whether they are within-subjects or between-subjects factors) you would take the following steps.

- Perform the appropriate ANOVA using the steps described in Section 6.5, but do not press the **OK** button at the end.
  - Click the **Options** button.
  - Move the main effects of interest to the **Display Means for** box.
  - Check the box next to **Compare main effects**.
  - Set the **Confidence interval adjustment** pulldown menu to the type of post-hoc analysis you want to perform.
  - Click the **Continue** button.
  - Click the **OK** button.
- Requesting post-hoc analyses will add the following sections to your ANOVA output.
    - *Estimates*. Reports the group means for each of the main effects you wanted to examine. It also provides the standard error of the mean, as well as a 95% confidence interval around the mean.
    - *Pairwise Comparisons*. Reports post-hoc comparisons testing the difference between each pair of factor levels for each main effect you wanted to examine. It also provides a confidence interval around the difference score.
  - To perform post-hoc analyses for interactions in SPSS (whether the interactions involve between-subjects factors, within-subjects factors, or both), you should take the following steps.
    - Use the steps described in Section 6.5 to set up your ANOVA, but do not press **OK** at the end.
    - Click the **Paste** button. SPSS will copy the syntax you need to run the ANOVA you defined into a syntax window.
    - Insert a blank line after the **/CRITERIA = ALPHA(.05)** statement.
    - Type **/EMMEANS =** on the blank line.
    - On the same line, type **TABLES** followed by the interaction you want to explore in parentheses.
    - On the same line, type **COMPARE** followed by the name of one of your IVs in parentheses. SPSS will compare the least squares means of this variable within each combination of the levels of the other IVs in the interaction.
    - On the same line, type **ADJ** followed by the name of the adjustment method for  $\alpha$  that you want to use. Acceptable values are **LSD** (which does not adjust your  $\alpha$ ), **BONFERRONI**, or **SIDAK**.
    - Select the entire block of syntax.
    - Choose **Run** → **Selection**, or click the button with the arrow on it.
  - The code you run should look something like the following.

**GLM**

```

ab1 a1b2 a1b3 a2b1 a2b2 a2b3 BY group
/WSEFACTOR = A 2 Polynomial B 3 Polynomial
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/CRITERIA = ALPHA(.05)
/EMMEANS = TABLES(A*group) COMPARE(A) ADJ(SIDAK)
/DESIGN = group.

```

- Using either of these post-hoc methods will add the following sections to your ANOVA output.
  - *Estimates*. Reports the least squares means within each cell of your interaction. It also provides the standard error of the mean, as well as a 95% confidence interval around the mean.
  - *Pairwise Comparisons*. Reports the post-hoc comparisons you requested. For a main effect, SPSS tests the difference between each pair of factor levels. For an interaction, SPSS tests the difference between each pair of levels for one factor within each combination of the other factors in the design.

### 6.6.3 Nonstatistical methods

- You might choose to simply display tables or graphs of the group means instead of testing for statistical differences.
- You can obtain means or graphs of the cell means from a mixed ANOVA in SPSS using the same procedures that you would for a multifactor between-subjects ANOVA. See Section 4.5.3 for specific details.



# References

- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modelling. *Sociological Methods & Research*, 16, 78-117.
- Carmer, S. G. & Swanson, M. R. (1972). An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. *Journal of the American Statistical Association*, 68, 66-74
- Goodnight, J. H. (1980). Tests of hypotheses in fixed effects linear models. *Communications in Statistics: Theory and Methods*, 9, 167-180.
- Moore, D. S., & McCabe, G. P. (2003). *Introduction to the Practice of Statistics* (4th edition). New York, NY: Freeman.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied Linear Statistical Models* (4th edition). Chicago, IL: Irwin.
- Page, M. C., Braver, S. L., & MacKinnon, D. P. (2003). *Levine's Guide to SPSS for Analysis of Variance* (2nd edition). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rencher, A. C., & Scott, D. T. (1990). Assessing the contribution of individual variables following rejection of a multivariate hypothesis. *Communications in Statistics: Simulation and Computation*, 19, 535-553.
- Sheskin, D. J. (2004). *Handbook of Parametric and Nonparametric Statistical Procedures* (3rd edition). Boca Raton, FL: Chapman & Hall.
- Sidak, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62, 626-633.