Performing a Two-Way Analysis of Variance

Recall that you conducted initial data explorations on **Reading3** previously. No particular concerns were noted about unusual data values or the distribution of the data. Let's use the GLM procedure to generate an interaction plot and the analysis of variance. The graph enables you to visually evaluate the interaction between the factors.

The PLOTS=ALL statement requests all plots that are relevant to the analysis. The UNPACK specification specifies that plots, which are normally produced with multiple plots per panel (or packed), should be unpacked and appear in multiple panels with one option in each panel. Let's submit the code.

```
title "MYDATA.school DATA SET";
proc glm data=mydata.school plots(unpack)=all;
   class school gender;
   model reading3=school|gender;
run;
quit;
title;
```

Examining the results, we see that the ANOVA table tests the hypothesis that the treatment group means are equal. The given *p*-value is 0.0016. Presuming an alpha equal to 0.05, you reject the null hypothesis and conclude that all treatment means are not equal. Which factors explain this difference?

The descriptive statistics indicate that the average test score for all observations is 47.76. The R square for this model is approximately 0.13. This indicates that this model accounts for approximately 13% of the variability in **Reading3** scores in the school district. Actually, four types of sums of squares are available in PROC GLM. However, for one-way analysis of variance, all types of sums of squares are the same. For two-way analysis of variance, all types of sums of squares are the same if the data is balanced. Balanced data is data that has an equal sample size for all treatment combinations. The four types of sums of squares are generally not identical for unbalanced data. Type I and Type III sums of squares are produced by default in PROC GLM and Type III sums of squares are the most commonly used sums of squares for unbalanced data ANOVA. Type III sums of squares measures the effect of an independent variable assuming all other independent variables are already in the model, whereas Type I sums of squares, also called "sequential sums of squares", measures the effect of the independent variable against only the variables that have been added in the model. Basically Type I adds in terms one at a time and measures their significance, where Type III adds all variables in at once and measures significance.

The sums of squares are used to test the null hypothesis that the effect of the terms in the model is insignificant. You should consider the test for the interaction first. The *p*-value is 0.0091. Presuming an alpha of 0.05, you reject the null hypothesis. You have sufficient evidence to conclude that there is an interaction between school and gender. The interaction is best illustrated by an interaction plot, which is produced by default for a two-way ANOVA.

The interaction is evident from the graph. Also evident is the large amount of variability in the scores around each mean. Boys and girls have the same **Reading3** scores at Maple, but the scores for boys and girls seem to differ at all of the other schools. The difference in average **Reading3** values for female students at Cottonwood and Dogwood (approximately -40) does not appear to be the same as the difference in average **Reading3** values for the male students at Cottonwood and Dogwood (approximately 4). Differences in **Reading3** values can also be found between other schools. Which of the differences are statistically significant? Because of the interaction, the tests for the individual factor effects might be misleading due to the masking of these effects by the interaction. Therefore, testing the means for the two factors separately might lead to erroneous conclusions.

Let's now use the LSMEANS statement to perform multiple comparisons or interactions. The PDIFF specification requests that *p*-values for differences of the LSMEANS be produced. ADJUST= specifies the multiple comparisons adjustment. ADJUST=TUKEY is the default. Finally, the CL specification requests confidence limits for the individual LS-means. If you specify the ADJUST= option, the confidence limits for the differences are adjusted for multiple inference, but the confidence intervals for individual means are not adjusted. Let's submit the code.

```
proc glm data=mydata.school;
   class school gender;
   model reading3=school|gender;
   lsmeans school*gender / pdiff adjust=tukey cl;
run;
quit;
```

This table presents the LSMEANS for each **School*Gender** combination and assigns a number to each of the treatment combinations. The means are displayed visually in the LS-Means plot, which is available from ODS Graphics for PROC GLM. From the plot, it appears that the girls at Pine have the highest **Reading3** scores, and the girls at Cottonwood seem to have the lowest. Dogwood girls, Maple boys, and Maple girls seem to have approximately equal **Reading3** scores. Recall that the data exhibited large variability around the average **Reading3** scores for the school by gender combinations. You should confirm significance with the results of the tests for the pairwise comparisons.

Because the PDIFF option was used in the LSMEANS statement, the ODS Graphics for PROC GLM includes a diffogram. This diffogram displays all pairwise LS-means differences and their significance. The display is also known as a mean-mean scatter plot or a pairwise-difference plot. Each line segment is centered at the intersection of two least squares means. The length of the line segments corresponds to the width of a 95% confidence interval for the difference between the two least squares means. The length of the segment is adjusted for the rotation. If a line segment crosses the dashed 45-degree line, the comparison between the two factor levels is not significant. Otherwise, it is significant. The horizontal and vertical axes of the plot are drawn in least squares means units, and the grid lines are placed at the values of the least squares means.

The background grid of the difference plot is drawn at the values of the least squares means for the eight **School*Gender** combinations. These grid lines are used to find a particular comparison by intersection. Also, the labels of the grid lines indicate the ordering of the least squares means. The diffogram visually displays the test results of the pairwise comparisons. It also confirms that at a Tukey-adjusted alpha level of 0.05, the **Reading3** scores for Cottonwood girls are significantly different from the scores for girls at Maple and Pine. No other comparisons are significant at this alpha level.

This table gives the *p*-values for a test of the null hypothesis that the group means are equal. Presuming an alpha equal to 0.05, the following conclusions can be drawn:

- The average **Reading3** test score for female students at Cottonwood is significantly different from that of female students at Dogwood and female students at Pine.

Additionally, significant differences can be found if the alpha level is set to be 0.10.

- The average **Reading3** test score for female students at Cottonwood is significantly different from that of female students at Maple and male students at Maple.
- The average **Reading3** test score for female students at Pine is significantly different from that of male students at Pine.

The table also shows the 95% confidence limits for the mean of each group. Recall that these limits are not adjusted to control the experimentwise error rate.

This table gives the 95% confidence limits for the differences between the groups. These intervals are adjusted with Tukey's method to control the experimentwise error rate. The only two comparisons that are significantly different from zero are the comparisons of group 1 (Cottonwood girls) to group 3 (Dogwood girls) and to group 7 (Pine girls).

Comparing group 1 to group 3, the average **Reading3** score for Cottonwood girls is almost 40 points lower (-39.58) than the average score for Dogwood girls. The simultaneous confidence interval estimates the difference in the scores for girls at these two schools to be between -76.53 points and -2.62 points. Comparing group 1 to group 7, the average **Reading3** score for Cottonwood girls is almost 50 points lower (-48.59) than the average score for Pine girls. The simultaneous confidence interval estimates the difference in the scores for girls at these two schools to be between -11.30 and -85.87 points.

When there is a significant interaction, interpreting main effects might not be appropriate. It might be of interest to test the effect of one factor within each level of the other factor. These are known as tests for simple effects. For example, you might want to determine whether there are significant differences between female students and male students for each school, or whether there are significant differences among the schools for each gender. Although some of these tests might be included with the PDIFF option, if you are interested only in this type of test, it is easier to use the SLICE= option.

The SLICE= option specifies effects within which to test for differences between interaction LS-mean effects. This can produce what are known as tests of simple effects.

Let's submit the code.

```
proc glm data=mydata.school;
   class school gender;
```

```
   model reading3=school|gender;
   lsmeans school*gender / slice=gender slice=school;
run;
quit;
```

The *p*-values for gender F (0.0007) is significant at the alpha = 0.05 level. Therefore, there is a significant difference among the schools' average **Reading3** scores for female students. However, there is not sufficient evidence to conclude that there is a difference among different schools for male students. Presuming an alpha level of 0.05, the average **Reading3** values for Pine school are significantly different between female students and male students. At a slightly higher alpha level (0.08), you find that the average **Reading3** values are significantly different between female students and male students for Cottonwood school and Dogwood school, respectively.

Close