Lecture 15

- Three-way ANOVA cells
  - Data
  - Model
  - Inference

Data for three-way ANOVA

- Y, the response variable
- Factor A with levels i = 1 to I
- Factor B with levels j = 1 to J
- Factor C with levels k = 1 to K
- Y_{ijkl} is the l^{th} observation in cell (i,j,k), l = 1 to L
- We can have unbalanced designs with L = L_{ijk}

KNNL Example

- KNNL p 1004
- Y is exercise tolerance, minutes until fatigue on a bicycle test
- A is gender, I=2 levels: male, female
- B is percent body fat, J=2 levels: low, high
- C is smoking history, K=2 levels: light, heavy
- L=3 persons aged 25-35 per (i,j,k) cell

Read and check the data

```
data a1;
  infile '…/CH24TA04.txt';
  input extol gender fat smoke;
proc print data=a1;
run;
```

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
. . .				
24	6.1	2	2	2

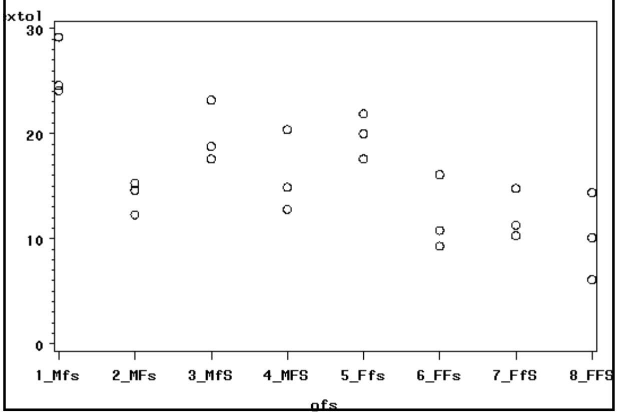
Prepare the data for a plot

```
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';
run;
```

Plot the data

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

Plot of the data



Find the means

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

Define fat*smoke

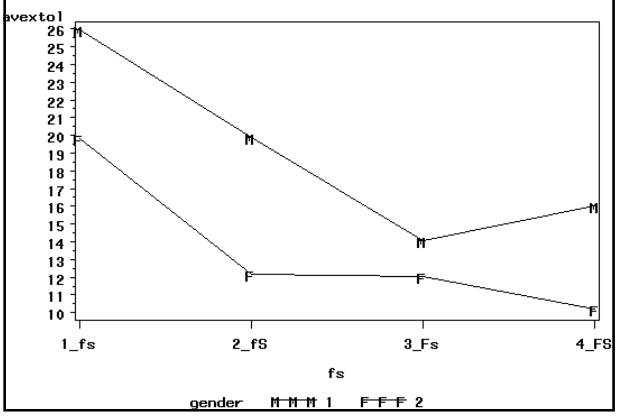
```
data a2; set a2;
   if (fat eq 1)*(smoke eq 1)
     then fs='1_fs';
   if (fat eq 1)*(smoke eq 2)
     then fs='2_fS';
   if (fat eq 2)*(smoke eq 1)
     then fs='3_Fs';
   if (fat eq 2)*(smoke eq 2)
     then fs='4_FS';
```

Plot the means

```
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;
run;
```

Obs	gen	fat	smoke	FR	avextol	fs
1	1	1	1	3	25.97	1_fs
2	2	1	1	3	19.83	1_fs
3	1	1	2	3	19.87	2_fS
4	2	1	2	3	12.13	2_fS
5	1	2	1	3	14.07	3_Fs
6	2	2	1	3	12.07	3_Fs
7	1	2	2	3	16.03	4_FS
8	2	2	2	3	10.20	4_FS

Plot of the means



Cell means model

- $Y_{ijkl} = \mu_{ijk} + \xi_{ijkl}$
  - where μ_{ijk} is the theoretical mean or expected value of all observations in cell (i,j,k)
  - the ξ_{ijkl} are iid $N(0, \sigma^2)$
  - $Y_{ijkl} \sim N(\mu_{ijk}, \sigma^2)$, independent

Estimates

- Estimate μ_{ijk} by the mean of the observations in cell (i,j,k), $Y_{ijk.}$
- $Y_{ijk.} = (\Sigma_l Y_{ijkl})/L_{ijk}$
- For each (i,j,k) combination, we can get an estimate of the variance
- $s_{ijk}^2 = (\Sigma(Y_{ijkl} - Y_{ijk.})^2)/(L_{ijk}-1)$
- We need to combine these to get an estimate of σ^2

Pooled estimate of σ^2

- We pool the s_{ijk}^2, giving weights proportional to the df, L_{ijk} -1
- The pooled estimate is
- $s^2 = (\Sigma (L_{ijk}-1)s_{ijk}^2) / (\Sigma (L_{ijk}-1))$

Factor effects model

- $\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$
- μ is the overall mean
- α_i, β_j, and γ_k are the main effects of A, B, and C
- $\alpha\beta_{ij}$, $\alpha\gamma_{ik}$, and $\beta\gamma_{jk}$ are the two-way interactions
- $\alpha\beta\gamma_{ijk}$ is the three-way interaction
- The usual constraints apply

Anova table

- Source of variation are the three main effects, the three two-way interactions, the three-way interaction, and error
- With balanced data the SS and DF add to the total
- Each effect is tested by an F statistic with MSE in the denominator

Run proc glm

```
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;
```

Model and error output

```
Source DF     MS    F      P
Model   7  84.08 9.01 0.0002
Error  16   9.33
Total  23
```

Factor effects output

```
Source            DF       F P
gender            1 18.92 0.0005
fat               1 25.98 0.0001
smoke             1  7.54 0.0144
gender*fat        1  1.46 0.2441
gender*smoke      1  1.19 0.2923
fat*smoke         1  7.76 0.0132
gender*fat*smoke  1  0.20 0.6604
```

Analytical Strategy

- First examine interactions
- Some options when one or more interactions are significant
  - Interpret the plot of means
  - Run analyses for each level of one factor, eg run A|B by C
  - Run as a one-way with IJK levels
  - Define a composite factor by combining two factors, eg AB with IJ levels

Analytical Strategy (2)

- Some options when no interactions are significant
  - Rerun without the interactions
  - Use a multiple comparison procedure for the main effects

Combine fat and smoke

```
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 glm

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

Model and error output

Source	DF	MS	F	P
Model	4	140.49	15.17	<.0001
Error	19	9.26		
Total	23			

Factor effects output

Source	DF	F Value	Pr > F
gender	1	19.07	0.0003
fs	3	13.87	<.0001

Means for gender

	Mean	N	gender
A	18.983	12	1
B	13.558	12	2

Tukey comparisons for fs

	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