Interpreting the Differences Among LSMEANS in Generalized Linear Models

Robin High, University of Nebraska Medical Center, Omaha, NE

Abstract

Through ODS Graphics, various SAS procedures now offer options to produce mean plots and diffograms for visual interpretation of Lsmeans and their differences in Generalized Linear Models. Compared with "lines" and line-by-line plots of differences in lsmeans, the diffogram is the only graphical display of differences that allows four inferential and two perceptual interpretations to be made. These plots are very informative and easy to interpret across the levels of single factors in an ANOVA. In the presence of interactions where one wishes to examine simple effects by slicing differences of one factor by the levels of the second factor, the diffogram can be difficult to visually interpret. The layout of the diffogram is described which will then allow enhanced examination of Lsmean differences.

Visual Displays of LSMeans in a General Linear Model

To display data summaries for continuous data, a standard graphical figure presents the arithmetic means classified by two or more levels of a categorical factor. The left portion of Figure 1 is a common example of this type of graph. The tops of shaded vertical bars represent the value of the respective group mean with the upper and lower confidence limits of the means emanating from it as error bars (e.g., produced with PROC GCHART); the origin of each bar begins at the horizontal axis, which usually implies 0 as a lower limit.

Figure 1. Examples of mean plots and their associated differences.
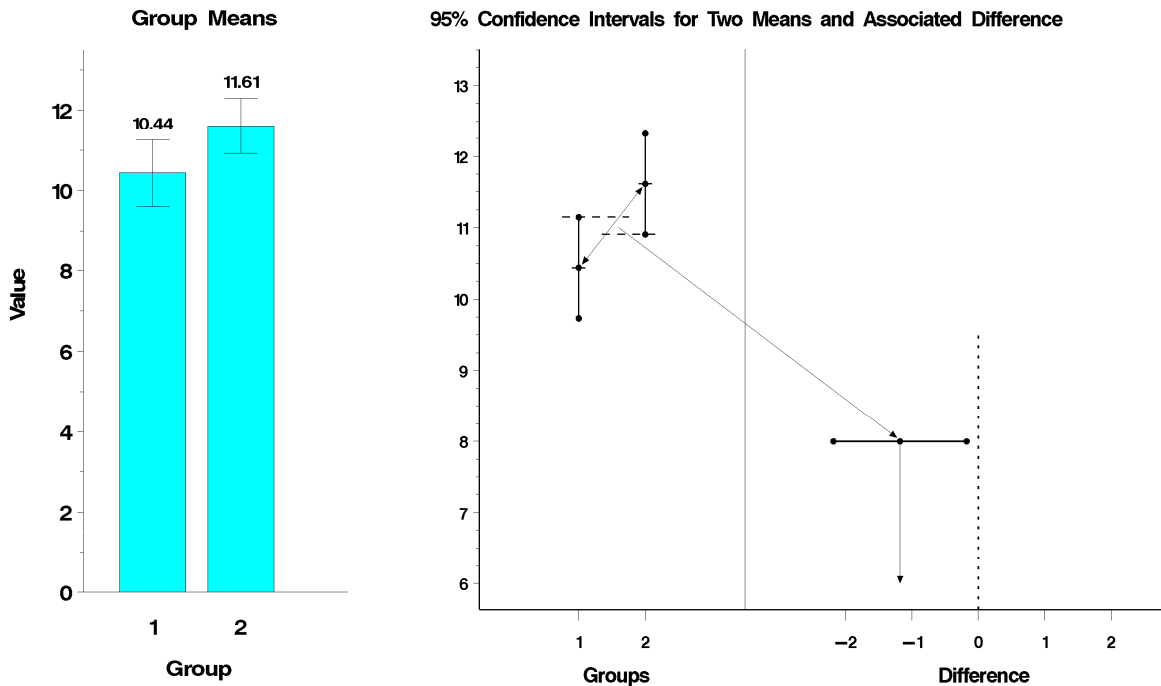
Consider these continuous data analyzed with a two sample t-test two having groups of equal sample sizes and the ideal conditions of normality, independence, and constant variance all satisfied. One important feature to note is that overlapping 95% confidence intervals of the two means (shown in the middle of Figure 1) can still produce a statistically significant difference between them (shown by the difference plot to the far right).

Julious (2004) compares the difference in two means from independent groups with equal sample sizes with the Z-test (variance known) and demonstrates how 84% confidence intervals (or smaller) drawn

around means that do not overlap ensure that statistical significance between the two groups will have significance of 0.05 (or lower) (i.e., a 95% or greater confidence interval for the difference in the two means does not contain 0).

When viewing plots of means with error bars, the casual observer may interpret significance of the difference between the means based only on ranges of the error bars placed around them. Interpretations become more problematic when three or means are shown in these graphs and when these plots are constructed with standard deviations computed from only the data belonging to each level of the classification factor and not with the pooled standard deviation computed from an ANOVA, from which statistical inferences about the differences are made (as GLIMMIX does with plot=mean option). The fact that the 95% error bars around the means in Figure 1 overlap, yet the 95% confidence interval for the difference does not contain 0 indicates why graphical displays of differences in means (evaluated with the correct standard error) is so important.

Evaluating differences among means from these vertical bar graphs becomes more difficult and even impossible when complicated analytic features such as unequal sample sizes, unequal variances, random effects, or repeated measures are present. Also, these plots do not allow multiple comparison adjustments when means from three or more categorical levels exist. Plots of means with confidence intervals convey no information about the strength of the correlations of the data for within subject effects in a repeated measures situation, so in essence, overlapping confidence intervals of the means can be inconclusive and, except for extreme cases, are essentially of little or no value for interpreting the significance of differences among means for fixed factors within subjects.

Graphical Representation of Lsmean Differences

In balanced ANOVA situations (equal group sample sizes) with only fixed effects and cross-classified designs, the Lsmeans are identical to ordinary arithmetic means. In more complicated scenarios, they may not be equal, though the hypothesis testing framework and searching for differences among means with statistical programs such as GLIMMIX has the same objective. John Tukey (1992) stated:

> multiple comparisons are somewhat complicated – as a result, graphical presentation of multiple comparison results is even more important than graphical representation of simpler results.

A good representation of multiple comparisons should, as the situation requires, allow the interpreter to make four inferential tasks between pairs of means. The inferential tasks are grouped into two sets.

The first two inferential tasks examine equality of means by looking for *significant differences* between groups:
- construct a confidence interval for $(\mu_i-\mu_j)$ defined by (L,U) where 0 is not in (L,U)
- make an accept/reject decision (i.e., $\mu_i < \mu_j$ or $\mu_i > \mu_j$)

And since equality testing is not equivalence testing (i.e., "absence of evidence is not evidence of absence"), a second pair of inferential tasks looks for *practical equivalence*:
- construct a confidence interval $(\mu_i - \mu_j)$ that is contained within (L,U) with 0 belonging to (L,U)
- make an accept/reject decision based on $-\delta_1 < \mu_i - \mu_j < \delta_2$ with $\delta_1, \delta_2 > 0$ prespecified

In addition to these four inferential tasks, intrinsic dependencies among the confidence intervals for the differences in means should be reflected in the comparisons. Two perceptual tasks should also be observable with graphical displays of differences. First, the elementary contrasts of means are *additive*:

$$\mu_i-\mu_k = (\mu_i-\mu_j) + (\mu_j-\mu_k)$$

This equation states that for any selected pairs of means with one group mean common to both contrasts, the graphical method should center the confidence interval for $\mu_i-\mu_k$ at the "sum" of the centers of the confidence intervals for $(\mu_i-\mu_j)$ and $(\mu_j-\mu_k)$.

The second perceptual task of a graphical display is the inherent *transitivity* present with significant differences; that is, if the multiple comparison method declares $\mu_i > \mu_j$ and also that $\mu_j > \mu_k$, then it must necessarily declare $\mu_i > \mu_k$.

Various visual methods currently exist to display differences among the lsmeans; among them are underlining (the "lines" option on the LSMEANS statement in PROC GLIMMIX), a line-by-line listing of the differences with a confidence interval (the cldiff option in PROC GLM or the diff option in GLIMMIX), comparison circles (available with JMP), or producing a graph of the confidence intervals by stacking them on top of one another and placing a vertical line to indicate where the value 0 lies in relation to each interval (see Figure 4).

The Diffogram

Another method to display differences in means is the diffogram (called the mean-mean scatter plot by Hsu, 1996) and beginning with Version 9.2 is available through ODS Graphics, first with the GLIMMIX procedure and now through various SAS procedures. The diffogram produced by GLIMMIX (Figure 2) is characterized by the following features:

- It is a two-dimensional plot with both the vertical and horizontal axes of the same length and having the same numerical range defined by the difference between largest and smallest lsmeans (with a slightly larger range needed to plot the confidence intervals for the differences within the plotting area).

- All possible pairs of the p LsMeans (LsMean$_i$ and LsMean$_j$, where i= 1,2,..p-1, j=2,..p, i NE j) are plotted with LsMean$_i$ on the horizontal axis (abscissa) and LsMean$_j$ on the vertical axis (ordinate).

- Within the plotting area the point of intersection of the horizontal and vertical line emanating from both axes for the two LsMeans represents the value of their difference:

$$diff_{ij} = LsMean_i - LsMean_j \quad (i= 1, 2, .. p-1, j=2, .. p, i \text{ NE } j)$$

- The upward sloping 45-degree dashed line in the plotting area represents points of equality for two means; that is, it shows how close the difference of the two lsmeans is to 0. For this reason both axes have the same length and scale and that all p means are not plotted on each axis, as the difference of any mean with itself is trivially 0 (you will see this point of intersection on the graph for the values of means that appear on both axes).

- A third axis (implied on the graph but not printed by ODS Graphics) runs at -45 degrees (from the upper left corner of the plotting area to the lower right corner). This axis represents the magnitude of the differences of the lsmeans depicted on the other two axes scaled in such a way [divided by SQRT(2)] that the confidence interval crosses the line of equality when the interval contains 0. It serves as the reference point for the differences between pairs of lsmeans and also gives the approximate values of the endpoints of the confidence interval.

- A significant difference between pairs of lsmeans (i.e., the adjusted pvalue is less than 0.05) occurs when the adjusted lower and upper endpoints of the confidence intervals are both positive or both negative; that is, the solid lines sloping at -45 degrees fall completely above or below the line of equality. This line of equality also identifies non-significant differences; when the line for the confidence interval intersects the diagonal (shown with a dashed line), the value 0 is included within the confidence interval.
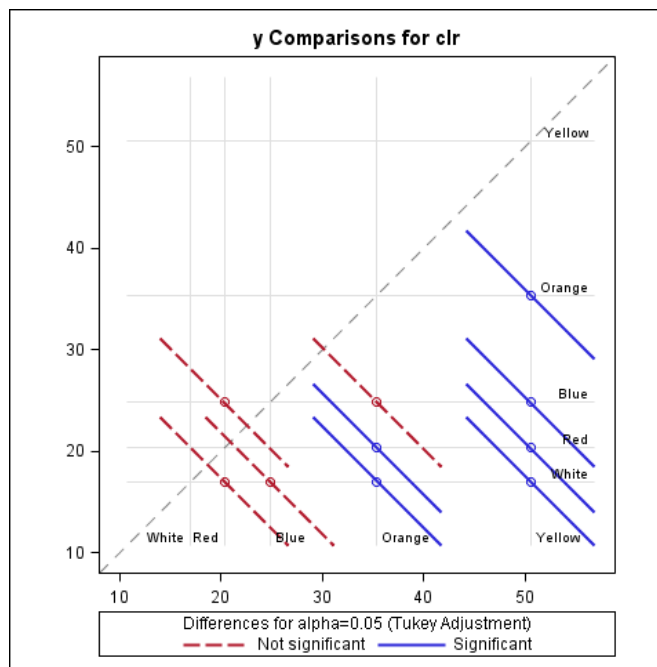
3

Figure 2. The diffogram produced by PROC GLIMMIX

The Diffogram in GLIMMIX

Options within GLIMMIX are available to produce plots for visual interpretation of the lsmeans [plot=mean() or plot=anom()] and the diffogram [plot=diff()] for the associated differences among the lsmeans when analyzing data with a Generalized Linear Model. The diffogram is the only visual representation of differences in LsMeans that allows one to make all four inferential tasks and the two perceptual tasks mentioned above.

```
ODS GRAPHICS;

PROC GLIMMIX DATA=testdata NoitPrint;
CLASS xx;
MODEL y = xx;
LSMEANS xx / cl diff PLOT=diff(NOabs center) adjust=tukey;
RUN;
ODS GRAPHICS off;
```

The LSMEANS statement contains the option to produce the diffogram, PLOT=diff() where the confidence intervals and pvalues for differences will be adjusted with the Tukey-Kramer method; it currently allows up to three options within the round brackets to affect the contents of the graphical display:

- "Noabs" plots the value of the center of the confidence interval (the pairwise difference in the lsmean) on the lower side of the diagonal line if positive and on the upper side, if negative (the default is "Abs" so that all differences are above the line)
- "center" places a circle at the center point of each confidence interval (Nocenter is the default)
- "Nolines" omits the downward sloping lines for the confidence intervals and plots only a single point for each difference

For non-normal distributions available with the distribution=<TYPE> option on the GLIMMIX MODEL statement, the differences displayed on the diffogram are on the scale of the specified or default link, e.g., the defaults are logit for binary data or log for poisson or negative binomial distributions.

4

Graphical Evaluation of the Difference in the LsMeans

Data for this example were taken from an experiment described by Wilson and Shade (1967) that reported on the relative attractiveness of five colors to insects (Yellow, Orange, Red, Blue, White). A data set with 6 replications for each of the five colors was simulated based on sample means and a pooled standard deviation from Table 5.1, p. 140, reported in Hsu (1996). The data are analyzed with a one-way ANOVA. The mean numbers of cereal leaf beetles trapped on each color are presented in Figure 3. The error bars in this figure are shown only above the means and are based on the standard deviation of the data within each color.
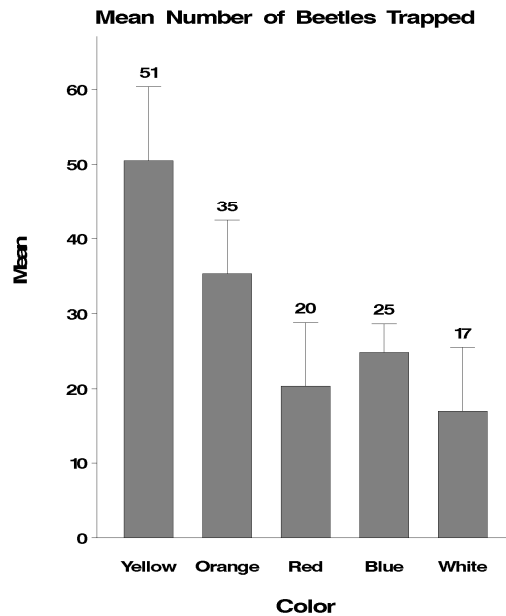

Figure 3. Mean Plots of Example Data Produced with PROC GPLOT

As demonstrated earlier, confidence intervals for means defined by error bars on rectangular bar graphs of means do not allow one to conclusively interpret the presence of significant differences among them. These data were entered into PROC GLIMMIX with an LSMEANS statement to produce the diffogram:

```
ODS GRAPHICS ;
ODS OUTPUT diffs=dfs(DROP=effect)
ODS SELECT tests3 lsmeans parameterestimates;

PROC GLIMMIX DATA=beetles NoitPrint;
CLASS clr;
MODEL count = clr / solution;
LSMEANS clr / cl diff PLOT=diff(NOabs center) adjust=tukey;
RUN;
ODS GRAPHICS close;
```

The Type III test of fixed effects for the factor color indicates that differences in lsmeans exist:

```
          Num    Den
Effect     DF     DF    F Value   Pr > F

clr         4     25     19.90    <.001
```

Table 1 contains the output from the diff option placed on the LSMEANS statement saved into an ODS output file (the same output is produced by PROC MIXED), which provides most of the information needed for construction of the diffogram. [This file, or a subset of its contents, can also be produced with the ESTIMATE or LSMESTIMATE statements.]

5

Table 1. Mean numbers of insects trapped with 95% Adjusted Confidence Intervals of the Differences

```
---------------------------------------------------------------------
|Color Comparisons         |Lower |Diffe-|Upper | Std  |   Adj    |
|                          |Limit |rence |Limit | Err  |  Pvalue  |
|--------------------------+------+------+------+------+----------|
|Compare     LsMean LsMean |      |      |      |      |          |
|Colors      Col 1  Col 2  |      |      |      |      |          |
|Yel vs Orn  50.5   35.3   |  2.5 | 15.2 | 27.8 | 4.30 |0.013 *   |
|Yel vs Red  50.5   20.3   | 17.5 | 30.2 | 42.8 | 4.30 |<.001 ***|
|Yel vs Blu  50.5   24.8   | 13.0 | 25.7 | 38.3 | 4.30 |<.001 ***|
|Yel vs Wht  50.5   17.0   | 20.9 | 33.5 | 46.1 | 4.30 |<.001 ***|
|Orn vs Red  35.3   20.3   |  2.4 | 15.0 | 27.6 | 4.30 |0.014 *   |
|Orn vs Blu  35.3   24.8   | -2.1 | 10.5 | 23.1 | 4.30 |0.14      |
|Orn vs Wht  35.3   17.0   |  5.7 | 18.3 | 31.0 | 4.30 |0.002 **  |
|Red vs Blu  20.3   24.8   |-17.1 | -4.5 |  8.1 | 4.30 |0.83      |
|Red vs Wht  20.3   17.0   | -9.3 |  3.3 | 16.0 | 4.30 |0.94      |
|Blu vs Wht  24.8   17.0   | -4.8 |  7.8 | 20.5 | 4.30 |0.38      |
---------------------------------------------------------------------
```

The factor color with five levels (p=5) produces (5*4)/2=10 pairwise comparisons of the LsMeans, represented in the rows of Table 1. Missing from the ODS output file containing estimated differences are the actual values of the LSMeans that compute these differences (the "diff" file only contains the index or classification levels). However, the values of the respective LsMeans (saved into an ODS LsMeans file) have been added to the table by saving the means into a format (with PROC FORMAT) and applying it to the two variables that define the classification levels.

For the first row of data, the value 15.2 near the center is the estimated difference between the two lsmeans for Yellow (50.5) and Orange (35.3). The columns surrounding the differences contain the upper and lower 95% confidence interval limits (e.g., 2.5, 27.8) and the associated pvalue (0.013) in the final column, all adjusted by the Tukey-Kramer method of multiple comparisons (other choices are available). The column next to the Adjusted Pvalue is the StdErr of the difference, which are all equal in this simple, balanced design, but will likely be unequal when the classification levels have different sample sizes. Also, with more complex designs the StdErr also takes for each comparison into account within subject correlations and/or random effects that are an essential component of the individual statistical test of the paired means.

Several observations about the differences in lsmeans among the five colors can be made through examination of this table. Of the four inferential and two perceptual tasks, statistical significance is easily determined by the presence or absence of 0 in the confidence interval or if the adjusted pvalue is less than 0.05 (accept/reject): Yellow is different than the other four colors; Orange is significantly different from all colors except for Blue. However, it is difficult to spatially consider the strength of these differences, including the placement of means and the width of the confidence intervals, so practical equivalence is less easily determined. It is more difficult to evaluate the perceptual differences among the means.

A line by line graph of the confidence intervals (presented in the order of the rows in Table 1) is made by plotting data from the diff file with GPLOT and is presented here in Figure 4 for comparison with the diffogram:
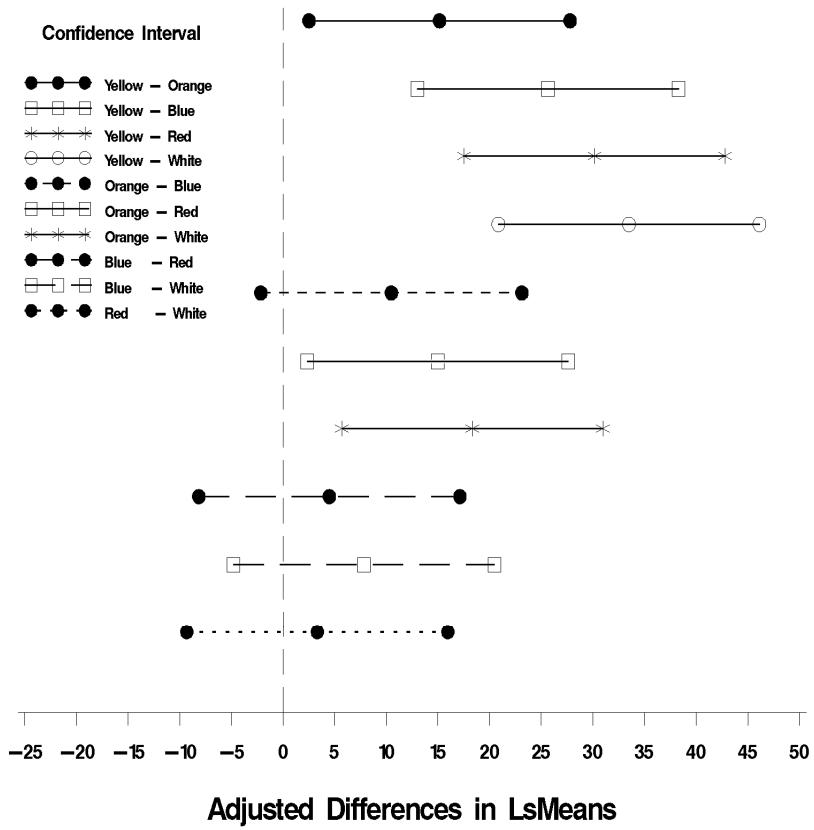
Confidence Intervals for LSMean Differences

Figure 4. Line by line plot of the differences in means and confidence intervals.

An enhanced diffogram (Figure 5) with data from the diff file allows a visual approximation of the entire contents of Table 1. Every piece of numerical information for each color, including the means, and differences in means with other colors table, and even the pvalues can be approximated from the graph. And, like the line-by-line plot of the intervals alone shown in Figure 4 it allows one to make the four inferential tasks; in addition, it allows interpretations of the two perceptual tasks.
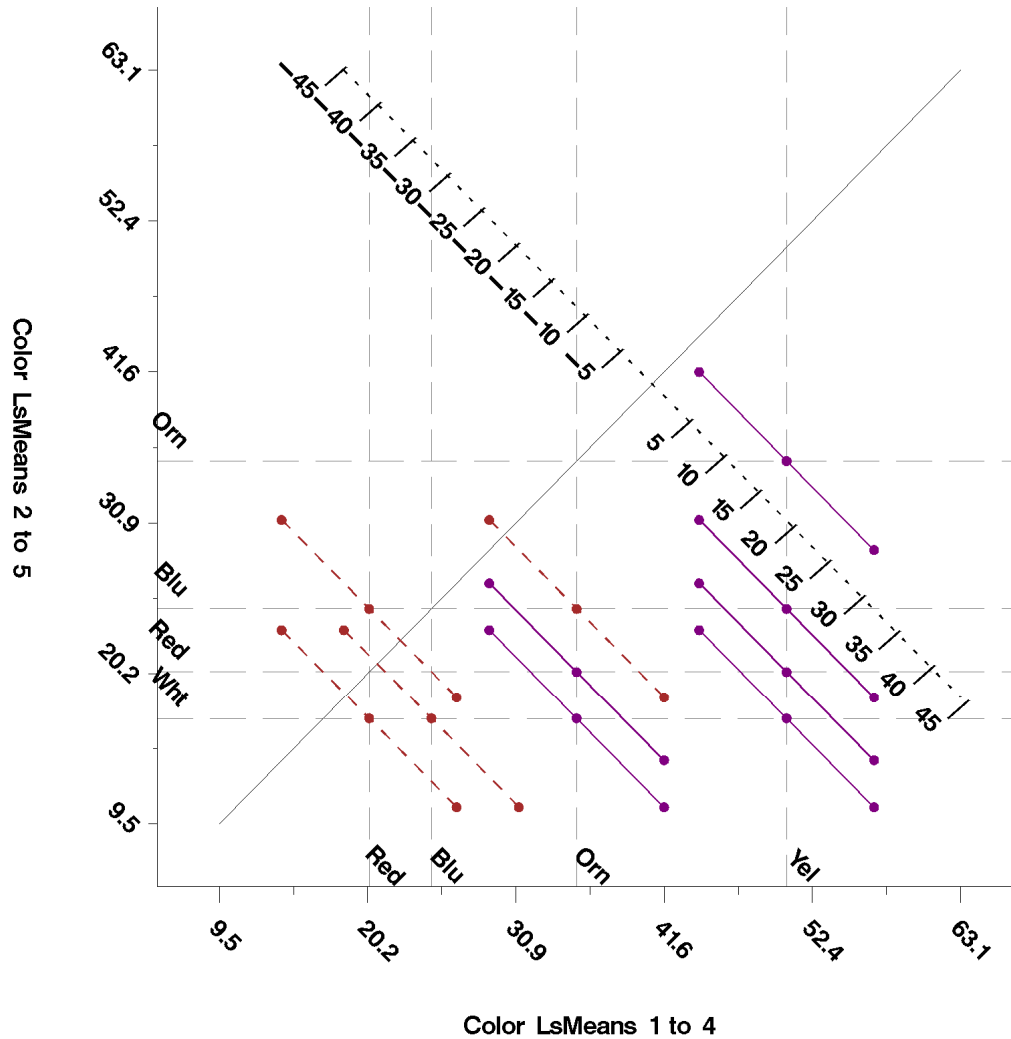
Figure 5. Enhanced diffogram produced with PROC GPLOT

A diffogram allows one to identify significant differences and practical equivalence among the pairs of lsmeans across the levels of a classification factor (e.g., a fixed effect) regardless of the complexity of the design or selected data analysis features.

Statistical Significance

Both the line-by-line plot and the diffogram of the confidence intervals allow one to make conclusions about the statistical significance of lsmean comparisons. Interpreting the diffogram requires that one first recognizes the horizontal axis lists four of the five colors (red, blue, orange, yellow) and the vertical axis also lists four colors (white, red, blue, orange), as labeled next to their respective axes. The lsmeans of each color are the points on the two axes and the value of the difference is represented where the vertical or horizontal dashed lines emanating from the lsmeans intersect; each of the ten lsmean comparisons are shown as the center point of each confidence interval, indicated by a -45 degree line. The line is solid if the difference in lsmeans is significant and dashed if it is not significant (i.e., it crosses the line of equality). The actual values of the differences and the endpoints of the intervals can be approximated by connecting these points by a 45 degree line to the third axis (sloping downward from left to right). If the difference in the two lsmeans is significant, the confidence intervals lie completely in either the lower right triangle (if the difference is positive as all the significant comparisons do in this example) and in the upper portion if the difference is negative.

8

Practical Equivalence

The length and position of the confidence intervals also determine whether practical equivalence of the two factor levels exists, that is, how tightly the confidence interval of the difference is placed around 0. Two additional lines parallel to the line of equality, at the specified upper and lower limits where the difference is trivial, can be drawn along the axis for the difference (not illustrated here). PROC TTEST evaluates equivalence of two means this way with its TOST option. The maximum values in each direction from the line of equality, such that the 90% confidence interval for the difference in the two means is completely contained within it are the minimum values where one could state the two treatments are practically equivalent.

Perceptual Tasks: Additivity and Transitivity

For interpretation of additivity refer to Figure 6 where any of the comparisons, whether they are significant or not, should display the confidence intervals in an additive sense. The placement of the differences for the Yellow/Red (30.2) and Red/Blue (-4.5) comparisons and the extended spatial placement of the Yellow/Blue comparison (25.7) further down and to the right in the lower right corner of the plotting area, where 25.7 = -4.5 + 30.2.

For transitivity, note that for pairs of significant differences, the mean for Yellow (50.5) is greater than the mean for Orange (35.3); also, the mean for Orange (35.3) is greater than the mean for White (17.0), which implies the mean for Yellow is greater than the mean for White, all represented by solid lines.
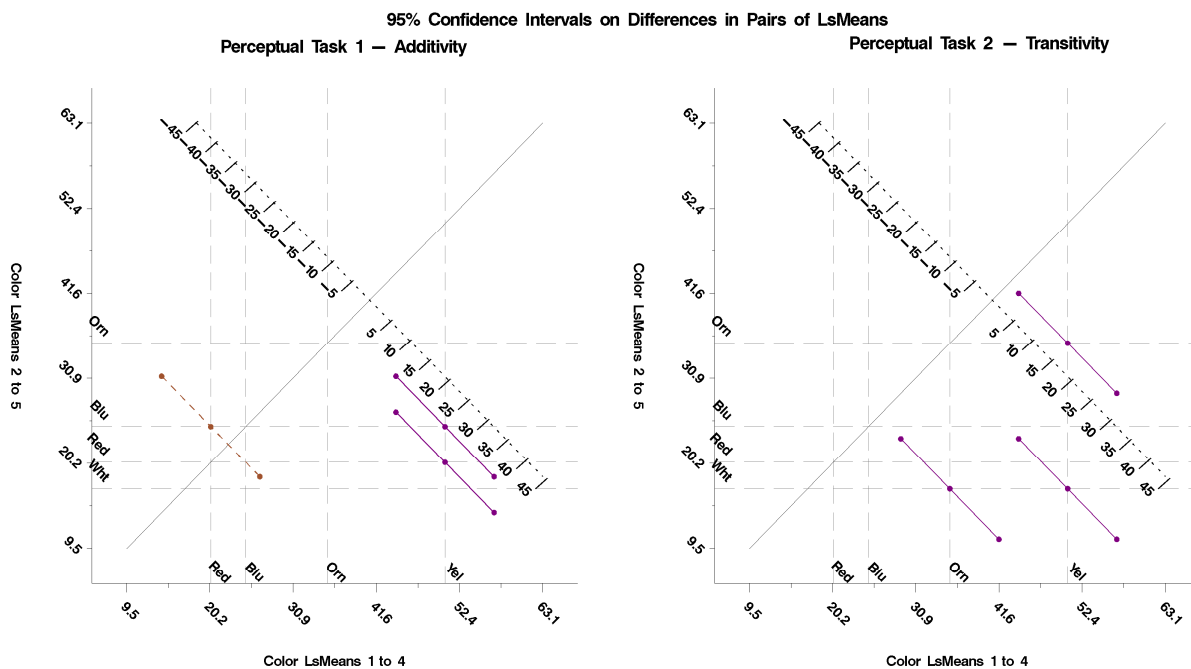

Figure 6. Diffogram Display of Perceptual Tasks Additivity and Transitivity

Difficulties Constructing Diffograms

As p gets large (e.g., p=6 implies 15 comparisons, p=7 implies 21 comparisons) or when the LsMeans are nearly equal (even with p = 3) lines plotted as confidence intervals for the differences may overlap. The placement of labels for the factor levels on the graph may also overstrike.

Another difficulty with the currently implementation of the diffogram is looking only at differences in adjusted confidence intervals of greatest inference based on a subset of all possible pairs with adjustments made based on a smaller number of comparisons (e.g., GLIMMIX offers a method to compare lsmeans of treatments with the lsmean of a control group).

In more complicated designs or with unequal variances, the differences should be adjusted for multiple comparisons with the simulate method:

```
LSMEANS factor_name / cl diff adjust= simulate(nsamp=40000 seed=121221) adjdfe=row ;
```

The basic algorithm and reasons for simulation are given in Chapter 5.2 of Westfall, et. Al. (2011).

Constructing a Diffogram with Interactions

For two classification factors with a significant interaction, the version of the diffogram from SAS 9.2 and 9.3 plots confidence intervals for all pairwise combinations among the LsMeans of both factors, so if factor A has 3 levels and factor B has 3 levels, there are 36 possible pairwise combinations of means for which to compare differences. Not only can this plot become difficult to evaluate, but it also may not easily convey the information intended that slicing interactions is intended to provide. That is, when interpreting a significant interaction, it may be more useful to fix one factor at each of its levels and evaluate the comparisons of adjusted confidence intervals for differences of lsmeans and their pvalues for the second factor.

```
ODS OUTPUT lsmeans=lsm(drop=effect alpha df tvalue probt)
           Slicediffs=slcdfs(drop=effect alpha);
PROC GLIMMIX DATA=indat;
< statements > ;
LSMEANS a*b / cl slice=a slicediff=a adjust=<enter choice> ;
RUN;
```

A file of specific differences among LSMEANs with ESTIMATE or LSMESTIMATE statements can be produced in these situations. The graphical display of specific subsets of differences is not currently available in SAS, but the author has written a SAS macro that constructs such graphs, adding the third axis which displays the values of the differences (as shown in Figure 5). Details of this macro are too numerous to explain here, but running the macro generally works well when the value of p is small for the level of interest.

References

Hsu, Jason, Mario Peruggia, (1994) *Graphical Representations of Tukey's Multiple Comparison Method*, Journal of Computational and Graphical Statistics, Volume 3, No. 2, pp. 143-161.

Hsu, Jason (1996) *Multiple Comparisons Theory and methods*, Chapman & Hall/CRC: Boca Raton.

Julious, Steven A. (2004) "Using confidence intervals around individual means to assess statistical significance between two means," Pharmaceutical Statistics, Vol. 3: 217-222.

Tukey, J. W. (1992). Where should multiple comparisons go next? In Hoppe, F. M. (ed.), *Multiple Comparisons, Selection, and Applications in Biometry: A Festschrift in Honor of Charles W. Dunnett*, Marcel Dekker, New York, Chapter 12, pp. 187-208.

Westfall, Peter H., Tobias, Randall D., and Wolfinger, Russell D., (2011). *Multiple Comparisons and Multiples Tests Using SAS*®, Cary, NC: SAS Institute Inc.

Wilson, M. C. and Shade, R. E. (1967). "Relative attractiveness of various luminescent colors to the cereal leaf beetle and the meadow spittlebug," Journal of Economic Entomology, 60:578-580.

Your comments and questions are valued and encouraged. Contact the author at:

Robin High
Statistical Coordinator
College of Public Health
Department of Biostatistics
University of Nebraska Medical Center
984375 Nebraska Medical Center
Omaha, NE 68198-4375
Phone: (402) 559-2929
email: rhigh@unmc.edu