Means vs. LSMeans and Type I vs. Type III Sums of Squares

An experiment was conducted to study the effect of storage time and storage temperature on the amount of active ingredient present in a drug at the end of storage. A total of 16 vials of the drug, each containing approximately 30 mg/mL of active ingredient were assigned (using a completely randomized design) to the following treatments:

1) Storage for 3 months at 20^{o} C
2) Storage for 3 months at 30^{o} C
3) Storage for 6 months at 20^{o} C
4) Storage for 6 months at 30^{o} C

	Storage Temperature	
Storage Time	**20^{o}C**	**30^{o}C**
3 months	2 5	9 12 15
6 months	6 6 7 7	16

Six of the 16 vials were damaged during shipment to the laboratory where the active ingredient was measured. Accurate measures of the amount of active ingredient could be obtained only for the 10 undamaged vials. The table above shows the amount of active ingredient lost during storage (in tenths of mg/mL) for each of the undamaged vials.

We call an experiment *balanced* if all treatments have the same number of experimental units. Although this experiment was designed to be balanced with 4 experimental units per treatment, it has become *unbalanced* because the number of measured experimental units varies with treatment.

SAS will compute either *means* or *lsmeans* for each treatment or each level of a factor. When an experiment is balanced, *means* and *lsmeans* agree. When data are unbalanced, however, there can be a large difference between a *mean* and an *lsmean.* A *mean* is simply either the average of observations associated with a treatment or the average of observations associated with a level of a factor. An *lsmean* (least-squares mean) is either the average of observations associated with a treatment or the average of treatment means associated with a level of a factor.

1. Find the *mean* and *lsmean* associated with each of the four treatments in this experiment.

2. Find the *mean* associated with each level of the factor *Storage Time.*

3. Find the *lsmean* associated with each level of the factor *Storage Time.*

4. For the purpose of understanding how *Storage Time* affects loss of the active ingredient, are *means* or *lsmeans* more appropriate? Explain.

5. The *Type I sums of squares* are called *sequential* sums of squares. They always add to the overall model sum of squares. The *Type I sum of squares* for a particular model term (factor, interaction between factors, or explanatory variable) tells how much the residual sum of squares is reduced by adding the particular term to the model that contains all other terms listed before it in the model statement. The *Type I sum of squares* for the first term tells how much the residual sum of squares for the model that has one common mean for all treatments is reduced by adding the first term to the model. Use the SAS output at the end of this handout to determine the residual sum of squares for each of the following models.

a) The model that has one common mean for all treatments.

b) The model that includes only the factor *Storage Time*.

c) The model that includes the factor *Storage Time* and the factor *Storage Temperature* but no interaction between these factors.

6. The *Type III sum of squares* for a particular model term (factor, interaction between factors, or explanatory variable) tells how much the residual sum of squares is reduced by adding the particular term to the model that contains all other terms in the model statement. Determine the residual sum of squares for the model that does not include *Storage Time* main effects but does include the factor *Storage Temperature* and the *Storage Time* Storage Temperature* interaction.

7. Conduct a *t*-test that is equivalent to the *F*-test for *Storarge Time* in the *Type III sum of squares* portion of the output. Note that this test is based on the difference between the *lsmeans* for the levels of the factor *Storage Time*. Also note the difference between the *Storarge Time* test results in the *Type I* and *Type III* portions of the output.

Answer for 7: To compare overall effects of storage time factor,

g = (6.5 + 16)/2 - (3.5+12)/2 = (1/2)6.5+(1/2)16+(-1/2)3.5+(-1/2)12
= 11.25 – 7.75
= 3.5
= lsmean for 6 months - lsmean for 3 months

$$SE(g) = \sqrt{MSW}\sqrt{\frac{(1/2)^2}{4} + \frac{(1/2)^2}{1} + \frac{(-1/2)^2}{2} + \frac{(-1/2)^2}{3}}$$

$$= \sqrt{3.9166667}\sqrt{\frac{1}{4}\left(\frac{1}{4} + \frac{1}{1} + \frac{1}{2} + \frac{1}{3}\right)} = 1.4283$$

$$t = \frac{g-0}{SE(g)} = \frac{3.5}{1.4283} = 2.45$$ df = 6 two-sided p-value = 0.0498

Note: $t^2 = (2.45)^2 = 6.01 = F$ (from SAS output, Type III SS)
F-test from Type III equivalent to t-test

```
data one;
  input time temp loss;
  cards;
3 20 2
3 20 5
3 30 9
3 30 12
3 30 15
6 20 6
6 20 6
6 20 7
6 20 7
6 30 16
;

proc glm;
  class time temp;
  model loss=time temp time*temp;
run;
```

```
The GLM Procedure

   Class Level Information

Class         Levels    Values
time               2    3 6
temp               2    20 30

Number of observations    10

Dependent Variable: loss
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	159.0000000	53.0000000	13.53	0.0044
Error	6	23.5000000	3.9166667		
Corrected Total	9	182.5000000			

R-Square	Coeff Var	Root MSE	loss Mean
0.871233	23.28302	1.979057	8.500000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
time	1	0.1000000	0.1000000	0.03	0.8783
temp	1	158.4200000	158.4200000	40.45	0.0007
time*temp	1	0.4800000	0.4800000	0.12	0.7382

Source	DF	Type III SS	Mean Square	F Value	Pr > F
time	1	23.5200000	23.5200000	6.01	0.0498
temp	1	155.5200000	155.5200000	39.71	0.0007
time*temp	1	0.4800000	0.4800000	0.12	0.7382