

Factorial Analysis of Variance  
Statistically Significant Interactions: What's the next step?  
T.C. Oshima and F. McCarty  
Georgia State University

The purpose of this handout is to assist the burgeoning statistician in analyzing (using SPSS) and interpreting the meaning of a statistically significant interaction in the context of factorial analysis of variance (ANOVA). We shall assume that the reader is already familiar with the results obtained when factorial ANOVA is the chosen analytic technique. However, just to be on the safe side, we will review the basics as we go through two examples demonstrating two of the methods that can be used as a follow-up to a statistically significant interaction effect. The two approaches that we will discuss are: 1. tests of simple main effects, and 2. statistical comparison of cell means.

Remember, an interaction effect exists when the effect of one independent variable on the dependent variable depends on the value (level) of some other independent variable included in the study design. In other words, the test of an interaction focuses on the question of whether or not the effect of one factor is the same for each level of the other factor. In either case, our interest shifts to the cell means, not the marginal means (row and column means). Remember, tests of the main effects focus on the marginal means.

When a statistically significant interaction is found we avoid interpreting the main effects and interpret the interaction effect. Interpreting the main effects in the presence of a statistically significant interaction could lead us to erroneous conclusions.

One approach to considering interaction effects calls on the researcher to specify a focal and a moderator independent variable. The focal independent variable (e.g., teaching method, counseling method) can be thought of as the variable whose effect on the dependent variable is thought to be moderated by some other independent variable included in the study (e.g., aptitude, gender). It may be useful to think of the focal independent variable as the variable of primary interest to the researcher. On the other hand, the moderator variable may be thought of as a variable of interest because of its potential impact on the effects of the independent variable that is the primary focus of the study. Another useful distinction that can be applied to the independent variables in a study is the assigned-active distinction. It will often be the case that an assigned (characteristic intrinsic to the participant) independent variable will be designated as the moderator variable and the active (characteristic determined by the researcher) independent variable will be considered the focal variable. In many cases, this designation can be made on conceptual grounds and can be quite useful when considering interaction effects. However, one should not worry if this designation cannot be easily made, because the results of the interaction analysis will be unaffected by the designation. We shall see in our second example, where both independent variables would be considered active independent variables (course length & homework status), that the focal-moderator distinction may not be needed to facilitate interpretation of the interaction. In this case, the groups formed by the combinations of the two independent variables can be meaningfully compared, therefore, the statistical comparison of cell means could be applied. However, many study designs will have a combination of active and assigned (focal and moderator) independent variables which leads us to the first approach to the follow-up of a significant interaction effect, tests of simple main effects.

## TESTS OF SIMPLE MAIN EFFECTS

Conducting a test of simple main effects involves examining the effect of one of the independent variables (focal, active) at a fixed level of another independent variable (moderator, assigned) included in the study. We can accomplish this by comparing the cell means on individual rows and/ or columns in a design with two factors (independent variables). Let's consider an example.

EXAMPLE 1: A researcher is interested in examining the effects of three methods of stress reduction (yoga, meditation, and exercise). Based on previous research, the researcher feels that gender may have some impact on the effects of stress reduction efforts. With this in mind, the researcher designed a study in which 90 participants (45 males and 45 females) were randomly assigned to one of the three stress reduction methods. At the conclusion of the study period, each participant completed a stress inventory (lower scores indicate less stress). The data were then analyzed using factorial ANOVA.

Dependent variable: amount of stress as measured by a stress inventory  
Independent variables: method of stress reduction - three levels (yoga, meditation, exercise)  
gender - two levels (male, female)  
Data set-up: SPSS data file with three variables - stress inventory score (score); method of stress reduction (method) with yoga coded as 1, meditation coded as 2, and exercise coded as 3; gender (gender) with male coded as 1 and female coded as 2. Note: The data for this example can be found in the Appendix.

Let's consider the initial analysis and look at the output obtained when SPSS 11.0 is used to analyze the data:

### Tests of Between-Subjects Effects

Dependent Variable: SCORE stress inventory score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2701.45 <sup>a</sup>	5	540.29	22.47	.000
Intercept	141722.63	1	141722.63	5893.27	.000
METHOD	1361.22	2	680.61	28.30	.000
GENDER	220.00	1	220.00	9.15	.003
METHOD * GENDER	1120.23	2	560.11	23.29	.000
Error	2020.05	84	24.05		
Total	146444.13	90			
Corrected Total	4721.50	89			

a. R Squared = .572 (Adjusted R Squared = .547)

If we look at the ANOVA table, we can see that the METHOD x GENDER interaction is significant ( $F=23.29$ ,  $p=.000$ ). In addition, both the method and gender main effects are significant ( $F=28.30$ ,  $p=.000$ ;  $F=9.15$ ,  $p=.003$ ). However, recall that when an interaction is significant, we focus our attention on the interaction and not the main effects.

Let's have a look at the means. The means that are in the lightly shaded area are the cell means and are of interest when considering an interaction effect. The marginal means, in the column and row labeled "Total", are the means that we are concerned with when looking at main effects. Again, the mean in the lower right corner (gray area) is the grand mean.

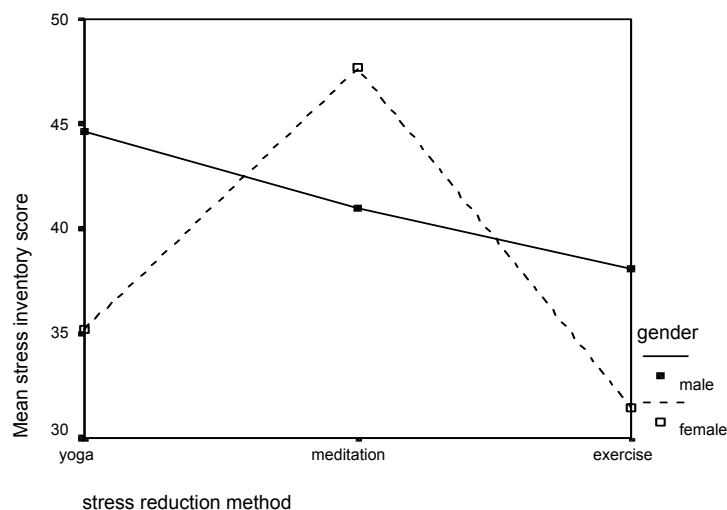
### Descriptive Statistics

Dependent Variable: SCORE stress inventory score

GENDER gender	Mean			
	METHOD stress reduction method			
	1 yoga	2 meditation	3 exercise	Total
1 male	44.68	40.95	38.11	41.25
2 female	35.22	47.67	31.47	38.12
Total	39.95	44.31	34.79	39.68

So, what do we do next? In this example, we will use the simple main effects approach to examine the interaction effect. First, let's make our focal and moderator distinction. Given the researcher's focus and the nature of the independent variables, it would seem appropriate to designate method of stress reduction our focal independent variable with gender as the moderator variable. This distinction will help us set up our tests of simple main effects. Recall from our previous discussion that an interaction exists when the effect of one independent variable is dependent on the level of the other independent variable in the design. Using our focal-moderator distinction, our initial analysis indicates that, for this example, the effect of method of stress reduction (focal) depends on the level of gender (moderator). While this is informative, we would generally like to know more. We will use the simple main effects approach to further elucidate the meaning of our significant interaction.

However, before we move on to the significance testing aspects of this approach we will first examine a graph of the cell means. This is a useful first step in considering an interaction effect. The following graph of the cell means, which helps us visualize the interaction effect, was obtained by using the 'Plots' option found in the General Linear Model(GLM)-Univariate procedure used to run the initial analysis:



As suggested previously, you can obtain a graph of the interaction when conducting the initial analysis (GLM-Univariate). After you have specified your dependent variable and the fixed factors, these additional steps will allow you to obtain the interaction graph:

--->Plots  
Horizontal axis: "method"  
Separate lines: "gender"  
--->Add --->Continue  
--->OK

Using SPSS 11.0, we will now go through the steps needed to produce the analysis necessary to conduct tests of simple main effects. The focus here will be on using a menu driven interface (i.e., pointing and clicking) to conduct the analysis. There are some time saving methods, included in the Appendix, that more advanced SPSS users may want to consider. However, as you will see, all analyses can be carried out successfully without employing the methods included in the Appendix.

**Note: Your original data file (similar to the one described above) should be open when you begin.**

1. The first step is to identify the variable that we will "hold constant." We have essentially done this by making our focal-moderator distinction. In most cases, you will want to perform the simple main effects analysis while holding the moderator variable constant. What this means is that we will be conducting a separate analysis of variance for each level of the moderator variable. In our example, we will analyze the effects of method of stress reduction for females and males, separately.

The easiest way to do this is to use the "split-file" option. We will split our file using the moderator variable. To do this (you need to be in the data file window):

Data—>Split file, Click the circle beside 'organize output by groups', the box labeled 'Groups based on' will become white, move the moderator variable (gender) over from your list on the left  
----> OK

Go back to the data file window, you should see "split file on" in the lower right corner. At this point, any analysis that you choose to run will be run on each level of the above chosen variable. For our example, all subsequent analyses will be run on females and males, separately.

2. Now, we are ready to run the analysis that we need to follow-up our significant interaction effect. We want to look at the effect of the focal independent variable at each level of the moderator variable. To do this, we simply need to run a one-way ANOVA for males and females, separately. There are several ways to accomplish this using SPSS, we will use the compare means procedure.

Analyze —>Compare Means —> One-way ANOVA  
Dependent List: "score"  
Factor: "method"  
---->Options, Click 'Descriptives' box  
---->Continue  
---->OK

We now have the analyses that we need , but there are some **calculations** that we need to do.

The output should look something like this:

**gender=male**

**Descriptives<sup>a</sup>**

SCORE stress inventory score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 yoga	15	44.68	5.55	1.43	41.60	47.75	38.33	54.68
2 meditation	15	40.95	4.53	1.17	38.44	43.46	31.83	47.61
3 exercise	15	38.11	4.03	1.04	35.88	40.34	28.13	42.66
Total	45	41.25	5.37	.80	39.63	42.86	28.13	54.68

1. GENDER gender = 1 male

**ANOVA<sup>a</sup>**

SCORE stress inventory score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	325.019	2	162.510	7.222	.002
Within Groups	945.135	42	22.503		
Total	1270.154	44			

1. GENDER gender = 1 male

**gender=female**

**Descriptives<sup>a</sup>**

SCORE stress inventory score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 yoga	15	35.22	4.63	1.20	32.65	37.78	27.19	42.60
2 meditation	15	47.67	5.09	1.32	44.85	50.49	38.45	55.17
3 exercise	15	31.47	5.42	1.40	28.47	34.47	22.69	40.23
Total	45	38.12	8.57	1.28	35.54	40.69	22.69	55.17

1. GENDER gender = 2 female

## ANOVA

SCORE stress inventory score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2156.429	2	1078.214	42.129	.000
Within Groups	1074.916	42	25.593		
Total	3231.345	44			

1. GENDER gender = 2 female

What we have are two separate ANOVA tables, one for males and one for females. Here's where the **calculations** come in. It is generally considered more appropriate to use the Mean Square Error (MSE) from the initial analysis in calculating the F for the simple main effects. If you look at the error terms for the three analyses you will note that they are different (initial analysis - 24.05, male analysis - 22.50, female analysis - 25.59). We tend to get a more efficient estimate of the population error variance when we use all relevant pieces of data to obtain the estimate, therefore the MSE from the initial analysis is the one we would prefer to use to compute the F for the simple main effects. We will use the Mean Square between from the separate analyses ( one for males and one for females). When we recompute the F, we also need to use an F table to obtain the appropriate critical value.

For our example:

$$\begin{aligned} \text{Recalculated F for males: } F &= 162.51 / 24.05 \\ &= 6.757 \end{aligned}$$

$$\begin{aligned} \text{Recalculated F for females: } F &= 1078.214 / 24.05 \\ &= 44.832 \end{aligned}$$

Since we have recalculated the F's, we have to look-up an F critical in an F table. The significance values in the simple effects analysis above are no longer accurate. For our example, the critical  $F_{\alpha=.05, 2, 84}$  is 3.15. Remember the degrees of freedom correspond to the numerator and denominator in our recalculation. In both cases, the F value exceeds the critical value suggesting that the stress reduction methods are different for both males and females. However, we still aren't finished. Since we are considering three methods, we need to apply a post-hoc procedure such as the Tukey test. Recall, if we had been comparing only two methods we would be finished. Again, we have to do some hand calculations.

For our example:

Tukey procedure:  $(\bar{x}_i - \bar{x}_j) \pm q_{\alpha, k, N-IJ} \sqrt{MSE / n}$ , where  $\bar{x}_i$  and  $\bar{x}_j$  are the means for any two groups, q is a tabled value (Studentized Range Statistic), and n=common group size.

We only need to compute the right side one time  $\rightarrow 3.399 \sqrt{24.05 / 15} = 4.304$ . The value for q (3.399) was based on  $\alpha = .05$ , k(number of groups)=3, and N(total sample)-IJ=84, where I is the number of levels of the first factor and J is the number of levels for the second factor. Recall, we are using the MSE from our initial analysis along with its associated degrees of freedom.

Now, we simply conduct all pairwise comparisons for males and then for females. One way to do this is to compute differences between all pairs of means and then compare these values to the value calculated above (4.304). When the absolute difference (disregard the sign) between the two means exceeds this value we can conclude that there is a statistically significant difference between the two means. This approach is similar to creating a 95% confidence interval (in this case,  $\alpha = .05$ ) and then determining whether the interval contains zero. Recall, when the interval contains zero, we cannot conclude that there is a statistically significant difference between the two means. The above mentioned approach is somewhat more straightforward and will lead us to the same conclusions. We will consider both approaches with the comparisons for the males:

### MALES

<u>Pairwise Comparison</u>	<u>Difference</u>	<u>95 % Confidence Interval</u> (difference $\pm$ 4.304)
yoga - meditation	44.68-40.95=3.73	(-.574, 8.034)
yoga - exercise	44.68-38.11=6.57*	(2.266, 10.874)*
meditation - exercise	40.95-38.11=2.84	(-1.464, 7.144)

For males, the mean for the yoga group is significantly higher than the mean for the exercise group (6.57 is greater than 4.304 and the 95% C.I. does not contain 0). The mean for the meditation group is not significantly different from the means for either the yoga or exercise groups ( the differences for both comparisons do not exceed 4.304 and their respective confidence intervals contain 0).

### FEMALES

<u>Pairwise Comparison</u>	<u>Difference</u>
yoga - meditation	35.22-47.67=-12.45*
yoga - exercise	35.22-31.47=3.75
meditation - exercise	47.67-31.47=16.2*

For females, we will compare the difference (absolute value) for each comparison to the value of 4.304 in order to determine statistically significant differences. We can see that the mean for the yoga group is significantly lower than the mean for the meditation group. The mean for the exercise group is also significantly lower than the mean for the meditation group. There does not appear to be a statistically significant difference between the means for the yoga and exercise groups.

So, let's put everything together and talk about our results. A two-way analysis of variance yielded a significant interaction between stress reduction method (method) and gender. This would suggest that the effect of method depended on the gender. A simple effects analysis for gender indicated that the means for the three method groups were significantly different for both males ( $F(2,84)=6.757, p<.05$ ) and females ( $F(2,84)=44.832, p<.05$ ). A post hoc analysis using Tukey's procedure ( $\alpha=.05$ ) further revealed that for males the mean of the yoga group was significantly higher than the mean for the exercise group. No significant difference was found when the yoga and meditation groups were compared or when the meditation and exercise groups were compared. For females, the mean for the meditation group was significantly higher than the means for both the yoga and exercise groups. No significant difference was found between the yoga and exercise groups. (Remember, lower scores indicate less stress.) These results would seem to suggest that for males exercise may be the more effective method of stress reduction when compared to yoga. For females, yoga and exercise both appear to be effective methods of stress reduction while meditation does not seem to be as effective as the other two methods.

## STATISTICAL COMPARISON OF CELL MEANS

We can also use all pairwise comparisons to interpret the meaning of a significant interaction effect. This approach may be most useful when the researcher's interest is in finding out which combination of the factors produces the most desirable results. An important consideration when using this approach is whether or not the groups created by the two independent variables provide for meaningful comparisons. This will often be the case when both of the independent variables are active, i.e., the researcher has control over the independent variable.

However, there are some drawbacks to this method that we should consider. This strategy is fairly conservative (i.e., power suffers) and it can produce comparisons that are not useful in interpreting the meaning of the interaction effect. Remember, we are comparing all of the cell means simultaneously. In the case of a 2x2 design, as with the example we will use, this is a reasonable approach. We only have four cell means and therefore only six unique comparisons. However, consider a 2x3 design, as seen with our first example, where we now have six cell means and 15 unique comparisons. Many of these comparisons may be of little substantive value in interpreting the interaction.

Conceptually, this approach is somewhat more straightforward than the simple main effects approach which we have already addressed. This method of follow-up calls for us to make all possible pairwise comparisons among the cell means. We can use a follow-up test, such as the Scheffe test, to accomplish this task. In considering this approach, we will use a 2 x 2 ANOVA design (the simplest case) to work through an example. Note that in this example both of the independent variables are active in nature.

EXAMPLE 2: A researcher was interested in examining the effects of homework assignments and course length on the learning of course content. The researcher designed a study in which 28 students were randomly assigned to either a homework or no homework condition as well as a 8-week session or a 16-week session. This assignment resulted in a total of 7 students participating in each of the four conditions. At the end of each session, students took a common final exam. The scores on the final exams were then analyzed via factorial ANOVA using the general linear model- univariate procedure in SPSS 11.0.

Dependent variable: learning of course content as measured by final exam score

Independent variables: homework status (homework, no homework)  
session length (8 wks, 16 wks)

Data set-up: SPSS data file with three variables - final exam score (score); homework status(hw) with homework coded as 1 and no homework coded as 2; session length (length) with 8wks coded as 1 and 16 wks coded as 2. Note: The data for this example can be found in the Appendix.



The following ANOVA table was obtained:

### Tests of Between-Subjects Effects

Dependent Variable: SCORE final exam score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1005.519 <sup>a</sup>	3	335.173	16.914	.000
Intercept	211691.1	1	211691.1	10682.681	.000
HW	276.901	1	276.901	13.973	.001
LENGTH	173.507	1	173.507	8.756	.007
HW * LENGTH	555.111	1	555.111	28.013	.000
Error	475.591	24	19.816		
Total	213172.2	28			
Corrected Total	1481.110	27			

a. R Squared = .679 (Adjusted R Squared = .639)

If we look at the shaded line, we can see that the interaction effect (HW x LENGTH) is statistically significant ( $F=28.013$ ,  $p=.000$ ). Further, both main effects are significant (HW:  $F=13.973$ ,  $p=.001$ ; LENGTH:  $F=8.756$ ,  $p=.007$ ). However, we should focus our attention on the interaction effect because interpretation of the main effects could lead us to incorrect conclusions. We should now move on to an interaction follow-up procedure. In this case, all pairwise comparisons.

The following 2x2 table of means was obtained when the initial analysis was run:

### Descriptive Statistics

Dependent Variable: SCORE final exam score

HW homework status	Mean		
	LENGTH course length		
	1 8-weeks	2 16-weeks	Total
1 homework	92.0585	88.1320	90.0953
2 no homework	76.8639	90.7477	83.8058
Total	84.4612	89.4399	86.9505

Let's review, with a 2x2 design we end up having four cell means (in the lightly shaded area) that are of interest in terms of examining the interaction effect. The four means that appear in the "total" column and row are the marginal means and are of interest with respect to the main effects. The number appearing in the gray area is the grand mean.

In using the all pairwise comparison approach, we are interested in obtaining the pairwise comparisons of the cell means. This can be done easily by converting our 2x2 design into a one-way design with four groups. Once we do this we are only a few “clicks” away from the follow-up tests that we need to interpret our interaction effect. Think of it as “stretching out” our initial design:

Example	8 wks-1	16wks-2
hw-1	1,1 (1)	1,2 (2)
no hw-2	2,1 (3)	2,2 (4)

1 (1,1)	2 (1,2)	3 (2,1)	4 (2,2)
---------	---------	---------	---------

**We first need to create a new variable that will identify our four groups.**

**Note: Your original data file (similar to the one described above) should be open.**

1. You need to be in the data file window. The following actions will create a new grouping variable (Anything appearing in double quotation marks should be typed in, in the appropriate area.):

Transform → Compute  
 Target Variable: “group”  
 Numeric expression: “1”  
 → If, Click “if case satisfies condition” box: “hw=1 & length=1”  
 →Continue  
 → OK

This has identified group 1. If you look at your data file you should see a new variable called group. Only those cases that satisfy the above condition will have a value (1). The other cases will not have a value. You need to repeat the above process for the other 3 groups.

Let’s go through the steps for group 2.

Transform → Compute  
 Target Variable: “group”  
 Numeric expression: “2”  
 → If, Click ‘if case satisfies condition’ box: “hw=1 & length=2”  
 →Continue  
 → OK

Repeat the above for groups 3 and 4, changing the numeric expression to 3 and 4 and the conditions to 2,1 and 2,2, respectively. You should now have a new variable called group. Every case should have a value for the new variable “group”.

2. You are now ready to run the follow-up tests for the interaction effect. (Again, I have used the variable names from our example.)

Analyze → Compare Means → One-way ANOVA

Dependent List: “score”  
 Factor: “group”  
 → Post Hoc, Click ‘Scheffe’ box  
 → Continue  
 → Options, Click ‘Descriptives’ box  
 → Continue  
 → OK

Now, let's look at the output that you should get:

### Descriptives

SCORE final exam score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 homework - 8 wks	7	92.06	2.76	1.04	89.50	94.61	87.27	95.60
2 homework - 16 wks	7	88.13	5.42	2.05	83.12	93.15	82.20	96.17
3 no homework - 8 wks	7	76.86	4.69	1.77	72.53	81.20	70.70	83.79
4 no homework - 16 wks	7	90.75	4.50	1.70	86.59	94.91	85.09	97.96
Total	28	86.95	7.41	1.40	84.08	89.82	70.70	97.96

First, note that the means in this table correspond to the cell means in our first table. This is good news! On to what we really need - our pairwise comparisons:

### Multiple Comparisons

Dependent Variable: SCORE final exam score

Scheffe

(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 homework - 8 wks	2 homework - 16 wks	3.93	2.38	.452	-3.22	11.08
	3 no homework - 8 wks	15.19*	2.38	.000	8.05	22.34
	4 no homework - 16 wks	1.31	2.38	.959	-5.84	8.46
2 homework - 16 wks	1 homework - 8 wks	-3.93	2.38	.452	-11.08	3.22
	3 no homework - 8 wks	11.27*	2.38	.001	4.12	18.42
	4 no homework - 16 wks	-2.62	2.38	.752	-9.76	4.53
3 no homework - 8 wks	1 homework - 8 wks	-15.19*	2.38	.000	-22.34	-8.05
	2 homework - 16 wks	-11.27*	2.38	.001	-18.42	-4.12
	4 no homework - 16 wks	-13.88*	2.38	.000	-21.03	-6.73
4 no homework - 16 wks	1 homework - 8 wks	-1.31	2.38	.959	-8.46	5.84
	2 homework - 16 wks	2.62	2.38	.752	-4.53	9.76
	3 no homework - 8 wks	13.88*	2.38	.000	6.73	21.03

\*. The mean difference is significant at the .05 level.

The above table is what we need to look at in order to interpret our significant interaction effect. This table displays the results of our pairwise comparisons using Scheffe's test. Notice that there are 12 comparisons reported in the table, only 6 comparisons are unique. We can look at the "Sig." column in order to obtain

the p-value for each of our comparisons. This information can then be used to decide whether or not the cell mean differences are statistically significant. Remember, if our p-value is less than our alpha level, we can conclude that the difference between the two means is statistically significant. We could also look at the "95% confidence interval" column to arrive at the same conclusions. Recall, if the interval does not contain zero, we can conclude that the difference between the two means is statistically significant, in this case, at the .05 level.

So, let's interpret our results....If we look at our multiple comparison table and our means, we can see that group 3, the no homework - 8 wk group, scored significantly lower than each of the other three groups. In other words, when group 3 is compared to group 1, group 2, and group 4, respectively, we find a statically significant difference using Scheffe's test. In our example, all other pairwise comparisons are not found to be statistically significant.

What does this mean for our interaction effect? Remember, an interaction occurs when the effect of one independent variable depends upon the level of the other independent variable under consideration. In our example we are considering the effects of homework status and course length on the learning of course content as measured by final exam scores. Our pairwise comparisons suggest that the homework-16 wk, no homework-16 wk, and homework-8 wk conditions produce similar results with respect to the learning of course content. However, the no homework-8 wk condition compared to the other three produces a less favorable outcome with respect to learning course content. Therefore, a recommendation based on this study would be to avoid an 8 week course that does not utilize homework.

#### ADDITIONAL CONSIDERATIONS

We have covered two approaches that can be used to follow-up a statistically significant interaction effect in the context of factorial ANOVA: tests of simple main effects and statistical comparison of cell means (all pairwise comparisons). In our discussion we have focused on designs with only two independent variables (factors). However, both of these approaches can be applied to designs with three or more factors. It gets slightly more complicated, but the logic remains the same. It should be noted that the all pairwise comparison approach may not be especially useful as the design gets more complicated. In addition, the soundness of this approach is highly questionable in designs with 3 or more factors. It is easy to see why this may be the case if we consider a 2x2x2 factorial design in which the 3-way interaction is found to be statistically significant. We have only added one additional factor with 2 levels, but we are now faced with 8 cell means and 28 unique comparisons. Clearly, this would present a challenge with regard to the interpretation of the interaction effect.

Conceptually, in the case of three or more factors, the simple main effects approach may be more useful. For example, when a 3-way interaction is found to be statistically significant, this approach would direct us to examine the 2-way interaction for two of the factors at each level of the third factor. Consider our first example where we were looking at the effects of stress reduction method and gender. Suppose that we include an additional independent variable in our design, personality type (type Y or type Z). In this case, our hypothesis might be that the effect of the stress reduction method is moderated by an individual's gender and personality type. If we found a significant 3-way interaction, one approach would be to run two separate two-way ANOVAs, one for each personality type, using the error term (MSE) from the original three-way ANOVA. If a two-way interaction is then found to be significant, we can continue to investigate the interaction effect using the techniques addressed in this handout.

Finally, while these two approaches are commonly used, they are not the only methods that can be employed to investigate an interaction effect. The decision regarding the follow-up approach should be based on the substantive question that the researcher is trying to address. The following two textbooks are recommended for the researcher who would like more information about interaction follow-up approaches: 1. Design and Analysis: A Researcher's Handbook by G. Keppel (1991), and 2. Experimental Design: Procedures for the Behavioral Sciences by R. Kirk (1995).

## Appendix

The following “short-cuts” are recommended for more advanced SPSS users. They require the user to be somewhat familiar with the use of SPSS syntax.

### **Two Options for conducting a simple main effects analysis using SYNTAX (these methods will reduce the number of hand calculations):**

- Start by identifying the focal and moderator variables. Again, this gives us direction with respect to the independent variable that will be held constant for the simple effects analysis.
- The following syntax templates can be used to conduct a simple effects analysis that will eliminate the need for recalculating the F's. Everything in bold represents SPSS commands that must be entered. Everything in italics represents the variables under consideration. Example 1 has been used to demonstrate both options.

OPTION 1: **MANOVA** *score* **BY** *gender* (1,2) *method* (1,3)  
**/DESIGN =** *gender*, *method* **W** *gender*(1), *method* **W** *gender*(2).

Note: The numbers in parentheses identify the levels of the independent variables (minimum and maximum values). Notice, on the DESIGN command line we are requesting that the analysis of method be conducted at each level of gender. These are our simple main effects.

The following output was obtained after running the above syntax file:

```
* * * * * Analysis of Variance -- design 1 * * * * *
Tests of Significance for SCORE using UNIQUE sums of squares
Source of Variation          SS          DF          MS          F          Sig of F
WITHIN+RESIDUAL              2020.05         84          24.05
GENDER                       220.00           1          220.00         9.15         .003
METHOD W GENDER(1)           325.02           2          162.51         6.76         .002
METHOD W GENDER(2)           2156.43           2          1078.21        44.84         .000

(Model)                      2701.45           5          540.29        22.47         .000
(Total)                      4721.50          89          53.05

R-Squared =                  .572
Adjusted R-Squared =        .547
```

Note that the F values for METHOD with GENDER(1) and METHOD with GENDER(2) correspond to the hand calculated F's that we obtained earlier. In this case, the next step would be to apply the Tukey procedure since the focal independent variable has three levels.

### OPTION 2: **UNIANOVA**

```
score BY method gender
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/EMMEANS = TABLES(method*gender) COMPARE(method) ADJ(BONFERRONI)
/CRITERIA = ALPHA(.05)
/DESIGN = method gender method*gender .
```

The following output was obtained after running the above syntax:

### Tests of Between-Subjects Effects

Dependent Variable: SCORE stress inventory score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2701.448 <sup>a</sup>	5	540.290	22.467	.000
Intercept	141722.629	1	141722.629	5893.268	.000
METHOD	1361.222	2	680.611	28.302	.000
GENDER	220.001	1	220.001	9.148	.003
METHOD * GENDER	1120.226	2	560.113	23.291	.000
Error	2020.051	84	24.048		
Total	146444.128	90			
Corrected Total	4721.499	89			

a. R Squared = .572 (Adjusted R Squared = .547)

Note: The above table is identical to the table obtained for the initial analysis.

### Univariate Tests

Dependent Variable: SCORE stress inventory score

gender		Sum of Squares	df	Mean Square	F	Sig.
1 male	Contrast	325.019	2	162.510	6.758	.002
	Error	2020.051	84	24.048		
2 female	Contrast	2156.429	2	1078.214	44.836	.000
	Error	2020.051	84	24.048		

Each F tests the simple effects of stress reduction method within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

Note: The F values in the above table are identical to the F values obtained using the MANOVA procedure and to the F values obtained by performing hand calculations.

In the table that follows, all pairwise comparisons are shown for each level of the moderator independent variable. Recall, pairwise comparisons would be needed if the focal independent variable had more than 2 levels. While the Tukey procedure was used earlier in presenting the calculation of pairwise comparisons, the options available with this syntax method include: LSD (Least Significant Difference), Bonferroni, and Sidak. The LSD method results in unadjusted probabilities. The Bonferroni method is the most conservative and multiplies the significance by the number of comparisons. The adjustment based on the Sidak method is less conservative than the Bonferroni adjustment and therefore may be the best option when an adjustment to the probabilities is desired.

### Pairwise Comparisons

Dependent Variable: SCORE stress inventory score

gender	(I) stress reduction meth	(J) stress reduction meth	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	5% Confidence Interval for Difference <sup>a</sup>	
						Lower Bound	Upper Bound
1 male	1 yoga	2 meditation	3.728	1.791	.116	-.635	8.090
		3 exercise	6.563*	1.791	.001	2.200	10.925
	2 meditation	1 yoga	-3.728	1.791	.116	-8.090	.635
		3 exercise	2.835	1.791	.312	-1.527	7.198
	3 exercise	1 yoga	-6.563*	1.791	.001	-10.925	-2.200
		2 meditation	-2.835	1.791	.312	-7.198	1.527
2 female	1 yoga	2 meditation	-12.449*	1.791	.000	-16.811	-8.086
		3 exercise	3.747	1.791	.114	-.616	8.109
	2 meditation	1 yoga	12.449*	1.791	.000	8.086	16.811
		3 exercise	16.195*	1.791	.000	11.833	20.558
	3 exercise	1 yoga	-3.747	1.791	.114	-8.109	.616
		2 meditation	-16.195*	1.791	.000	-20.558	-11.833

Based on estimated marginal means

\*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Sidak.

### Using SYNTAX to create a new "group" variable to conduct *all pairwise comparisons*:

The following syntax template can be used:

General form for 2x2 design:

IF (A=1 & B=1) group = 1 .

IF (A=1 & B=2) group = 2 .

IF (A=2 & B=1) group = 3 .

IF (A=2 & B=2) group = 4 .

EXECUTE .

For our example:

IF (hw=1 & length=1) group = 1 .

IF (hw=1 & length=2) group = 2 .

IF (hw=2 & length=1) group = 3 .

IF (hw=2 & length=2) group = 4 .

EXECUTE.

Note: Looking at the first line for our example, we are saying that if hw=1 and length =1, our new variable "group" will equal 1, etc.

**DATA FILES:** The data for the two examples was generated by using a random variable function in SPSS. The data was created for demonstration purposes only and in no way represents real data.

**EXAMPLE 1:**

METHOD	GENDER	SCORE	METHOD	GENDER	SCORE
1	1	42.18	2	2	38.45
1	1	46.91	2	2	42.06
1	1	53.42	2	2	52.04
1	1	45.38	2	2	52.43
1	1	48.70	2	2	52.84
1	1	38.33	2	2	46.17
1	1	39.82	2	2	48.90
1	1	39.84	2	2	44.17
1	1	43.87	2	2	43.63
1	1	46.30	2	2	41.14
1	1	38.68	2	2	46.17
1	1	54.68	2	2	47.46
1	1	51.94	2	2	50.72
1	1	41.49	2	2	53.64
1	1	38.59	2	2	55.17
1	2	39.35	3	1	37.58
1	2	27.19	3	1	32.28
1	2	34.19	3	1	37.65
1	2	36.43	3	1	41.09
1	2	37.03	3	1	33.72
1	2	41.99	3	1	37.42
1	2	29.22	3	1	42.14
1	2	37.80	3	1	38.90
1	2	36.13	3	1	41.35
1	2	35.20	3	1	42.66
1	2	42.60	3	1	28.13
1	2	34.31	3	1	38.80
1	2	32.66	3	1	38.19
1	2	27.51	3	1	41.19
1	2	36.67	3	1	40.61
2	1	42.16	3	2	27.71
2	1	41.19	3	2	24.20
2	1	34.96	3	2	33.79
2	1	45.30	3	2	34.76
2	1	39.09	3	2	22.69
2	1	37.20	3	2	33.42
2	1	47.26	3	2	29.48
2	1	43.42	3	2	28.26
2	1	40.11	3	2	31.71
2	1	44.22	3	2	37.08
2	1	47.61	3	2	22.98
2	1	38.82	3	2	37.02
2	1	31.83	3	2	35.92
2	1	37.09	3	2	40.23
2	1	43.95	3	2	32.82

N=90

**EXAMPLE 2:**

LENGTH	HW	SCORE	GROUP	LENGTH	HW	SCORE	GROUP
1	1	94.59	1	2	1	96.17	3
1	1	92.40	1	2	1	87.07	3
1	1	90.94	1	2	1	86.42	3
1	1	95.60	1	2	1	86.30	3
1	1	87.27	1	2	1	82.20	3
1	1	90.69	1	2	1	95.14	3
1	1	92.93	1	2	1	83.64	3
1	2	83.79	2	2	2	92.27	4
1	2	80.06	2	2	2	85.09	4
1	2	76.96	2	2	2	87.14	4
1	2	71.18	2	2	2	90.00	4
1	2	78.71	2	2	2	97.96	4
1	2	70.70	2	2	2	94.59	4
1	2	76.66	2	2	2	88.19	4

N=28