The scaffold data revisited

February 18, 2011

Remember how I carefully changed my lecture notes to replace `means` by `lsmeans` in `proc glm`? Well, I shouldn't have done, because you don't have the `lines` option on SAS 9.1 (the version that's in Mathlab), and without `lines`, it doesn't do much good. So here's how things were before, which is the way you can make this assignment work.

Let's go back to the jumping rats first. The data looked like this:

```
Control 1 611
Control 1 621
Control 1 614
Control 1 593
Control 1 593
Control 1 653
Control 1 600
Control 1 554
Control 1 603
Control 1 569
Lowjump 2 635
Lowjump 2 605
Lowjump 2 638
Lowjump 2 594
Lowjump 2 599
Lowjump 2 632
Lowjump 2 631
Lowjump 2 588
Lowjump 2 607
Lowjump 2 596
Highjump 3 650
Highjump 3 622
Highjump 3 626
Highjump 3 626
Highjump 3 631
Highjump 3 622
Highjump 3 643
Highjump 3 674
```

```
Highjump 3 643
Highjump 3 650
```

with the variables being the amount of jumping done by each rat, a numeric version of the group name, and a measure of bone density, all separated by tabs. Thus:

```
  data jumping;
  infile "jumping.dat" delimiter='09'x;
  input group $ g density;
```

will read this in, and the following does Tukey and Bonferroni multiple comparisons.

```
proc glm;
  class group;
  model density=group;
  means group / tukey lines;
  means group / bon lines;
```

There's no problem with `lines` here, and the output is as shown below. This should be the same for you. Notice that `lsmeans` has changed to `means`, and there's no `adjust=` in there any more.

```
The GLM Procedure
             Class Level Information
Class         Levels    Values
group              3    Control Highjump Lowjump

Number of Observations Read          30
Number of Observations Used          30

The GLM Procedure
Dependent Variable: density
                                        Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F
Model                        2      7433.86667      3716.93333       7.98    0.0019
Error                       27     12579.50000       465.90741
Corrected Total             29     20013.36667

R-Square     Coeff Var      Root MSE    density Mean
0.371445      3.495906      21.58489        617.4333

Source                      DF       Type I SS     Mean Square    F Value    Pr > F
group                        2     7433.866667     3716.933333       7.98    0.0019

Source                      DF     Type III SS     Mean Square    F Value    Pr > F
```

```
group                          2     7433.866667     3716.933333       7.98    0.0019

The GLM Procedure
Tukey's Studentized Range (HSD) Test for density
NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher
II error rate than REGWQ.

Alpha                                    0.05
Error Degrees of Freedom                   27
Error Mean Square                    465.9074
Critical Value of Studentized Range  3.50643
Minimum Significant Difference        23.934

Means with the same letter are not significantly different.

Tukey
Grouping

           Mean      N    group
A       638.700     10    Highjump
B       612.500     10    Lowjump
B
B       601.100     10    Control

The GLM Procedure
Bonferroni (Dunn) t Tests for density
NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher
II error rate than REGWQ.

Alpha                               0.05
Error Degrees of Freedom              27
Error Mean Square               465.9074
Critical Value of t              2.55246
Minimum Significant Difference    24.639

Means with the same letter are not significantly different.

Bonferroni
Grouping
           Mean      N    group
A       638.700     10    Highjump
B       612.500     10    Lowjump
B
B       601.100     10    Control
```

Now, the problem comes when you have a significant interaction, and you want to do Tukey for that. There seems to be no good alternative than to define all the treatment combinations using `cat` (just like we did to get the survival plot) and then doing an analysis on *that*. This is after we've found that the interaction is significant:

```
  data scaffold;
  infile "scaffold.dat";
  input material $ weeks gpi;
  mw=cat(material,"-",weeks);

proc glm;
  class mw;
  model gpi=mw;
  means mw / tukey lines;
```

The logic here is that we create a new variable (here `mw`) that contains all the material-weeks combinations, and then we run a *one-way* analysis of variance predicting `gpi` from that. If you check, the appropriate sums of squares from the output below match the ones in the output you saw before (model SS and error SS).

Here's the output:

```
The GLM Procedure
                                      Class Level Information
Class       Levels  Values
mw              18  ecm1    -2 ecm1    -4 ecm1    -8 ecm2    -2 ecm2    -4 ecm2    -8 ecm3
                    ecm3    -4 ecm3    -8 mat1    -2 mat1    -4 mat1    -8 mat2    -2 mat2
                    mat2    -8 mat3    -2 mat3    -4 mat3    -8

Number of Observations Read          54
Number of Observations Used          54

The GLM Procedure
Dependent Variable: gpi
                                         Sum of
Source                     DF          Squares     Mean Square    F Value    Pr > F
Model                      17      37609.25926      2212.30937      86.88    <.0001
Error                      36        916.66667        25.46296
Corrected Total            53      38525.92593

R-Square     Coeff Var      Root MSE      gpi Mean
0.976206      11.74520      5.046084      42.96296

Source                     DF        Type I SS     Mean Square    F Value    Pr > F
mw                         17      37609.25926      2212.30937      86.88    <.0001
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
mw	17	37609.25926	2212.30937	86.88	<.0001

The GLM Procedure

Tukey's Studentized Range (HSD) Test for gpi

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher
II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	36
Error Mean Square	25.46296
Critical Value of Studentized Range	5.30380
Minimum Significant Difference	15.452

Means with the same letter are not significantly different.

Tukey Grouping			Mean	N	mw	
	A		73.333	3	ecm3	-8
	A					
	A		73.333	3	ecm3	-4
	A					
	A		71.667	3	ecm3	-2
	A					
	A		70.000	3	ecm1	-2
	A					
	A		65.000	3	ecm1	-4
	A					
	A		65.000	3	ecm2	-2
	A					
B	A		63.333	3	ecm1	-8
B	A					
B	A		63.333	3	ecm2	-8
B	A					
B	A		63.333	3	ecm2	-4
B						
B			48.333	3	mat1	-2
	C		26.667	3	mat3	-2
	C					
D	C		23.333	3	mat1	-4
D	C					
D	C	E	21.667	3	mat1	-8
D	C	E				
D	C	E	11.667	3	mat3	-4
D		E				
D		E	10.000	3	mat2	-2

```
D         E
D         E        10.000      3    mat3    -8
          E
          E         6.667      3    mat2    -8
          E
          E         6.667      3    mat2    -4
```

The last line of the code above gets the means for all the material-weeks combinations, and does Tukey on them. Since `mw` is not itself an interaction (as far as SAS is concerned), SAS can do it OK.

The last part of this example used only the ECM materials. That goes like this:

```
  data scaffold;
  infile "scaffold2.dat";
  input material $ weeks gpi;

proc glm;
  class material weeks;
  model gpi=weeks material;
  means material weeks / tukey lines;

run;
```

and produces the same output as we had before:

```
The GLM Procedure
        Class Level Information
Class         Levels    Values
material           3    ecm1 ecm2 ecm3
weeks              3    2 4 8

Number of Observations Read          27
Number of Observations Used          27

The GLM Procedure
Dependent Variable: gpi
                                        Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F
Model                        4      409.259259      102.314815       3.17    0.0335
Error                       22      709.259259       32.239057
Corrected Total             26     1118.518519

R-Square     Coeff Var      Root MSE      gpi Mean
0.365894      8.400247      5.677945      67.59259
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weeks	2	24.0740741	12.0370370	0.37	0.6927
material	2	385.1851852	192.5925926	5.97	0.0085

Source	DF	Type III SS	Mean Square	F Value	Pr > F
weeks	2	24.0740741	12.0370370	0.37	0.6927
material	2	385.1851852	192.5925926	5.97	0.0085

The GLM Procedure
Tukey's Studentized Range (HSD) Test for gpi
NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher
II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	22
Error Mean Square	32.23906
Critical Value of Studentized Range	3.55259
Minimum Significant Difference	6.7238

Means with the same letter are not significantly different.

Tukey Grouping		Mean	N	material
	A	72.778	9	ecm3
	A			
B	A	66.111	9	ecm1
B				
B		63.889	9	ecm2

The GLM Procedure
Tukey's Studentized Range (HSD) Test for gpi
NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher
II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	22
Error Mean Square	32.23906
Critical Value of Studentized Range	3.55259
Minimum Significant Difference	6.7238

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	weeks

A	68.889	9	2
A			
A	67.222	9	4
A			
A	66.667	9	8