SAS Lab 2

Hao Zhang

Objective of the Lab

- Run SAS for the two-way models
- Test hypothesis for the interaction effects
- Test hypothesis for the mean effects
- Multiple comparisons for the main effects
- Use `estimate` statement for any contrast
- Know the difference of the ordering of factors in the `class` statement
- Know the difference between `lsmeans and means statements`

An experiment was conducted to determine the effects of three different pesticides of the yield of fruit from two different varieties of a citrus tree. Six trees of each variety were randomly selected from an orchard. The four pesticides were then randomly assigned to two trees of each variety, and applications were made according to recommended levels. Yields of fruits (in bushels per tree) were obtained after the test period.

	Pesticide		
Variety	**1**	**2**	**3**
1	49	50	38
	39	55	*
2	55	67	53
	41	58	42

*Missing: results in unequal sample sizes

* Note: This tree was accidently hit by a vehicle and has no yield data.

Two-way Complete Model

1. This is a completely randomized design. Without any knowledge about the interaction, we will employ the two way complete model with Pesticide as Factor A (with 3 levels) and Variety as Factor B (with two levels).

pesticide

variety

```
data pest;
input variety   pesticide   yield;
lines;
1   1   49
1   1   39
1   2   50
1   2   55
1   3   38
2   1   55
2   1   41
2   2   58
2   3   53
2   3   42
2   2   67
;
proc print;
run;

proc glm data=pest;
class pesticide variety;
model yield=variety pesticide variety*pesticide;
run;
```

skip missing

Default $\alpha=0.05$

A B

interaction

(1) (2) (3)

2. Is there a significant interaction effect? (3)

The p-values 0.82. Fail to reject the null hypothesis. No significant interaction effects.

Always use the p-value that corresponds to the Type III SS. (Sum Squares)

3. Are there significant differences in the mean yields among the three pesticides? Use $\alpha = 0.10$

The p-values is 0.089. We reject the null hypothesis at $\alpha = 0.1$ level of significance and believe there is a significance among the three main effects of pesticide.

4. Are there significant differences in the mean yields between the two varieties? Use $\alpha = 0.10$.

The p-value is ~~0.1998~~ 0.1429 and we fail to reject the null hypothesis which says no difference between the main effects of variety.

5. Find 90% simultaneous confidence intervals for the pairwise comparisons of the main effects of pesticide.

Use Tukey's method. The SAS code is (under the model statement)

Least Squares Means for Effect pesticide				
i	j	Difference Between Means	Simultaneous 90% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-11.500000	-24.940808	1.940808
1	3	3.250000	-11.777280	18.277280
2	3	14.750000	-0.277280	29.777280

```
proc glm data=pest;
class variety pesticide;
model yield=variety pesticide variety*pesticide;
lsmeans pesticide/cl pdiff adjust=Tukey;
run;
```

Default $\alpha=0.05$
Otherwise specify

lsmeans pesticide / cl pdiff adjust=tukey alpha=0.10;

There are 3 main effects and consequently 3 all pairwise comparisons for the main effects. The SAS output is below.

		Difference Between	Simultaneous 95% Confidence Limits
j	j	Means	for LSMean(i)-LSMean(j)
1	2	-11.500000	-28.139482, 5.139482
1	3	3.250000	-15.353507, 21.853507
2	3	14.750000	-3.853507, 33.353507

Note what each option in the `lsmeans` statement does:

"Confidence + least"

- `cl` : To produce the least squares estimates and the confidence interval for each treatment means (these are not simultaneous confidence intervals).

"pair differences"

- `pdiff` : To provide confidence intervals for pairwise differences. If a method for simultaneous confidence intervals are provided through `adjust=` , then simultaneous confidence intervals are provided. Compare the difference if you use the following statement:

```
lsmeans pesticide/cl pdiff;
```

You will see that the confidence intervals differ from those provided by $\boxed{\text{Tukey's method.}}$ Why?

6. Find a 95% confidence interval for the difference between the mean yield for the two varieties (variety 2-variety 1) when pesticide 1 is applied.

This requires some careful work. We are comparing two specific treatment means in a two-way model. It is doable but not as straightforward as in the one-way model.

First we must know how SAS coded the treatments. Since `pesticide` is entered into the `class` statement before `variety` , `pesticide` is Factor A and `variety` is Factor B. We are therefore comparing μ_{12} with μ_{11}

types (2) variety (B)

types (3) pesticide (A)

μ_{AB}

(A) (B) interaction (A) (B) interaction

$$\mu_{12} - \mu_{11} = (\cancel{\mu} + \cancel{\alpha_1} + \beta_2 + (\alpha\beta)_{12}) - (\cancel{\mu} + \cancel{\alpha_1} + \beta_1 + (\alpha\beta)_{11})$$
$$= (-\beta_1 + \beta_2) + (-(\alpha\beta)_{11} + (\alpha\beta)_{12}).$$

Therefore, this contrast involves the main effects of Factor B (variety) and the interaction effects. The SAS code is

11 21
12 22
13 23 } 6 total interaction

```
estimate "v2-v1|pesticide=1" variety -1 1 variety*pesticide -1 1 0 0 0 0;
```

1 2 3   11 12 21 22 31 32

mean yield | applied   *specify variables name first

SAS output provides estimate=4.0000000 and standard error=7.23187389. We will need to find the t critical value $t_{5,0.025} = 2.5706$. We therefore have the 95% confidence interval 4.0 ± 18.59. Estimate $\pm (t_{df,\alpha/2})(SE)$

t-table

6. Find a 95% confidence interval for the difference between the mean yield for pesticide 2 and pesticide 3 when they are applied to Variety 1 trees.

Here we are estimating $\mu_{21} - \mu_{31}$, for which we must first express it interms of main-effects and interaction effects.

variety (B)

pesticide (A)

$$\mu_{21} - \mu_{31} = (\cancel{\mu} + \alpha_2 + \cancel{\beta_1} + (\alpha\beta)_{21}) - (\cancel{\mu} + \alpha_3 + \cancel{\beta_1} + (\alpha\beta)_{31})$$
$$= (\alpha_2 - \alpha_3) + ((\alpha\beta)_{21} - (\alpha\beta)_{31}).$$

The SAS code is

(1,2) (1,2,3)

A 1 2 3 B A   11 12 21 22 31 32

```
estimate "P2-P3|v=1" pesticide 0 1 -1 variety*pesticide 0 0 1 0 -1 0;
```

variety=1

Code follows AxB format depend which variables comes before "class".

The estimate and standard error are 14.5000000 and 8.85720046, respectively. The confidence interval can be given similar to the previous question.

7. In this problem, I will show you the order that the factors appear in the `class` statement is important. Compare the output from the following program:

```
proc glm data=pest;
class pesticide variety;
model yield=variety pesticide variety*pesticide/solution;
run;
```

A B

Shows detailed parameter

with those from

```
proc glm data=pest;
             A          B
class variety pesticide;
model yield=variety pesticide variety*pesticide/solution ;
run;
```

We see that the interaction plots are different and the treatments are coded differently.

Two-way Main-Effects Model

Suppose the experimenter has knowledge that no interaction effects exist. Then the two-way main effects model can be used. Let us see how to get the simultaneous confidence intervals in Problem 5 and the confidence interval in Problem 6.

```
proc glm data=pest;
             A          B
class pesticide variety;
model yield=pesticide variety;
lsmeans pesticide/cl pdiff adjust=Tukey;
estimate 'v2-v1' -1 1;
run;     B2-B1
```

Note how the p-values changed! The confidence intervals for the pairwise comparison become shorter.

Download Print

Activity Details

Task: View this topic