

Annotated SAS Code for Multiple Comparisons

In Proc GLM, the LSMEANS statement generates treatment means. To get p-values for comparing all pairs of means, we add a "pdiff" option on the LSMEANS statement. Including an "adjust =" command after the "pdiff" will make an adjustment to the p-values to control the Type I error rate for multiple comparisons. The code below illustrates the 'adjust' option for Tukey, Bonferroni, Scheffe and Dunnett. If no adjustment is specified, SAS generates p-values corresponding to Fisher's LSD. Note that Dunnett's adjustment also requires that we specify which group is the control group. The code below uses the simulated data from Analysis of Variance, Part 5.

```
data simulated;
input treatment value @@;
datalines;
1 97 1 99 1 97 1 100 1 101 1 101 1 100 1 102
2 97 2 101 2 101 2 103 2 98 2 103 2 96 2 102
3 100 3 101 3 100 3 99 3 99 3 98 3 100 3 98
4 98 4 100 4 99 4 101 4 101 4 99 4 102 4 101
5 103 5 106 5 100 5 100 5 102 5 100 5 101 5 100
;

proc glm data=simulated;
class treatment;
model value = treatment ;
lsmeans treatment / pdiff; * equivalent to Fisher's LSD;
lsmeans treatment / pdiff adjust=tukey;
lsmeans treatment / pdiff adjust=bon;
lsmeans treatment / pdiff adjust=scheffe;
lsmeans treatment / pdiff=control('1') adjust=dunnett;
means treatment / lsd lines;
means treatment / Tukey lines;
run;
quit;
```

An alternative to the LSMEANS statement is the MEANS statement. It can compute Fisher's LSD or Tukey's HSD and put means in order with letters beside them indicating groups of means that are not significantly different. This is explained more fully in the output.

Edited Output

Each of the LSMEANS statements will produce the following table. It shows the sample mean for each treatment, and also assigns a 'LSMEAN Number' to each treatment. In subsequent comparisons of means, SAS refers to the LSMEAN numbers rather than the treatments themselves.

treatment	value LSMEAN	LSMEAN Number
1	99.625000	1
2	100.125000	2
3	99.375000	3
4	100.125000	4
5	101.500000	5

In addition to the table of LSMEANS (shown above), each of the LSMEANS statements generates a table similar to the one shown below. This is a matrix of p-values for testing the equality of each pairs of means. For example, the p-value 0.6064 (in the first row and second column) is testing the null hypothesis that LSMEANS 1 and 2 are equal (versus the alternative that they are not equal). To find out which treatments correspond to LSMEANS 1 and 2, refer to the table above. Note that the matrix of p-values is symmetric, that is, the result for testing means 1 and 2 is the same as the result for testing means 2 and 1.

Least Squares Means for effect treatment Pr > t for H0: LSMean(i)=LSMean(j)					
Dependent Variable: value					
i/j	1	2	3	4	5
1		0.6064	0.7964	0.6064	0.0593
2	0.6064		0.4407	1.0000	0.1617
3	0.7964	0.4407		0.4407	0.0338
4	0.6064	1.0000	0.4407		0.1617
5	0.0593	0.1617	0.0338	0.1617	

(Fisher adjusted p-values)

This particular table of pairwise differences is the result of the SAS statement

```
lsmeans treatment / pdiff;
```

Since there is no "adjust=" option on this statement, the p-values are derived using Fisher's LSD method.

Tables of adjusted p-values are shown on the next page.

Note that SAS prints nothing in these tables to indicate the p-values have been adjusted, or the method of adjustment. This information is given in the header of the SAS output, but not in the table itself.

Least Squares Means for effect treatment Pr > t for H0: LSMean(i)=LSMean(j)					
Dependent Variable: value					
i/j	1	2	3	4	5
1		0.9848	0.9989	0.9848	0.3113
2	0.9848		0.9348	1.0000	0.6133
3	0.9989	0.9348		0.9348	0.2000
4	0.9848	1.0000	0.9348		0.6133
5	0.3113	0.6133	0.2000	0.6133	

(Tukey adjusted p-values)

Least Squares Means for effect treatment Pr > t for H0: LSMean(i)=LSMean(j)					
Dependent Variable: value					
i/j	1	2	3	4	5
1		1.0000	1.0000	1.0000	0.5928
2	1.0000		1.0000	1.0000	1.0000
3	1.0000	1.0000		1.0000	0.3378
4	1.0000	1.0000	1.0000		1.0000
5	0.5928	1.0000	0.3378	1.0000	

(Bonferroni adjusted p-values)

Least Squares Means for effect treatment Pr > t for H0: LSMean(i)=LSMean(j)					
Dependent Variable: value					
i/j	1	2	3	4	5
1		0.9913	0.9994	0.9913	0.4468
2	0.9913		0.9608	1.0000	0.7280
3	0.9994	0.9608		0.9608	0.3198
4	0.9913	1.0000	0.9608		0.7280
5	0.4468	0.7280	0.3198	0.7280	

(Scheffe adjusted p-values)

The SAS output for Dunnett's adjustment (below) is different than the others, since Dunnett's method is used only to compare each treatment to a single control treatment. We specified treatment 1 as the control, so there is no p-value for treatment 1.

treatment	value LSMEAN	H0:LSMean=Control
		Pr > t
1	99.625000	
2	100.125000	0.9587
3	99.375000	0.9968
4	100.125000	0.9587
5	101.500000	0.1781

(Dunnett adjusted p-values)

We now discuss the output generated by the 'means' statements.

The output shown below is for the first MEANS statement:

```
means treatment / lsd lines;
```

Alpha	0.05
Error Degrees of Freedom	35
Error Mean Square	3.7
Critical Value of t	2.03011
Least Significant Difference	1.9525

Fisher's LSD

Means with the same letter are not significantly different.				
t Grouping		Mean	N	treatment
	A	101.5000	8	5
	A			
B	A	100.1250	8	2
B	A			
B	A	100.1250	8	4
B	A			
B	A	99.6250	8	1
B				
B		99.3750	8	3

The ' lines' option generates the 't Grouping' column in the second table.

Treatments in the 'A' group are not significantly different (according to Fisher's criterion), and treatments in the 'B' group are not significantly different. These groupings may seem inconsistent, but remember that these are the results of statistical tests and not algebraic manipulation. In algebra, if we know $A = B$ and $B = C$, then it will always be true that $A = C$. In statistics, knowing that A is similar to B and that B is similar to C, it does not follow that A must be similar to C.

The output shown below is for the second MEANS statement

```
means treatment / Tukey lines;
```

Alpha	0.05
Error Degrees of Freedom	35
Error Mean Square	3.7
Critical Value of Studentized Range	4.06586
Minimum Significant Difference	2.7651

Tukey's HSD

Means with the same letter are not significantly different.			
Tukey Grouping	Mean	N	treatment
A	101.5000	8	5
A			
A	100.1250	8	2
A			
A	100.1250	8	4
A			
A	99.6250	8	1
A			
A	99.3750	8	3

With Tukey's adjustment, none of the means are significantly different (they all fall into the same grouping). Recall that we are using simulated data for this example, so we know that the means are all the same.

In general, Tukey's method performs better than Fisher's whenever all pairs of means are being compared.