Topic 26 - Two-Way Random Effects

STAT 525 - Fall 2013

Outline

- Two-way Random effects
  - Data
  - Model
  - Variance component estimation
  - F-tests

Data for Two-Way Design

- Same data structure as fixed design
- Y is the response variable
- Factor A has levels $i = 1, 2, ..., a$
- Factor B has levels $j = 1, 2, ..., b$
- Y_{ijk} is the k^{th} observation of cell (i, j) with $k = 1, 2, ..., n_{ij}$
- Balanced when $n_{ij} = n$

Example Page 1080

- Interested in assessing the fuel efficiency (mpg) of a specific model of car
- Two **random** factors
  - Factor A: Driver
  - Factor B: Car
- How much of the overall variability is due to driver and/or car?
- Each driver drove each car twice ($n = 2$) over same 40 mile course

SAS Commands

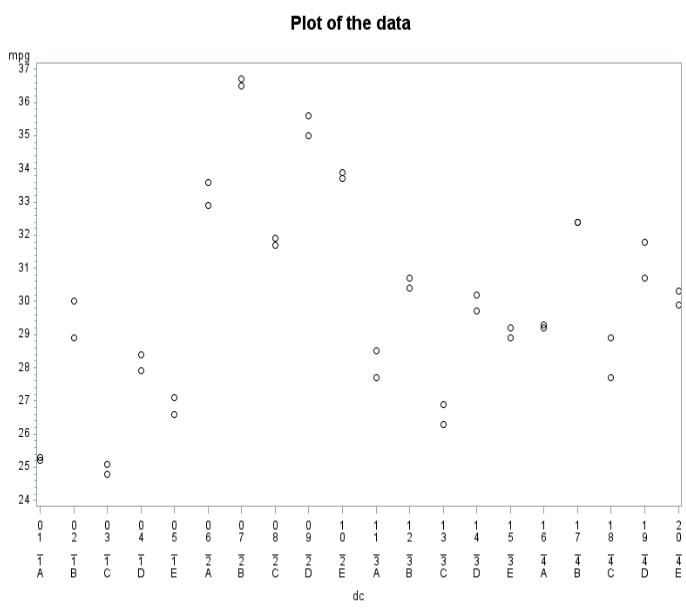
```
data a1; infile 'u:\.www\datasets525\CH25PR15.txt';
   input mpg driver car;
proc print data=a1; run;

data a1; set a1;
   if (driver eq 1)*(car eq 1) then dc='01_1A';
   if (driver eq 1)*(car eq 2) then dc='02_1B';
   if (driver eq 1)*(car eq 3) then dc='03_1C';
   if (driver eq 1)*(car eq 4) then dc='04_1D';
   if (driver eq 1)*(car eq 5) then dc='05_1E';
   if (driver eq 2)*(car eq 1) then dc='06_2A';
   if (driver eq 2)*(car eq 2) then dc='07_2B';
    :
   if (driver eq 4)*(car eq 4) then dc='19_4D';
   if (driver eq 4)*(car eq 5) then dc='20_4E';

proc gplot data=a1;
   plot mpg*dc/frame; run;
```


Output

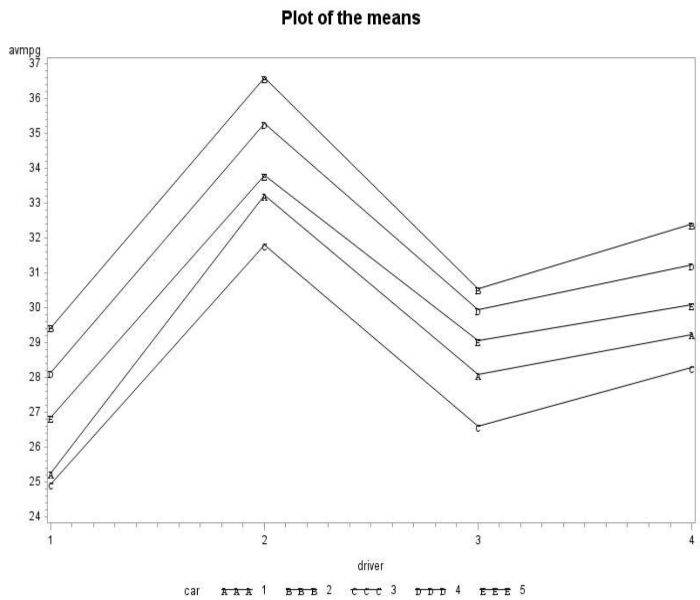
Obs	mpg	driver	car
1	25.3	1	1
2	25.2	1	1
3	28.9	1	2
4	30.0	1	2
5	24.8	1	3
6	25.1	1	3
7	28.4	1	4
8	27.9	1	4
9	27.1	1	5
10	26.6	1	5
11	33.6	2	1
12	32.9	2	1
:	:	:	:
37	31.8	4	4
38	30.7	4	4
39	30.3	4	5
40	29.9	4	5



SAS Commands

```
proc means data=a1;
   output out=a2 mean=avmpg;
   var mpg;
   by driver car;

title1 'Plot of the means';
symbol1 v='A' i=join c=black;
symbol2 v='B' i=join c=black;
symbol3 v='C' i=join c=black;
symbol4 v='D' i=join c=black;
symbol5 v='E' i=join c=black;
proc gplot data=a2;
   plot avmpg*driver=car/frame;
run;
```



Random Effects Model

- Expressed as

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

$$\mu_{ij} \sim N(\mu, \sigma^2_\mu)$$

$$\varepsilon_{ijk} \sim N(0, \sigma^2)$$

$$\mu_{ik} \text{ and } \varepsilon_{ijk} \text{ independent}$$

- Not all observations independent
- Will separate mean variances into factor variances

Random Factor Effects Model

- Expressed as

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

$$\mu \text{ - overall mean}$$

$$\alpha_i \sim N(0, \sigma^2_\alpha)$$

$$\beta_j \sim N(0, \sigma^2_\beta)$$

$$(\alpha\beta)_{ij} \sim N(0, \sigma^2_{\alpha\beta})$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

- There are <u>FOUR</u> parameters/variances

Covariance Structure

- Covariances:

$$\begin{aligned}
Cov(Y_{ijk}, Y_{ijk}) &= \sigma^2 + \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} \\
Cov(Y_{ijk}, Y_{ijk^*}) &= \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} \\
Cov(Y_{ijk}, Y_{ij^*k}) &= \sigma^2_\alpha \\
Cov(Y_{ijk}, Y_{i^*jk}) &= \sigma^2_\beta \\
Cov(Y_{ijk}, Y_{i^*j^*k}) &= 0
\end{aligned}$$

- Can look at percentage of variability due to factors
- Could look at percentage of total variability or percentage of cell means variability (i.e., ignoring error variance).
- Approach to confidence intervals same as before

ANOVA Table

- Terms and layout of ANOVA table the same as that used in the fixed effects case
- The expected means squares (EMS) are different because of the additional random effects
- Results in different F tests
- Use EMS as guide for tests $\longrightarrow$ determine denominator MS

Expected Mean Squares

- Same partition of Total Sum of Squares
- Assuming balanced design

  E(MSE)=σ^2

  E(MSAB)=$\sigma^2 + n\sigma^2_{\alpha\beta}$

  E(MSB)=$\sigma^2 + n\sigma^2_{\alpha\beta} + an\sigma^2_{\beta}$

  E(MSA)=$\sigma^2 + n\sigma^2_{\alpha\beta} + bn\sigma^2_{\alpha}$

- Estimates of variances can be obtained from these equations or other methods

Model Estimates

- Using mean squares (ANOVA estimates)

  $\hat{\sigma}^2 = \text{MSE}$

  $\hat{\sigma}^2_{\alpha\beta} = (\text{MSAB} - \text{MSE})/\text{n}$

  $\hat{\sigma}^2_{\beta} = (\text{MSB} - \text{MSAB})/\text{an}$

  $\hat{\sigma}^2_{\beta} = (\text{MSA} - \text{MSAB})/\text{bn}$

- Estimates can be negative
- Similar adjustments used

Hypothesis Tests

- Three tests of variance

  H_{0A}: $\sigma^2_{\alpha} = 0$ vs H_{1A}: $\sigma^2_{\alpha} > 0$

  H_{0B}: $\sigma^2_{\beta} = 0$ vs H_{1B}: $\sigma^2_{\beta} > 0$

  H_{0AB}: $\sigma^2_{\alpha\beta} = 0$ vs H_{1AB}: $\sigma^2_{\alpha\beta} > 0$

- **No hierarchy** in terms of testing
- Not all tests use MSE in denominator
- To test σ^2_{α} or σ^2_{β} use MSAB
- Will alter denominator DF too

SAS Commands

```
proc glm data=a1;
   class driver car;
   model mpg=driver car driver*car;
   random driver car driver*car/test;
run;

proc mixed data=a1 cl;
   class car driver;
   model mpg=;
   random car driver car*driver/vcorr;
run;
```


Output

```
                          Sum of
Source             DF       Squares  Mean Square  F Value  Pr > F
Model              19   377.4447500   19.8655132   113.03  <.0001
Error              20     3.5150000    0.1757500
Corrected Total    39   380.9597500

R-Square     Coeff Var      Root MSE      mpg Mean
0.990773      1.395209      0.419225      30.04750

Source             DF     Type I SS  Mean Square  F Value  Pr > F
driver              3   280.2847500   93.4282500   531.60  <.0001
car                 4    94.7135000   23.6783750   134.73  <.0001
driver*car         12     2.4465000    0.2038750     1.16  0.3715

Source       Type III Expected Mean Square
driver       Var(Error) + 2 Var(driver*car) + 10 Var(driver)
car          Var(Error) + 2 Var(driver*car) + 8 Var(car)
driver*car   Var(Error) + 2 Var(driver*car)
```


Output

```
Proc GLM assumes all factors are fixed effects...in ANOVA table
all terms tested over MSE

With random statement and test option, will perform tests based
on EMS

Source    DF  Type III SS  Mean Square  F Value  Pr > F
driver     3   280.284750    93.428250   458.26  <.0001
car        4    94.713500    23.678375   116.14  <.0001
Error     12     2.446500     0.203875
Error: MS(driver*car)


Source           DF Type III SS Mean Square  F Value  Pr > F
driver*car       12    2.446500    0.203875     1.16  0.3715
Error: MS(Error) 20    3.515000    0.175750
```


Output

```
                  Model Information
Data Set                     WORK.A1
Dependent Variable           mpg
Covariance Structure         Variance Components
Estimation Method            REML

            Dimensions
Covariance Parameters            4
Columns in X                     1
Columns in Z                    29
Subjects                         1
Max Obs Per Subject             40
Total Observations              40

                     Iteration History
Iteration    Evaluations    -2 Res Log Like       Criterion
        0              1       203.25223618
        1              1        86.77908149      0.00000000
                   Convergence criteria met.
```

Output

```
Estimated V Correlation Matrix      ***Summarized entries***
Same observation: 1.0000
Same i and j    : 0.9859
Same j          : 0.7490
Same i          : 0.2358

             Covariance Parameter Estimates
Cov Parm       Estimate     Alpha        Lower       Upper
car              2.9343      0.05       1.0464     24.9038
driver           9.3224      0.05       2.9864      130.79
car*driver      0.01406      0.05     0.001345    3.592E17
Residual         0.1757      0.05       0.1029      0.3665

           Fit Statistics
-2 Res Log Likelihood            86.8
AIC (smaller is better)          94.8
AICC (smaller is better)         96.0
BIC (smaller is better)          93.2
```


Background Reading

- KNNL Section 25.2-25.6
- knnl1080.sas
- KNNL Sections 25.2-25.6