Topic 24 - Multifactor Studies

STAT 525 - Fall 2013

Outline

- 3-factor studies
  - Data
  - Model
  - Parameter Estimates
  - Inference
- Unequal sample size

Data for Three Factor ANOVA

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

Example Page 1005

- Influence of several factors on exercise tolerance
- Considered age range : 25-35 years old
- Three factors were
  - Gender $(a = 2)$
  - Percent body fat $(b = 2)$
  - Smoking history $(c = 2)$
- Y is exercise tolerance (minutes until fatigue) when doing a bicycle test
- Set up as balanced design $(n = 3)$

General Plan

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

SAS Commands

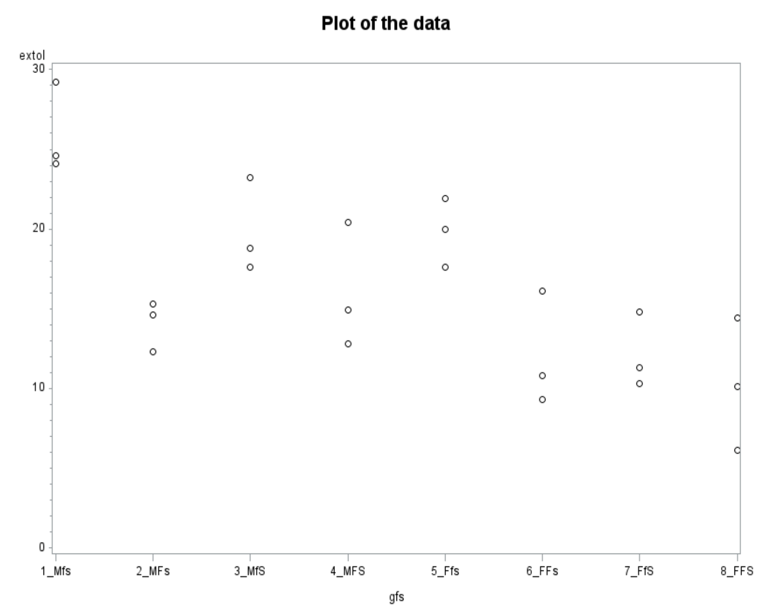
```
data a1; infile 'u:\.www\datasets525\CH24TA04.txt';
   input extol gender fat smoke;
proc print;

data a1; set a1;
   if (gender eq 1)*(fat eq 1)*(smoke eq 1) then gfs='1_Mfs';
   if (gender eq 1)*(fat eq 2)*(smoke eq 1) then gfs='2_MFs';
   if (gender eq 1)*(fat eq 1)*(smoke eq 2) then gfs='3_MfS';
   if (gender eq 1)*(fat eq 2)*(smoke eq 2) then gfs='4_MFS';
   if (gender eq 2)*(fat eq 1)*(smoke eq 1) then gfs='5_Ffs';
   if (gender eq 2)*(fat eq 2)*(smoke eq 1) then gfs='6_FFs';
   if (gender eq 2)*(fat eq 1)*(smoke eq 2) then gfs='7_FfS';
   if (gender eq 2)*(fat eq 2)*(smoke eq 2) then gfs='8_FFS';

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


Output

Obs	extol	gender	fat	smoke
1	24.1	1	1	1
2	29.2	1	1	1
3	24.6	1	1	1
4	20.0	2	1	1
5	21.9	2	1	1
6	17.6	2	1	1
7	14.6	1	2	1
8	15.3	1	2	1
9	12.3	1	2	1
10	16.1	2	2	1
11	9.3	2	2	1
12	10.8	2	2	1
13	17.6	1	1	2
14	18.8	1	1	2
15	23.2	1	1	2
16	14.8	2	1	2
17	10.3	2	1	2
18	11.3	2	1	2
19	14.9	1	2	2
20	20.4	1	2	2
21	12.8	1	2	2
22	10.1	2	2	2
23	14.4	2	2	2
24	6.1	2	2	2



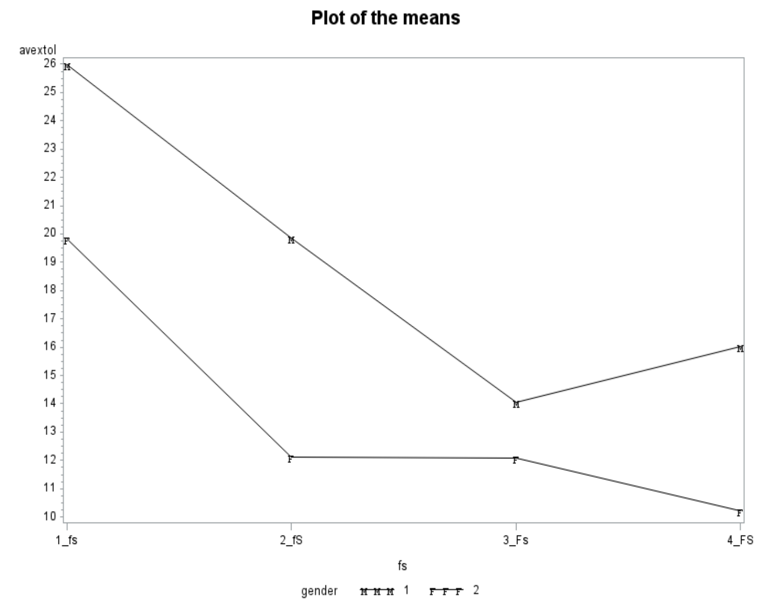
SAS Commands

```
proc sort data=a1; by gender fat smoke;

proc means data=a1;
   output out=a2 mean=avextol;
   by gender fat smoke;

data a2; set a2;
   if (fat eq 1)*(smoke eq 1) then fs='1_fs';
   if (fat eq 1)*(smoke eq 2) then fs='2_fS';
   if (fat eq 2)*(smoke eq 1) then fs='3_Fs';
   if (fat eq 2)*(smoke eq 2) then fs='4_FS';
run;

proc sort data=a2; by fs; title1 'Plot of the means';
symbol1 v='M' i=join c=black; symbol2 v='F' i=join c=black;
proc gplot data=a2;
   plot avextol*fs=gender/frame;
run;
```

The Cell Means Model

- Expressed numerically

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

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

- The ε_{ijkl} are iid $N(0, \sigma^2)$ which implies the Y_{ijkl} are independent $N(\mu_{ijk}, \sigma^2)$

- Parameters
  - $\{\mu_{ijk}\}$, $i = 1, 2, .., a$, $j = 1, 2, ..., b$, $k = 1, 2, ..., c$
  - σ^2

Estimates

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

$$\hat{\mu}_{ijk} = \overline{Y}_{ijk.}$$

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

$$s^2_{ijk} = \sum (Y_{ijkl} - \overline{Y}_{ijk.})^2/(n_{ijk} - 1)$$

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

Factor Effects Model

- Statistical model is

$$\begin{aligned} Y_{ijkl} &= \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \\ &\quad (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl} \end{aligned}$$

  μ - grand mean

  $\alpha_i, \beta_j, \gamma_k$ - main effects of A, B, and C

  $(\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}$ are the two-factor (first-order) interactions

  $(\alpha\beta\gamma)_{ijk}$ is the three-factor (second-order) interaction

- Over-parameterized model.
- Extension of usual model constraints.

ANOVA Table

- Sources of variation are three main effects, three first-order interactions, one second-order interaction, and error
- If balanced
  - SS add up to model SS
  - Type I and Type III the same
- Each effect tested over MSE

SAS Commands

```
proc glm data=a1;
   class gender fat smoke;
   model extol=gender fat smoke
     gender*fat gender*smoke fat*smoke
     gender*fat*smoke;
   means gender*fat*smoke;
run;
```


Output

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	588.5829167	84.0832738	9.01	0.0002
Error	16	149.3666667	9.3354167		
Corrected Total	23	737.9495833			

R-Square	Coeff Var	Root MSE	extol Mean
0.797592	18.77833	3.055391	16.27083

Source	DF	Type I SS	Mean Square	F Value	Pr > F
gender	1	176.5837500	176.5837500	18.92	0.0005
fat	1	242.5704167	242.5704167	25.98	0.0001
smoke	1	70.3837500	70.3837500	7.54	0.0144
gender*fat	1	13.6504167	13.6504167	1.46	0.2441
gender*smoke	1	11.0704167	11.0704167	1.19	0.2923
fat*smoke	1	72.4537500	72.4537500	7.76	0.0132
gender*fat*smoke	1	1.8704167	1.8704167	0.20	0.6604

Comments

- First examine interactions
- Some options when one or more interactions significant
  - Interpret the plot of means
  - Run analyses for each level of one factor (slice)
  - Run as one-way with abc levels
  - Run as two-way with a and bc levels
  - Use contrasts
- If no interactions
  - Use contrasts
  - Multiple comparison procedure

SAS Commands

```
data a1; set a1;
   if (fat eq 1)*(smoke eq 1) then fs='1_fs';
   if (fat eq 1)*(smoke eq 2) then fs='2_fS';
   if (fat eq 2)*(smoke eq 1) then fs='3_Fs';
   if (fat eq 2)*(smoke eq 2) then fs='4_FS';
run;

proc glm data=a1;
   class gender fs;
   model extol=gender fs;
   means gender fs/tukey;
run;
```


Output

```
                           Sum of
Source           DF       Squares  Mean Square  F Value  Pr > F
Model             4   561.9916667  140.4979167    15.17  <.0001
Error            19   175.9579167    9.2609430
Corrected Total  23   737.9495833

R-Square     Coeff Var      Root MSE    extol Mean
0.761558      18.70328      3.043180      16.27083

Source           DF    Type I SS  Mean Square  F Value  Pr > F
gender            1  176.5837500  176.5837500    19.07  0.0003
fs                3  385.4079167  128.4693056    13.87  <.0001

Source           DF  Type III SS  Mean Square  F Value  Pr > F
gender            1  176.5837500  176.5837500    19.07  0.0003
fs                3  385.4079167  128.4693056    13.87  <.0001
```


Output

```
Tukey's Studentized Range (HSD) Test for extol

Alpha                                     0.05
Error Degrees of Freedom                    19
Error Mean Square                     9.260943
Critical Value of Studentized Range    3.97655
Minimum Significant Difference          4.9404

           Mean      N    fs

A        22.900      6    1_fs

B        16.000      6    2_fS
B
B        13.117      6    4_FS
B
B        13.067      6    3_Fs
```

Unequal Sample Size

- Similar approach as two-way ANOVA
- Type I and Type III SS different
- Type III more commonly used
- lsmeans used for comparsions

Background Reading

- KNNL Chapter 24
- knnl1005.sas
- KNNL Section 25.1