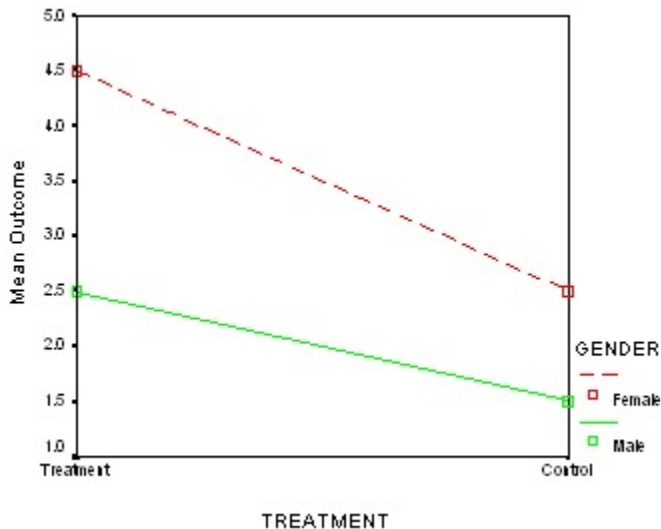# Interaction Effects in ANOVA

This handout is designed to provide some background and information on the analysis and interpretation of interaction effects in the Analysis of Variance (ANOVA). This is a complex topic and the handout is necessarily incomplete. In practice, be sure to consult the text and other references on ANOVA (Kirk, 1982; Rosenthal & Rosnow, 1991; Stevens, 1990; Winer, Brown, & Michels, 1991) for additional information.

Interaction effects represent the combined effects of factors on the dependent measure. When an interaction effect is present, the impact of one factor <u>depends</u> on the level of the other factor. Part of the power of ANOVA is the ability to estimate and test interaction effects. As Pedhazur and Schmelkin note, the idea that multiple effects should be studied in research rather than the isolated effects of single variables is one of the important contributions of Sir Ronald Fisher. When interaction effects are present, it means that interpretation of the main effects is incomplete or misleading.

<u>Kinds of Interactions</u>. For example, imagine a study that tests the effects of a treatment on an outcome measure. The treatment variable is composed of two groups, treatment and control. The results are that the mean for the treatment group is higher than the mean for the control group. But what if the researcher is also interested in whether the treatment is equally effective for females and males. That is, is there a difference in treatment depending on gender group? This is a question of interaction. Inspect the results below. Interaction results whose lines do not cross (as in the figure at left) are called "ordinal" interactions. If the slope of lines is not parallel in an ordinal interaction, the interaction effect will be significant, given enough statistical power. If the lines are parallel, then there is no interaction effect. In this case, a difference in level between the two lines would indicate a main effect of gender; a difference in level for both lines between treatment and control would indicate a main effect of treatment. When ordinal interactions are significant, it is necessary to follow up the omnibus F-test with one of the focused comparison procedures described below. Usually, the main effects can also be interpreted and tested further when they are significant.

However, when an interaction is significant and "disordinal", main effects can not be sensibly interpreted. The first graph below shows an example of a <u>disordinal</u> interaction. Disordinal interactions involve crossing lines. Generally speaking, one should <u>not</u> interpret main effects in the presence of a significant disordinal interaction. For example, in the results shown in the next graph
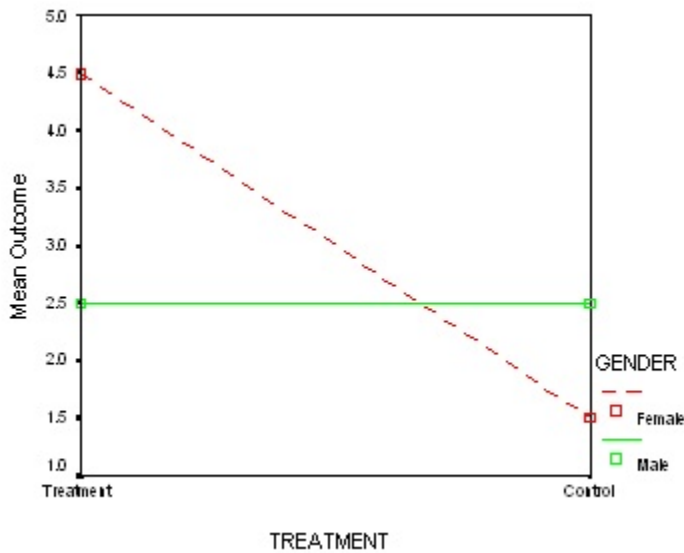
at bottom left, the results of the main effect of treatment seem to show that the treatment mean is higher than the control mean. However, one really can't say whether the treatment is effective or not in general – the effect of treatment must be qualified depending on which gender is considered; as shown in the graph of the interaction at left. Similarly, differences in gender overall are misleading. While the female mean (averaged over treatments) is higher than the male mean (averaged over treatments (see below right), this main effect result is not really the true state of affairs. Whether females score higher than males depends on the treatment condition. This is the essence of an interaction effect: results and interpretations of one variable's effect or impact must be qualified in terms of the impact of the second variable. This phenomenon is especially pronounced in the case of disordinal
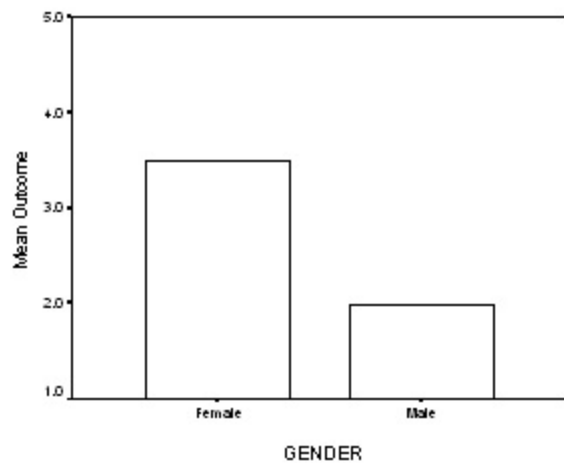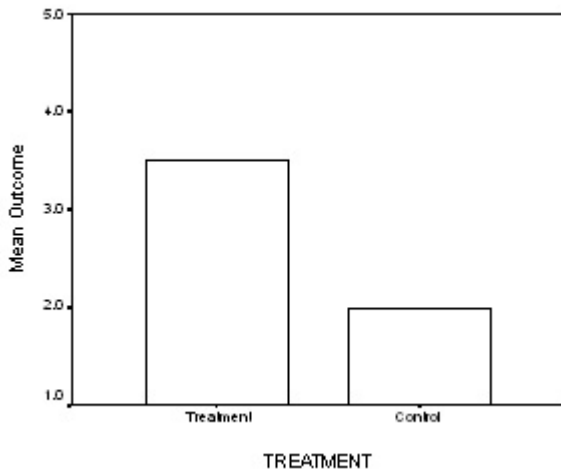
interactions and as a result, one should avoid interpreting or

discussing main effects when significant disordinal interactions are present. For example, in the present case, results for the F tests of the main effects should be reported, but interpretation should be limited to the significant interaction effect. To determine exactly which parts of the interaction are significant, the omnibus F test must be followed by more focused tests or comparisons.

FOCUSED TESTS OF INTERACTIONS

Whenever interactions are significant, the next question that arises is exactly where are the significant differences? This situation is similar to the issues in finding focused tests of main effects, but it is also more complex in that interactions represent the combined effects of two forces or dimensions in the data not just one. The following section describes four common approaches to obtaining more focused, specific information on where differences are in the interaction effect.

Method 1. Oneway ANOVA. In essence this method assumes that all relevant variance is located in the cells and there is no meaningful variance associated with the main effects. Given this assumption, it is reasonable to analyze the difference among the a by b cell means as though they are separate groups in a one-factor design. To accomplish this analysis in SPSS it is necessary to recode the ab cells into a one factor design by creating a new grouping variable. For the example above testing the interaction of gender and treatment, the ONEWAY analysis of the eight cell means starts by creating a new, four-level variable named "interact":

```
compute interact=0.
formats interact (F5.0).
if (gender eq 1 and treatment eq 1) interact=1.
if (gender eq 1 and treatment eq 2) interact=2.
if (gender eq 2 and treatment eq 1) interact=3.
if (gender eq 2 and treatment eq 2) interact=4.
execute.
```

The SPSS ONEWAY procedure (or a one factor ANOVA using GLM-Univariate) can now be used to analyze the relationship of the four-group variable "interact" with the dependent measure. Contrast procedures or post hoc procedures can be requested in SPSS to test specific differences between cell means. The advantage of this procedure is that it is easily implemented in SPSS. The disadvantage of the procedure is that it fails to separately partition out variance due to the main effects from variance due to the interaction effect (a complex topic somewhat beyond the scope of this class) and it loses the structural or dimensional information about the interaction (i.e., in this case, they're not really four groups but two different dimensions or facets of the data).

Method 2. Post Hoc Tests. This method is a direct extension of the application of post hoc tests for main effects. When the omnibus F test for the interaction is significant, it may be followed by the application of a post hoc procedure to explore which pairs of cell means are significantly different. Tukey's can be implemented using the oneway procedures described above. For most general applications, Tukey's HSD procedure is the post hoc procedure of choice as it provides true correction of alpha slippage for the number of comparisons made but does not sacrifice statistical power. As the number of cell means becomes large, however, the procedure gets cumbersome. The formula for Tukey's is:

$$HSD = q\sqrt{MSE/n^*}$$

where HSD is the minimum difference between cell means for significance, q = the studentized range statistic value for the desired alpha and the given number of cell means and denominator degrees of freedom, MSE is the denominator for the F test for the interaction effect, and n^* is the number of scores per cell (when cell sizes are unequal the harmonic mean may be used). In SPSS you must use the oneway coding of the interaction described above to perform Tukey's or other post hoc procedures. SPSS will not compute post hoc tests on interaction effects. One disadvantage of pair-wise, post hoc tests is that they do not distinguish the two-dimensional structure (in our two-factor examples) of the interaction effect in any way; all pairs of means are treated and tested equally.

Method 3. Simple Effects Tests. Simple effects procedures attempt to maintain the essential structure or nature of the interaction effect. This approach essentially breaks the interaction effect into component parts and then tests the separate parts for significance. In our present example, differences in levels of Factor A (treatment) would be tested at B_1 (female) and then at B_2 (male). Differences in levels of Factor B (gender) would then be tested at A_1 (treatment) and then at A_2 (control). The most convenient way to implement simple effects tests in SPSS is using the following syntax language which invokes an older analysis procedure, "MANOVA":

```
MANOVA score BY gender (1,2) treatment (1,2)
  /PRINT=CELLINFO (MEANS)
  /DESIGN = gender treatment gender BY treatment
  /DESIGN = gender treatment WITHIN gender(1), treatment WITHIN gender(2)
  /DESIGN = treatment gender WITHIN treatment(1), gender WITHIN treatment(2).
```

Since multiple tests are being performed, some adjustment to alpha is usually recommended. One common procedure to protect alpha by dividing the desired alpha level by the number of simple effects tests performed within each factor (see Pedhazur & Schmelkin, p. 527). In the present example, there are two simple effects tests within the gender effect so alpha = .05 / 2 = .025. The same is true for the treatment factor in this particular example. Any of the simple effects tests with a p-value less than .025 would be considered significant at a protected alpha level of .05 (note that this procedure for alpha adjustment is an application of Bonferroni's procedure).

Method 4. Planned Comparisons. Another alternative to the procedures above is the use of planned comparisons instead of the omnibus F test for the interaction. These procedures can also be implemented using the oneway procedures described above. As we have discussed previously, if a set of comparisons is orthogonal, no alpha adjustment is needed. If the comparisons are nonorthogonal, Bonferroni's or Sidak's adjustment procedures should be used.