Random Effects in CRD

Design of Experiments - Montgomery
Section 12-1

Random Effects vs Fixed Effects

- Consider factor with numerous possible levels
- Want to draw inference on population of levels
- Not concerned with any specific levels
- Example of difference (1=fixed, 2=random)
  1. Compare reading ability of 10 2nd grade classes in NY
  2. Compare variability among all 2nd grade classes in NY
     1. Select $a = 10$ specific classes of interest. Randomly choose $n$ students from each classroom. Want to compare $\tau_i$ (class-specific effects).
     2. Randomly choose $a = 10$ classes from large number of classes. Randomly choose $n$ students from each classroom. Want to assess $\sigma^2$ (class to class variability).

- Inference broader in random effects case
- Levels chosen randomly $\rightarrow$ inference on population

Random Effects Model (CRD)

- Same model as in the fixed case

  $$y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad \{ i = 1, 2 \ldots a \} \quad \{ j = 1, 2 \ldots n_i \}$$

  $\mu$ - grand mean
  $\tau_i$ - $i$th treatment effect
  $\epsilon_{ij} \sim N(0, \sigma^2)$

  But view number of treatment levels as infinite
- Instead of $\sum \tau_i = 0$, assume
  $\tau_i \sim N(0, \sigma^2)$
  $\{ \tau_i \}$ and $\{ \epsilon_{ij} \}$ independent
- $\text{Var}(y_{ij}) = \sigma^2 + \sigma^2$ (HW #1 Problem 4)

Random Effects Model

- The hypotheses are:
  $$H_0 : \sigma^2 = 0$$
  $$H_1 : \sigma^2 > 0$$

- Partitioning of Total Sum of Squares identical
  $$E(\text{MSE}) = \sigma^2$$
  $$E(\text{MS} \text{Treatment}) = \sigma^2 + n_\sigma^2$$
- Under $H_0$, $F_0 \sim F_{a,a-1,N-a}$

- Same test as before
- Direct comparison of variabilities (between vs within)
- Conclusions, however, pertain to entire population
Model Estimates

- Usually interested in estimating variances
- Use mean squares (known as ANOVA method)

\[ \hat{\sigma}^2 = \text{MS}_E \]

\[ \hat{\sigma}^2 = (\text{MS}_{\text{Treatment}} - \text{MS}_E)/n \]

If unbalanced, replace \( n \) with \( n_0 = ((\sum n_i)^2 - \sum n_i^2)/((a-1)\sum n_i) \)

- Estimate of \( \hat{\sigma}^2 \) can be negative
  - Supports \( H_0 \)? Use zero as estimate?
  - Validity of model? Nonlinear?
  - Bayesian approach (nonnegative prior)

Confidence intervals

- \( \sigma^2 \): Same as fixed case

\[ \frac{(N-a)\text{MS}_E}{\sigma^2} \sim \chi^2_{N-a} \]

\[ \frac{(N-a)\text{MS}_E}{\chi^2_{a/2,N-a}} \leq \sigma^2 \leq \frac{(N-a)\text{MS}_E}{\chi^2_{1-a/2,N-a}} \]

- \( \sigma^2 \): Linear combination of \( \chi^2 \)

\[ \frac{(a-1)\text{MS}_{\text{Treat}}}{\sigma^2 + n\sigma_i^2} \sim \chi^2_{a-1} \]

so

\[ f(\sigma^2) = \frac{\sigma^2 + n\sigma_i^2}{n(a-1)}\chi^2_{a-1} - \frac{\sigma^2}{a(N-a)}\chi^2_{N-a} \]

No closed form expression for this distribution

Approximations available (Section 12-7)

Example

A supplier delivers several hundred batches of raw material to a company each year. The company is interested in a high yield from each batch of raw material (percent usable). Therefore, to investigate the consistency of this supplier, an experiment is done where five batches were selected at random and three yield determinations were made on each batch.

<table>
<thead>
<tr>
<th>Batch</th>
<th>( \text{Yield} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>74, 68, 75</td>
</tr>
<tr>
<td>2</td>
<td>76, 71, 77</td>
</tr>
<tr>
<td>3</td>
<td>75, 72, 73</td>
</tr>
<tr>
<td>4</td>
<td>72, 74, 81</td>
</tr>
<tr>
<td>5</td>
<td>79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>( F_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>147.73</td>
<td>4</td>
<td>36.93</td>
<td>20.5</td>
</tr>
<tr>
<td>Within</td>
<td>18.00</td>
<td>10</td>
<td>1.80</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>165.73</td>
<td>14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Highly significant result \((F_{0.05,4,10} = 3.48)\)

\[ \hat{\sigma}^2 = (36.93 - 1.80)/3 = 11.71 \]

86.7% (=11.71/(11.71+1.80)) is attributable to batch differences

Time to improve consistency of the batches

\[ \bar{\gamma} = \frac{1}{a}(\bar{\gamma}_1 + \bar{\gamma}_2 + \cdots + \bar{\gamma}_a) \]

\( \bar{\gamma} \) iid Normal but what is variance?

CI for \( \mu \): \( \bar{\gamma} \pm t\sqrt{\text{Var}(\bar{\gamma})} \)
Example
Confidence Intervals

- 95% CI for \( \sigma^2 \)

\[
\frac{SS_E}{\chi^2_{0.025,10}} \leq \sigma^2 \leq \frac{SS_E}{\chi^2_{0.975,10}} = (18.00/20.48, 18.00/3.25) = (0.879, 5.538)
\]

- 95% CI for Intraclass Correlation

\[
\left( \frac{20.52-4.47}{20.52+3(1-1)/1.847}, \frac{20.52-1/8.84}{20.52+3(1-1)/1.847} \right) \approx (0.545, 0.984)
\]

using property that

\[
F_{1,n-2,\nu,\nu} = 1/F_{\nu,2,\nu,\nu}
\]

Using SAS

```
options nocenter ps=35 ls=72;
data example;
input batch percent;
cards;
17 4
17 6
17 5
26 8
6 9;
proc glm;
class batch;
model percent=batch;
random batch;
output out=diag r=res p=pred;
proc plot;
plot res*pred;
proc varcomp method = type1;
class batch;
model percent = batch;
proc mixed cl;
class batch;
model percent = batch;
random batch;
run;
```

```
Dependent Variable: PERCENT

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>147.73333</td>
<td>36.93333</td>
<td>20.52</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>10</td>
<td>18.00000</td>
<td>1.80000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>14</td>
<td>165.73333</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source | Type III Expected Mean Square | Var(Error) + 3 Var(BATCH)
BATCH   | Source | Type I SS | Mean Square | F Value | Pr > F |
BATCH   | 4      | 147.73333 | 36.93333    | 20.52   | 0.0001 |

Variance Components Estimation Procedure

```
Dependent Variable: PERCENT

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Type I MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATCH</td>
<td>4</td>
<td>147.73333</td>
<td>36.93333</td>
</tr>
<tr>
<td>Error</td>
<td>10</td>
<td>18.000000</td>
<td>1.800000</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>14</td>
<td>165.73333</td>
<td></td>
</tr>
</tbody>
</table>

Source | Expected Mean Square | Var(Error) + 3 Var(BATCH) | Var(Error) | Var(BATCH)
BATCH   | 4      | 11.7111111  | 1.80000000 |
BATCH   | 6      | 11.7111111  | 1.80000000 |

Covariance Parameter Estimates (REML)

```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
Negative $\sigma^2_T$ Estimate Example

options nocenter ps=39 ls=64;
data new;
input class subj score @@;
cards;
1 1 74.62 1 2 73.90 1 3 72.27 1 4 71.60 1 5 73.80
1 6 77.42 1 7 72.16 1 8 76.69 1 9 75.84 1 10 70.35
2 1 72.55 2 2 71.44 2 3 72.67 2 4 72.59 2 5 71.25
2 6 68.99 2 7 69.61 2 8 77.44 2 9 73.99 2 10 73.90
3 1 76.66 3 2 74.76 3 3 70.47 3 4 75.38 3 5 68.32
3 6 76.69 3 7 73.34 3 8 68.24 3 9 69.33 3 10 78.22;
proc glm;
class class;
model score = class;
random class / test;
proc varcomp method = type1;
class class;
model score = class;
proc varcomp method = reml;
class class;
model score = class;
proc mixed cl;
class class;
model score = ;
random class;
run;

General Linear Models Procedure
Dependent Variable: SCORE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>10.11154667</td>
<td>0.60</td>
<td>0.5557</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>227.34895000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>29</td>
<td>237.46049667</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square 0.042582 C.V. SCORE Mean 73.1496667

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>2</td>
<td>10.11154667</td>
<td>0.60</td>
<td>0.5557</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>2</td>
<td>10.11154667</td>
<td>0.60</td>
<td>0.5557</td>
</tr>
</tbody>
</table>

General Linear Models Procedure

Tests of Hypotheses for Random Model Analysis of Variance
Source: CLASS
Error: MS(Error)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III MS</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>2</td>
<td>8.18829299</td>
<td>0.60</td>
<td>0.5557</td>
</tr>
</tbody>
</table>

The MIXED Procedure

Class Level Information

<table>
<thead>
<tr>
<th>Class Levels Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS 1 2 3</td>
</tr>
</tbody>
</table>

REML Estimation Iteration History

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Evaluations</th>
<th>Objective</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>93.37965543</td>
<td>0.00000000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>93.37965543</td>
<td>0.00000000</td>
</tr>
</tbody>
</table>

Convergence criteria met.

Covariance Parameter Estimates (REML)

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Estimate</th>
<th>Alpha</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>0.00000000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>8.18829299</td>
<td>0.05</td>
<td>5.1935</td>
<td>14.7977</td>
</tr>
</tbody>
</table>

Model Fitting Information for SCORE

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
<th>Observations</th>
<th>30.0000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Res Log Likelihood</td>
<td>-73.3390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akaike's Information Criterion</td>
<td>-75.3390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schwarz's Bayesian Criterion</td>
<td>-76.7063</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2 Res Log Likelihood</td>
<td>146.6781</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Variance Components Estimation Procedure
Class Levels Values
CLASS 3 1 2 3

Number of observations in data set = 30

Variance Components Estimation Procedure
Dependent Variable: SCORE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Type I MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>2</td>
<td>10.11154667</td>
<td>5.05577333</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>227.34895000</td>
<td>8.42033148</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>29</td>
<td>237.46049667</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS</td>
<td>Var(Error) + 10 Var(CLASS)</td>
</tr>
<tr>
<td>Error</td>
<td>Var(Error)</td>
</tr>
</tbody>
</table>

Variance Components Estimate

<table>
<thead>
<tr>
<th>Variance Component</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var(CLASS)</td>
<td>-0.33645581</td>
</tr>
<tr>
<td>Var(Error)</td>
<td>8.42033148</td>
</tr>
</tbody>
</table>

REML Procedure
Dependent Variable: SCORE

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Objective</th>
<th>Var(CLASS)</th>
<th>Var(Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>60.97845805</td>
<td>0</td>
<td>8.18829299</td>
</tr>
<tr>
<td>1</td>
<td>60.97845805</td>
<td>0</td>
<td>8.18829299</td>
</tr>
</tbody>
</table>

Convergence criteria met.

<table>
<thead>
<tr>
<th>Asymptotic Covariance Matrix of Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var(CLASS)</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Var(CLASS)</td>
</tr>
<tr>
<td>Var(Error)</td>
</tr>
</tbody>
</table>