Two-Way ANOVA
Part 5: SAS Statements

STAT 705: Regression and Analysis of Variance

Introduction

- So far, we have performed calculations ‘by hand’.
- In this lesson, we will use SAS to generate estimates and perform inference for two-way ANOVA.
- We will continue to use the Fabric data, so we can compare our hand calculations to the SAS output.
- At the and of this lesson, you should understand SAS syntax for two-way ANOVA and be able to interpret the output.

SAS Statements

```
proc glm data=subset;
 class level salt;
 model temperature = level | salt /ss3 solution;
 lsmeans level salt level*salt / stderr cl;
run;
```

- Use Proc GLM (instead of Proc Reg), because Proc Reg does not allow categorical predictors.
- The class statement identifies both ‘level’ and ‘salt’ as categorical predictors. SAS automatically generates the appropriate indicator variables.

. . . continued . . .

SAS Model Statement

```
proc glm data=subset;
 class level salt;
 model temperature = level | salt /ss3;
 lsmeans level salt level*salt / stderr cl;
run;
```

- The model statement defines ‘temperature’ as the response.
- The vertical bar between ‘level’ and ‘salt’ tells SAS to generate a model with main effects for ‘level’ and ‘salt’ and also include the interaction between them.
- After the slash (on the model statement), we specify the ‘ss3’ option. This instructs SAS to calculate Type 3 sums of squares.

. . . continued . . .

SAS LSmeans Statement

```
proc glm data=subset;
 class level salt;
 model temperature = level | salt /ss3;
 LSmeans level salt level*salt / stderr cl;
run;
```

- The LSmeans statement generates least squares estimates for the means that are listed in the statement. In this example, we will calculate marginal means (‘level’ and ‘salt’) and also the treatment means (‘level*salt’).
- We specify two options on the LSmeans statement (after the slash)
  - ‘stderr’ generates the standard errors of the means
  - ‘cl’ generates the confidence limits (i.e. confidence interval)

Output: Model Statement

The model statement generates 3 tables

1) The standard ANOVA table
2) The standard output summarizing the fit
3) A new table that gives details for main effects and interaction

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	79149.61111	15829.92222	30.45	<.0001
Error	12	6238.66667	519.88889		
Corrected Total	17	85388.27778			

R-Square	Coeff Var	Root MSE	temperature Mean
0.926938	2.870266	22.80107	794.3889

Source	DF	Type III SS	Mean Square	F Value	Pr > F
level	1	17734.72222	17734.72222	34.11	<.0001
salt	2	60928.77778	30464.38889	58.60	<.0001
level*salt	2	486.11111	243.05556	0.47	0.6375

The new table provides the partitioned sum of squares for the model.

The sum of squares for level, salt and the interaction add up to the Model SS.

The degrees of freedom for level, salt and the interaction add up to the Model df.

Output: LSmeans Statement

```
LSmeans level salt level*salt / stderr cl;
```

This statement generates estimates for (1) the marginal means of level, (2) the marginal means of salt, and (3) treatment means for level*salt. Since we included the 'stderr' option, the standard errors are also computed. The confidence limit are given in other tables.

1

level	temperature LSMEAN	Standard Error	Pr > |t|
1	763.000000	7.600357	<.0001
2	825.777778	7.600357	<.0001

2

salt	temperature LSMEAN	Standard Error	Pr > |t|
CaCO3	752.833333	9.308499	<.0001
CaCl2	753.666667	9.308499	<.0001
Untreated	876.666667	9.308499	<.0001

3

level	salt	temperature LSMEAN	Standard Error	Pr > |t|
1	CaCO3	727.000000	13.164205	<.0001
1	CaCl2	723.666667	13.164205	<.0001
1	Untreated	838.333333	13.164205	<.0001
2	CaCO3	778.666667	13.164205	<.0001
2	CaCl2	783.666667	13.164205	<.0001
2	Untreated	915.000000	13.164205	<.0001

More Output: LSmeans Statement

```
LSmeans level salt level*salt / stderr cl;
```

The ‘cl’ option on the lsmeans statement generates the confidence limits (i.e. confidence intervals) for each mean. These are given in separate tables, one table for each effect.

level	temperature LSMEAN	95% Confidence Limits	
1	763.000000	746.440244	779.559756
2	825.777778	809.218022	842.337534

salt	temperature LSMEAN	95% Confidence Limits	
CaCO3	752.833333	732.551857	773.114810
CaCl2	753.666667	733.385190	773.948143
Untreated	876.666667	856.385190	896.948143

level	salt	temperature LSMEAN	95% Confidence Limits	
1	CaCO3	727.000000	698.317661	755.682339
1	CaCl2	723.666667	694.984328	752.349006
1	Untreated	838.333333	809.650994	867.015672
2	CaCO3	778.666667	749.984328	807.349006
2	CaCl2	783.666667	754.984328	812.349006
2	Untreated	915.000000	886.317661	943.682339

Pairwise Differences

- To get tests of pairwise differences in SAS, we include the 'pdiff' option on the LSmeans statement
- Syntax:

```
LSmeans level salt level*salt / stderr cl pdiff;
```

- This will generate three additional tables of pairwise differences (one for level, one for salt and one for level*salt)
- The table for level*salt is shown on the next slide.

Example of Pairwise Differences

Least Squares treatment means ('level*salt' in SAS). Use this to get the LSMEAN numbers (last column).

level	salt	temperature LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	CaCO3	727.000000	13.164205	<.0001	1
1	CaCl2	723.666667	13.164205	<.0001	2
1	Untreated	838.333333	13.164205	<.0001	3
2	CaCO3	778.666667	13.164205	<.0001	4
2	CaCl2	783.666667	13.164205	<.0001	5
2	Untreated	915.000000	13.164205	<.0001	6

Least Squares Means for effect level*salt
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: temperature

i/j	1	2	3	4	5	6
1		0.8609	<.0001	0.0168	0.0102	<.0001
2	0.8609		<.0001	0.0120	0.0073	<.0001
3	<.0001	<.0001		0.0076	0.0125	0.0014
4	0.0168	0.0120	0.0076		0.7928	<.0001
5	0.0102	0.0073	0.0125	0.7928		<.0001
6	<.0001	<.0001	0.0014	<.0001	<.0001	

We are usually interested in comparing two levels of one factor, while keeping the other factor level fixed. For example, compare $CaCl_2$ to Untreated, both at Level 1. These are LSmeans 2 and 3, p-value <.0001.
We would probably not want to test LSmeans 3 and 5 because both factor levels are different.

Contrasts in SAS

- To specify the contrast in SAS, we need to
  1) re-write the contrast in terms of the effects (μ, α's and β's), and
  2) use the factor levels in the order that SAS defines them
- The order of the factor levels is specified in the 'Class Level Information' table

Class Level Information		
Class	Levels	Values
level	2	1 2
salt	3	CaCO3 CaCl2 Untreated

- The cell means are denoted by μ_{ij} (with i for 'Level' and j for 'Salt'). In SAS, j=1 corresponds to $CaCO_3$, j=2 is $CaCl_2$ and j=3 is Untreated.

Contrast Coefficients in SAS

- Recall that we are working with the interaction model, in which the cell means are defined by

  $$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

- For the interaction terms, the order of the subscripts is

  (i, j): (1,1), (1,2), (1,3), (2,1), (2,2), (2,3)

- In the previous lesson, we tested the contrast that compared the average temperature for salts $CaCO_3$ and $CaCl_2$ at Level 1 to the average temperature of $CaCO_3$ and $CaCl_2$ at Level 2.

- Using SAS's subscripts, the contrast is

  $$\tfrac{1}{2}\mu_{11} + \tfrac{1}{2}\mu_{12} - \tfrac{1}{2}\mu_{21} - \tfrac{1}{2}\mu_{22}$$

Contrasts in SAS

- The contrast is $\frac{1}{2}\mu_{11} + \frac{1}{2}\mu_{12} - \frac{1}{2}\mu_{21} - \frac{1}{2}\mu_{22}$
- We need to write this in terms of the effects
  - Substitute $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$
- The contrast is

$$\tfrac{1}{2}(\mu+\alpha_1+\beta_{11}+(\alpha\beta)_{11}) + \tfrac{1}{2}(\mu+\alpha_1+\beta_2+(\alpha\beta)_{12})$$

$$-\tfrac{1}{2}(\mu+\alpha_2+\beta_1+(\alpha\beta)_{21}) - \tfrac{1}{2}(\mu+\alpha_2+\beta_2+(\alpha\beta)_{22})$$

$$= \alpha_1 - \alpha_2 + \tfrac{1}{2}(\alpha\beta)_{11} + \tfrac{1}{2}(\alpha\beta)_{12} - \tfrac{1}{2}(\alpha\beta)_{21} - \tfrac{1}{2}(\alpha\beta)_{22}$$

- Note that $(\alpha\beta)_{13}$ and $(\alpha\beta)_{23}$ are missing – these will have coefficients 0

Syntax of SAS Contrast Statement

```
contrast 'ContrastName' level 1 -1 level*salt .5 .5 0 -.5 -.5 0;
```

- `'ContrastName'` is any name you want to give to this contrast. It will be printed in the output.
  - (This is handy if you have more than one contrast.)
- The rest of the statement defines the coefficients for the contrast: $\alpha_1 - \alpha_2 + ½(\alpha\beta)_{11} + ½(\alpha\beta)_{12} – ½(\alpha\beta)_{21} – ½(\alpha\beta)_{22}$
- Be careful with subscripts. They must be in the correct order.
- SAS Output:

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
ContrastName	1	9352.083333	9352.083333	17.99	0.0011

- SAS reports F = 17.99; by hand, we calculated t = -4.241
  - $(-4.241)^2 = 17.99$ (i.e. $t^2 = F$) so these are the same

What You Should Know

- SAS output duplicates much of what we calculated by hand
- The complete SAS code is provided as part of this lesson. You should run this code and examine the SAS output.
- Be able to match our hand calculations with SAS output for these values:
  - treatment means and their standard errors
  - marginal means and their standard errors
  - the difference of two means and its standard error
  - contrasts and their standard errors
  - confidence intervals and hypothesis tests for all of these
- There will be more examples in the next lesson