

Topic 23 - Unequal Replication

STAT 525 - Fall 2013

Outline

- Data
- Model
- Parameter Estimates
- Inference

Data for Two Factor ANOVA

- Y is the response variable
- Factor A has levels $i = 1, 2, \dots, a$
- Factor B has levels $j = 1, 2, \dots, b$
- Y_{ijk} is the k^{th} observation from cell (i, j)
- Now $k = 1, 2, \dots, n_{ij}$

Example Page 954

- Synthetic growth hormone administered to growth hormone deficient pre-pubescent children
- Interested in two factors
 - Gender ($a = 2$)
 - Bone development level ($b = 3$)
- Y is the difference between growth rate during treatment and prior to treatment
- Set up as balanced design ($n = 3$) but four children were unable to complete the study
 - $i = 1, 2$ and $j = 1, 2, 3$
 - $n_{ij} = 3, 2, 2, 1, 3, 3$

General Plan

- Construct scatterplot / interaction plot
- Run full model
- Check assumptions
 - Residual plots
 - Histogram / QQplot
 - Ordered residuals plot
- Check significance of interaction

SAS Commands

```
options nocenter;
data a1; infile 'u:\.www\datasets525\CH23TA01.txt';
input growth gender bone;
proc print data=a1; run;

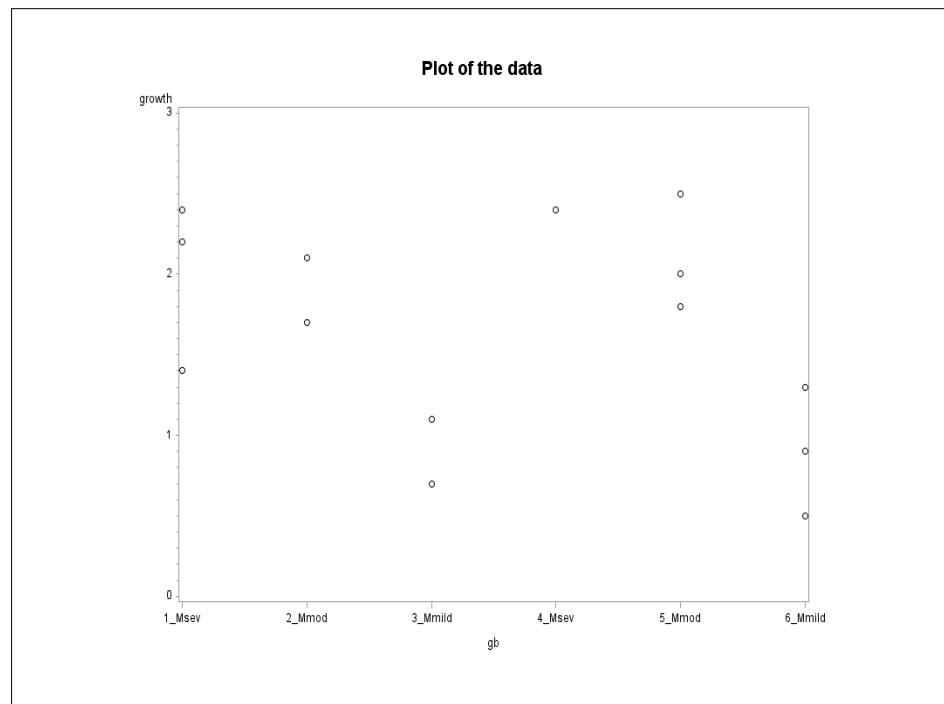
data a1; set a1;
if (gender eq 1)*(bone eq 1) then gb='1_Msev ';
if (gender eq 1)*(bone eq 2) then gb='2_Mmod ';
if (gender eq 1)*(bone eq 3) then gb='3_Mmild';
if (gender eq 2)*(bone eq 1) then gb='4_Msev ';
if (gender eq 2)*(bone eq 2) then gb='5_Mmod ';
if (gender eq 2)*(bone eq 3) then gb='6_Mmild';

title1 'Plot of the data';
symbol1 v=circle i=none c=black;
proc gplot data=a1;                      *****Scatterplot;
plot growth*gb/frame;
run;
```

```
proc means data=a1;
output out=a2 mean=avgrowth;
by gender bone;
title1 'Plot of the means';
symbol1 v='M' i=join c=black;          *****Interaction plot;
symbol2 v='F' i=join c=black;
proc gplot data=a2;
plot avgrowth*bone=gender/frame;
run;
```

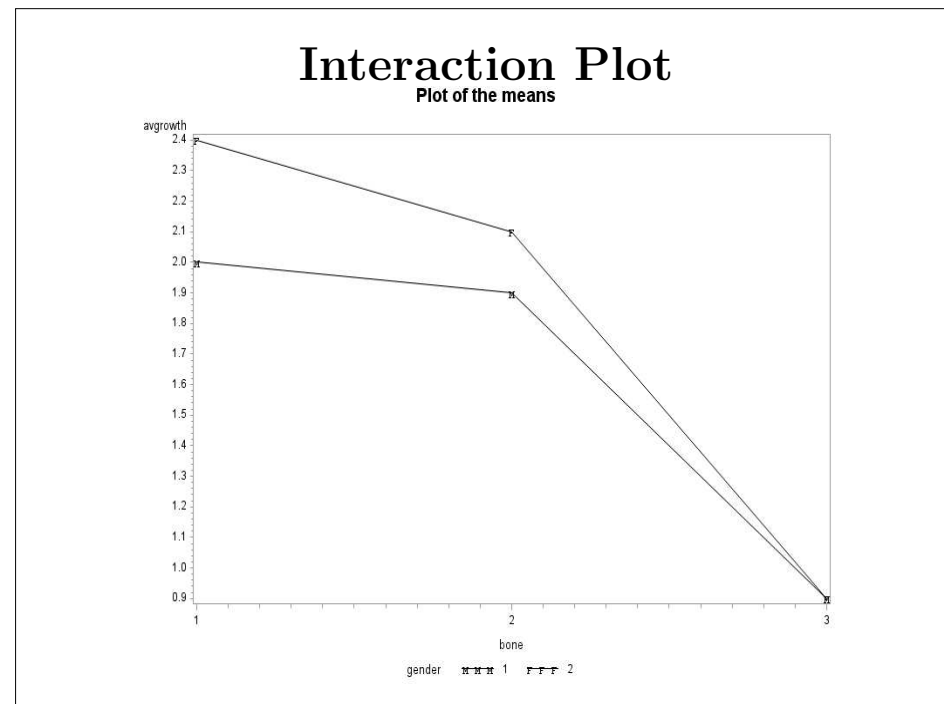
Output

Obs	growth	gender	bone
1	1.4	1	1
2	2.4	1	1
3	2.2	1	1
4	2.1	1	2
5	1.7	1	2
6	0.7	1	3
7	1.1	1	3
8	2.4	2	1
9	2.5	2	2
10	1.8	2	2
11	2.0	2	2
12	0.5	2	3
13	0.9	2	3
14	1.3	2	3



Topic 23

9



Topic 23

10

The Cell Means Model

- Expressed numerically

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

where μ_{ij} is the theoretical mean or expected value of all observations in cell (i, j)

- The ε_{ijk} are iid $N(0, \sigma^2)$ which implies the Y_{ijk} are independent $N(\mu_{ij}, \sigma^2)$
- Parameters
 - $\{\mu_{ij}\}$, $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$
 - σ^2

Topic 23

11

Estimates

- Estimate μ_{ij} by the sample mean of the observations in cell (i, j)

$$\hat{\mu}_{ij} = \bar{Y}_{ij}.$$

- For each cell (i, j) , also estimate of the variance

$$s_{ij}^2 = \sum (Y_{ijk} - \bar{Y}_{ij})^2 / (n_{ij} - 1)$$

- These s_{ij}^2 are pooled to estimate σ^2

Topic 23

12

SAS Commands

```
proc glm data=a1;
  class gender bone;
  model growth=gender|bone / solution;
  means gender*bone;

  contrast 'gender Type III'
    gender 3 -3    gender*bone 1 1 1 -1 -1 -1;
  contrast 'gender Type I'
    gender 7 -7    bone 2 -1 -1 gender*bone 3 2 2 -1 -3 -3;
  contrast 'bone Type III'
    bone 2 -2 0    gender*bone 1 -1 0 1 -1 0,
    bone 0 2 -2    gender*bone 0 1 -1 0 1 -1;
  contrast 'gender*bone Type I and III'
    gender*bone 1 -1 0 -1 1 0,
    gender*bone 0 1 -1 0 -1 1;
```

```
***Will now pool to increase error df***;

proc glm data=a1;
  class gender bone;
  model growth=gender bone/solution;
  means gender bone/ tukey lines;
  lsmeans gender bone/ adjust=tukey pdiff;
```

Parameter Estimates

- The **solution** option gives parameter estimates for the factor effects model under the GLM restriction
 - $\alpha_2 = 0$
 - $\beta_3 = 0$
 - $(\alpha\beta)_{13} = (\alpha\beta)_{23} = (\alpha\beta)_{21} = (\alpha\beta)_{22} = 0$
- Produces the cell means in the usual way

Output

		Sum of			
Source	DF	Squares	Mean Square	F Value	Pr > F
Model	5	4.47428571	0.89485714	5.51	0.0172
Error	8	1.30000000	0.16250000		
Corrected Total	13	5.77428571			
Source	DF	Type I SS	Mean Square	F Value	Pr > F
gender	1	0.00285714	0.00285714	0.02	0.8978
bone	2	4.39600000	2.19800000	13.53	0.0027
gender*bone	2	0.07542857	0.03771429	0.23	0.7980
Source	DF	Type III SS	Mean Square	F Value	Pr > F
gender	1	0.12000000	0.12000000	0.74	0.4152
bone	2	4.18971429	2.09485714	12.89	0.0031
gender*bone	2	0.07542857	0.03771429	0.23	0.7980

Parameter		Estimate	Std Error	t Value	Pr > t
Intercept		0.9000 B	0.23273733	3.87	0.0048
gender	1	-0.0000 B	0.36799004	-0.00	1.0000
gender	2	0.0000 B	.	.	.
bone	1	1.5000 B	0.46547467	3.22	0.0122
bone	2	1.2000 B	0.32914029	3.65	0.0065
bone	3	0.0000 B	.	.	.
gender*bone	1 1	-0.4000 B	0.59336610	-0.67	0.5192
gender*bone	1 2	-0.2000 B	0.52041650	-0.38	0.7108
gender*bone	1 3	0.0000 B	.	.	.
gender*bone	2 1	0.0000 B	.	.	.
gender*bone	2 2	0.0000 B	.	.	.
gender*bone	2 3	0.0000 B	.	.	.

Comments

- Type I and Type III SS's are different
 - Type I: $0.003 + 4.396 + 0.075 = 4.474 = \text{Model SS}$
 - Type III: $0.120 + 4.190 + 0.075 = 4.385 \neq \text{Model SS}$
- $A * B$ same because last term in model
- Type I and Type III hypotheses are different
- Most prefer Type III analysis
- Can be misleading if n_{ij} widely different
- Use contrasts to understand the difference

Contrast for $A * B$

- Same for Type I and Type III
- Null Hypothesis is that the mean profiles are parallel (recall interaction plot)
- Null hypothesis can be expressed

$$\begin{aligned}\mu_{12} - \mu_{11} &= \mu_{22} - \mu_{21} \\ \text{and} \\ \mu_{13} - \mu_{12} &= \mu_{23} - \mu_{22}\end{aligned}$$

```
contrast 'gender*bone Type I and III'
  gender*bone 1 -1 0 -1 1 0,
  gender*bone 0 1 -1 0 -1 1;
```

Type III Contrast for A

- Null Hypothesis is that the marginal gender means are the same
- Null hypothesis can be expressed

$$\begin{aligned}1\mu_{11} &= 1(\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}) \\ 1\mu_{12} &= 1(\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}) \\ 1\mu_{13} &= 1(\mu + \alpha_1 + \beta_3 + (\alpha\beta)_{13}) \\ -1\mu_{21} &= -1(\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}) \\ -1\mu_{22} &= -1(\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}) \\ -1\mu_{23} &= -1(\mu + \alpha_2 + \beta_3 + (\alpha\beta)_{23}) \\ \hline &= 3\alpha_1 - 3\alpha_2 + (\alpha\beta)_1 - (\alpha\beta)_2.\end{aligned}$$

```
contrast 'gender Type III'
  gender 3 -3  gender*bone 1 1 1 -1 -1 -1;
```

Type I Contrast for A

- Null Hypothesis is that the “weighted” marginal gender means are the same

$$\begin{aligned}
 3\mu_{11} &= 3(\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}) \\
 2\mu_{12} &= 2(\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}) \\
 2\mu_{13} &= 2(\mu + \alpha_1 + \beta_3 + (\alpha\beta)_{13}) \\
 -1\mu_{21} &= -1(\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}) \\
 -3\mu_{22} &= -3(\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}) \\
 -3\mu_{23} &= -3(\mu + \alpha_2 + \beta_3 + (\alpha\beta)_{23}) \\
 \hline
 &= 7\alpha_1 - 7\alpha_2 + 2\beta_1 - \beta_2 - \beta_3 + \\
 &\quad 3(\alpha\beta)_{11} + 2(\alpha\beta)_{12} + 2(\alpha\beta)_{13} - \\
 &\quad (\alpha\beta)_{21} - 3(\alpha\beta)_{22} - 3(\alpha\beta)_{23}
 \end{aligned}$$

```
contrast 'gender Type I'
  gender 7 -7   bone 2 -1 -1
  gender*bone 3 2 2 -1 -3 -3;
```

Output

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
gender Type III	1	0.12000000	0.12000000	0.74	0.4152
gender Type I	1	0.00285714	0.00285714	0.02	0.8978
bone Type III	2	4.18971429	2.09485714	12.89	0.0031
gender*bone	2	0.07542857	0.03771429	0.23	0.7980

Parameter	Estimate	Standard		Pr > t
		Error	t Value	
gender Type III	-0.60000000	0.69821200	-0.86	0.4152
gender Type I	0.20000000	1.50831031	0.13	0.8978

Interaction Not Significant

- Determine whether pooling is beneficial
- If yes, rerun analysis without interaction
- Check significance of main effects
- If factor insignificant, determine whether pooling is beneficial
- If yes, rerun analysis as one-way ANOVA
- If statistically significant factor has more than two levels, use multiple comparison procedure to assess differences
- Contrasts and linear combinations can also be used

Output

Source	DF	Sum of		F Value	Pr > F
		Squares	Mean Square		
Model	3	4.39885714	1.46628571	10.66	0.0019
Error	10	1.37542857	0.13754286		
Corrected Total	13	5.77428571			

R-Square	Coeff Var	Root MSE	growth Mean
0.761801	22.57456	0.370868	1.642857

Source	DF	Type I SS	Mean Square	F Value	Pr > F
gender	1	0.00285714	0.00285714	0.02	0.8883
bone	2	4.39600000	2.19800000	15.98	0.0008

Source	DF	Type III SS	Mean Square	F Value	Pr > F
gender	1	0.09257143	0.09257143	0.67	0.4311
bone	2	4.39600000	2.19800000	15.98	0.0008

Output

*** MEANS STATEMENT ***
Tukey's Studentized Range (HSD) Test for growth

Alpha	0.05
Error Degrees of Freedom	10
Error Mean Square	0.137543
Critical Value of Studentized Range	3.15106
Minimum Significant Difference	0.4417

	Mean	N	gender
A	1.6571	7	1
A	1.6286	7	2

Alpha	0.05
Error Degrees of Freedom	10
Error Mean Square	0.137543
Critical Value of Studentized Range	3.87676
Minimum Significant Difference	0.6692
Harmonic Mean of Cell Sizes	4.615385

	Mean	N	bone
A	2.1000	4	1
A	2.0200	5	2
B	0.9000	5	3

Output

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

HO:LSMean1=		
gender	growth LSMEAN	LSMean2 Pr > t
1	1.59047619	0.4311
2	1.76190476	

growth		LSMEAN
bone	LSMEAN	Number
1	2.14285714	1
2	2.00285714	2
3	0.88285714	3

Least Squares Means for effect bone
Pr > |t| for HO: LSMean(i)=LSMean(j)

Dependent Variable: growth			
i/j	1	2	3
1		0.8538	0.0017
2	0.8538		0.0020
3	0.0017	0.0020	

Comments

- The means and lsmeans not the same
 - means : raw sample mean - similar to a weighted average of cell means (Type I)
 - lsmeans: uses parameter estimates - similar to Type III approach
- lsmeans most commonly used

Background Reading

- KNNL Chapter 23
- knnl954.sas
- KNNL Chapter 24