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 $i = 1, 2, \ldots, a$
- Factor $B$ has levels $j = 1, 2, \ldots, b$
- $Y_{ijk}$ is the $k^{th}$ observation of cell $(i, j)$ with $k = 1, 2, \ldots, n_{ij}$
- 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\CH2SPR15.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 2)*(car eq 3) then dc='08_2C';
    if (driver eq 2)*(car eq 4) then dc='09_2D';
    if (driver eq 2)*(car eq 5) then dc='10_2E';
    if (driver eq 3)*(car eq 1) then dc='11_3A';
    if (driver eq 3)*(car eq 2) then dc='12_3B';
    if (driver eq 3)*(car eq 3) then dc='13_3C';
    if (driver eq 3)*(car eq 4) then dc='14_3D';
    if (driver eq 3)*(car eq 5) then dc='15_3E';
    if (driver eq 4)*(car eq 1) then dc='16_4A';
    if (driver eq 4)*(car eq 2) then dc='17_4B';
    if (driver eq 4)*(car eq 3) then dc='18_4C';
    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;

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

\[ Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \]

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

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

\( \mu_k \) and \( \varepsilon_{ijk} \) independent

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

Random Factor Effects Model

- Expressed as

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

\( \mu \) - 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 FOUR parameters/variances

Covariance Structure

- Covariances:

\[ Cov(Y_{ijk}, Y_{ijk'}) = \sigma^2 + \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} \]

\[ Cov(Y_{ijk}, Y_{ijk*}) = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} \]

\[ Cov(Y_{ijk}, Y_{ij*k}) = \sigma^2_\alpha \]

\[ Cov(Y_{ijk}, Y_{i*j*k}) = \sigma^2_\beta \]

\[ Cov(Y_{ijk}, Y_{i*j*k}) = 0 \]

- 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 → determine denominator MS

### Expected Mean Squares

- Same partition of Total Sum of Squares
- Assuming balanced design
  
  $E(MSE) = \sigma^2$
  
  $E(MSAB) = \sigma^2 + n\sigma_{\alpha\beta}^2$
  
  $E(MSB) = \sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_{\beta}^2$
  
  $E(MSA) = \sigma^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_{\alpha}^2$

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

### Model Estimates

- Using mean squares (ANOVA estimates)
  
  $\hat{\sigma}^2 = MSE$
  
  $\hat{\sigma}_{\alpha\beta}^2 = (MSAB - MSE)/n$
  
  $\hat{\sigma}_{\beta}^2 = (MSB - MSAB)/an$
  
  $\hat{\sigma}_{\alpha}^2 = (MSA - MSAB)/bn$

- Estimates can be negative
- Similar adjustments used

### Hypothesis Tests

- Three tests of variance
  
  $H_{0A}: \sigma_{\alpha}^2 = 0$ vs $H_{1A}: \sigma_{\alpha}^2 > 0$
  
  $H_{0B}: \sigma_{\beta}^2 = 0$ vs $H_{1B}: \sigma_{\beta}^2 > 0$
  
  $H_{0AB}: \sigma_{\alpha\beta}^2 = 0$ vs $H_{1AB}: \sigma_{\alpha\beta}^2 > 0$

- **No hierarchy** in terms of testing
- Not all tests use MSE in denominator
- To test $\sigma_{\alpha}^2$ or $\sigma_{\beta}^2$ use MSAB
- Will alter denominator DF too
SAS Commands

```sas
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

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

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>driver</td>
<td>3</td>
<td>280.284750</td>
<td>93.428250</td>
<td>458.26</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>car</td>
<td>4</td>
<td>94.713500</td>
<td>23.678375</td>
<td>116.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>12</td>
<td>2.446500</td>
<td>0.203875</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Error: MS(driver*car)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>driver*car</td>
<td></td>
<td>2.446500</td>
<td>0.203875</td>
<td>1.16</td>
<td>0.3715</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>3.515000</td>
<td>0.175750</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Error: MS(Error)

Output

Model Information

<table>
<thead>
<tr>
<th>Data Set</th>
<th>WORK.A1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
<td>mpg</td>
</tr>
<tr>
<td>Covariance Structure</td>
<td>Variance Components</td>
</tr>
<tr>
<td>Estimation Method</td>
<td>REML</td>
</tr>
</tbody>
</table>

Dimensions

<table>
<thead>
<tr>
<th>Covariance Parameters</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columns in X</td>
<td>1</td>
</tr>
<tr>
<td>Columns in Z</td>
<td>29</td>
</tr>
<tr>
<td>Subjects</td>
<td>1</td>
</tr>
<tr>
<td>Max Obs Per Subject</td>
<td>40</td>
</tr>
<tr>
<td>Total Observations</td>
<td>40</td>
</tr>
</tbody>
</table>

Iteration History

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Evaluations</th>
<th>-2 Res Log Like</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>203.25223618</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>86.77908149</td>
<td>0.00000000</td>
</tr>
</tbody>
</table>

Convergence criteria met.
Output

Estimated V Correlation Matrix

Same observation: 1.0000
Same i and j : 0.9859
Same j : 0.7490
Same i : 0.2358

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Estimate</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>2.9343</td>
<td>0.05</td>
<td>1.0464</td>
<td>24.9038</td>
</tr>
<tr>
<td>driver</td>
<td>9.3224</td>
<td>0.05</td>
<td>2.9864</td>
<td>130.79</td>
</tr>
<tr>
<td>car*driver</td>
<td>0.01406</td>
<td>0.05</td>
<td>0.001345</td>
<td>3.592E17</td>
</tr>
<tr>
<td>Residual</td>
<td>0.1757</td>
<td>0.05</td>
<td>0.1029</td>
<td>0.3665</td>
</tr>
</tbody>
</table>

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