Lecture 4. Random Effects in Completely Randomized Design

Montgomery: 3.9, 13.1 and 13.7
Random Effects vs Fixed Effects

- Consider factor with numerous possible levels
- Want to draw inference on population of levels
- Not just concerned with levels in experiment

- Example of differences
  - **Fixed**: Compare reading ability of 10 2nd grade classes in NY
    * Select $a = 10$ specific classes of interest.
    * Randomly choose $n$ students from each classroom.
    * Want to compare $\tau_i$ (class-specific effects).
  - **Random**: Compare variability among all 2nd grade classes in NY
    * Randomly choose $a = 10$ classes from large number of classes.
    * Randomly choose $n$ students from each classroom.
    * Want to assess $\sigma_\tau^2$ (class to class variability).

- Inference broader in random effects case: inference on population with randomly chosen levels
Random Effects Model (CRD)

- Same model as in fixed effects case

\[ y_{ij} = \mu + \tau_i + \varepsilon_{ij} \]

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

But view treatment effects in different way

- Instead of \( \sum \tau_i = 0 \), assume
  \( \tau_i \overset{iid}{\sim} N(0, \sigma^2_{\tau}) \)
  \( \{\tau_i\} \) and \( \{\varepsilon_{ij}\} \) independent

- \( \text{Var}(y_{ij}) = \sigma^2_{\tau} + \sigma^2 \)
Random Effects Model

- The hypotheses are:

\[ H_0 : \sigma^2 \tau = 0 \]
\[ H_1 : \sigma^2 \tau > 0 \]

- Partitioning of Total Sum of Squares identical

\[ E(MS_E) = \sigma^2 \]
\[ E(MS_{Treatment}) = \sigma^2 + n\sigma^2 \tau \]

- Under \( H_0 \), \( F_0 = MS_{Treatment} / MS_E \sim F_{a-1, N-a} \)

- 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)

\[ \hat{\sigma}^2 = MS_E \]

\[ \hat{\sigma}_\tau^2 = (MS_{\text{Treatment}} - MS_E)/n \]

If unbalanced, replace \( n \) with

\[ n_0 = ((\sum n_i)^2 - \sum n_i^2)/((a - 1)\sum n_i) \]

- Estimate of \( \sigma^2_\tau \) can be negative
  - Supports \( H_0 \)? Use zero as estimate?
  - Is the model reasonable?
  - Bayesian approach (nonnegative prior)
  - Use estimation method that only gives nonnegative estimates
Confidence intervals

- $\sigma^2$: Same as fixed case

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

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

- $\sigma^2_\tau$: Approximate Confidence Interval

\[
\hat{\sigma}^2_\tau = \frac{(\text{MS}_{\text{Trt}} - \text{MS}_E)}{n}
\]

- No exact calculation of CI available.
- Approximate CI based on Satterthwaite’s Approximation:

\[
r \frac{\hat{\sigma}^2_\tau}{\chi^2_{\alpha/2,r}} \leq \sigma^2_\tau \leq r \frac{\hat{\sigma}^2_\tau}{\chi^2_{1-\alpha/2,r}}
\]

\[
r = \frac{(\text{MS}_{\text{Trt}} - \text{MS}_E)^2}{\text{MS}_{\text{Trt}}^2/(a-1) + \text{MS}_E^2/(N-a)}
\]
Approximate Confidence interval

- CI for a variance component: \( \sigma_0^2 = E[MS' - MS'']/k \)
  - \( MS' = MS_r + \cdots + MS_s \)
  - \( MS'' = MS_u + \cdots + MS_v \)
  - No common mean squares terms shared by \( MS' \) and \( MS'' \).
  - Note \( f_i MS_i/\sigma_i^2 = SS_i/\sigma_i^2 \overset{ind}{\sim} \chi^2_{f_i} \)
  - Point estimate of \( \sigma_0^2 \): \( \hat{\sigma}_0^2 = (MS' - MS'')/k \)

- Satterthwaite’s Approximation: \( r\hat{\sigma}_0^2/\sigma_0^2 \sim \chi^2_r \)

\[
r = (\hat{\sigma}_0^2)^2 \left/ \sum_i \frac{MS_i^2}{k^2 f_i} \right.
\]

- Approximate \( (1 - \alpha) \times 100\% \) CI of \( \sigma_0^2 \)

\[
r\hat{\sigma}_0^2/\chi^2_{\alpha/2,r} \leq \sigma_0^2 \leq r\hat{\sigma}_0^2/\chi^2_{1-\alpha/2,r}
\]
• Proportion of $\sigma^2_T$ in Var($y_{ij}$), i.e., Intraclass Correlation Coefficient (ICC)

Common estimate if goal is to reduce variance

Uses ratio of two $\chi^2$ distributions (i.e., $F$ dist)

$$\frac{L}{L + 1} \leq \frac{\sigma^2_T}{\sigma^2 + \sigma^2_T} \leq \frac{U}{U + 1}$$

$$L = \frac{1}{n} \left( \frac{MS_{Trt}}{MS_E} \frac{F_{\alpha/2,a-1,N-a}}{F_{\alpha/2,a-1,N-a}} - 1 \right) = \frac{1}{n} \left( \frac{F_0}{F_{\alpha/2,a-1,N-a}} - 1 \right)$$

$$U = \frac{1}{n} \left( \frac{MS_{Trt}}{MS_E} \frac{F_{1-\alpha/2,a-1,N-a}}{F_{1-\alpha/2,a-1,N-a}} - 1 \right) = \frac{1}{n} \left( \frac{F_0}{F_{1-\alpha/2,a-1,N-a}} - 1 \right)$$

• Grand mean: $\mu$

Example: Average reading ability of 2nd grade class,

$$\bar{y}_{..} = (\bar{y}_{1.} + \bar{y}_{2.} + \ldots + \bar{y}_{a.})/a.$$ 

$\bar{y}_{i.}$ iid Normal. But what is the variance?

CI for $\mu : \bar{y}_{..} \pm t_{\alpha/2,a-1} \sqrt{MS_{Trt}/(an)}$
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>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>74</td>
<td>68</td>
<td>75</td>
<td>72</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>76</td>
<td>71</td>
<td>77</td>
<td>74</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>72</td>
<td>77</td>
<td>73</td>
<td>79</td>
</tr>
<tr>
<td>Source of Variation</td>
<td>Sum of Squares</td>
<td>Degrees of Freedom</td>
<td>Mean Square</td>
<td>$F_0$</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Between</td>
<td>147.73</td>
<td>4</td>
<td>36.93</td>
<td>20.5</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>18.00</td>
<td>10</td>
<td>1.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>165.73</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Highly significant result ($F_{0.05,4,10} = 3.48$)
- $\hat{\sigma}^2_t = (36.93 - 1.80)/3 = 11.71$
- 86.7% (=11.71/(11.71+1.80)) is attributable to batch differences
- Time to improve consistency of the batches
• 95% CI for $\sigma^2$

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

• 95% CI for $\sigma^2_\tau$

\[
r = \frac{(36.93 - 1.80)^2}{36.93^2/4 + 1.80^2/10} = 3.62
\]

\[
\chi^2_{0.025,3.62} = (1 - .62)\chi^2_{.025,3} + .62\chi^2_{.025,4}
\]

\[
= .38 \times 9.35 + .62 \times 11.14 = 10.4598
\]

\[
\chi^2_{0.975,3.62} = (1 - .62)\chi^2_{.975,3} + .62\chi^2_{.075,4}
\]

\[
= .38 \times .22 + .62 \times .48 = .3812
\]

\[
(3.62 \times 11.71/10.4598, 3.62 \times 11.71/.3812)
\]

\[
= (4.0527, 111.2020)
\]

– SAS allows noninteger degrees of freedom for $\chi^2$ with: \texttt{CINV(p, df)}
• 95% CI for Intraclass Correlation

\[
L = \frac{1}{3} \times \left( \frac{20.52}{4.47} - 1 \right) = 1.1969 \implies \frac{L}{L + 1} = 0.5448
\]

\[
U = \frac{1}{3} \times \left( \frac{20.52}{1/8.84} - 1 \right) = 60.1323 \implies \frac{U}{U + 1} = 0.9836
\]

95% CI: (0.545, 0.984)

using property that

\[
F_{1-\alpha/2,v_1,v_2} = 1/F_{\alpha/2,v_2,v_1}
\]
Using SAS

data example; input batch percent @@; cards;
   1 74  1 76  1 75  2 68  2 71
   2 72  3 75  3 77  3 77  4 72
   4 74  4 73  5 79  5 81  5 79
;

   proc glm data=example;
      class batch;
      model percent=batch;
      random batch;
      output out=diag r=res p=pred;
   
   proc gplot data=diag;
      plot res*pred;
   
   proc mixed data=example cl;
      class batch;
      model percent = ;
      random batch / vcorr; run;
Dependent Variable: PERCENT

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</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

<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>BATCH</td>
<td>4</td>
<td>147.73333</td>
<td>36.93333</td>
<td>20.52</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Source

Type III Expected Mean Square

BATCH Var(Error) + 3 Var(BATCH)

- GLM: Uses least squares for estimation
  - Not designed for random effects...Must compute variance estimates / CIs by hand

- MIXED: Uses restricted maximum likelihood (REML)
  - Often preferred to ML because it produces unbiased estimates of covariance parameters by taking into account the loss of degrees of freedom in estimating fixed effects
  - Usually residual variance profiled out of the likelihood
The MIXED Procedure

Estimated V Correlation Matrix for batch 1

<table>
<thead>
<tr>
<th>Row</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0000</td>
<td>0.8668</td>
<td>0.8668</td>
</tr>
<tr>
<td>2</td>
<td>0.8668</td>
<td>1.0000</td>
<td>0.8668</td>
</tr>
<tr>
<td>3</td>
<td>0.8668</td>
<td>0.8668</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

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>BATCH</td>
<td>11.71111111</td>
<td>0.05</td>
<td>4.0450</td>
<td>114.2090</td>
</tr>
<tr>
<td>Residual</td>
<td>1.80000000</td>
<td>0.05</td>
<td>0.8788</td>
<td>5.5436</td>
</tr>
</tbody>
</table>

Fit Statistics

-2 Res Log Likelihood 62.8
AIC (smaller is better) 66.8
AICC (smaller is better) 67.8
BIC (smaller is better) 66.0
Negative $\sigma^2_T$ Estimate Example

data new;
  input school 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 data=new;
  class school;
  model score = school;
  random school;

proc mixed data=new cl;
  class school;
  model score = ;
  random school; run;
### The GLM Procedure

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2</td>
<td>10.1115467</td>
<td>5.0557733</td>
<td>0.60</td>
<td>0.5557</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>227.3489500</td>
<td>8.4203315</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>29</td>
<td>237.4604967</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-Square</th>
<th>Coeff Var</th>
<th>Root MSE</th>
<th>score Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.042582</td>
<td>3.966909</td>
<td>2.901781</td>
<td>73.14967</td>
</tr>
</tbody>
</table>

<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>school</td>
<td>2</td>
<td>10.1115467</td>
<td>5.0557733</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>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>school</td>
<td>2</td>
<td>10.1115467</td>
<td>5.0557733</td>
<td>0.60</td>
<td>0.5557</td>
</tr>
</tbody>
</table>

Parameter Estimates using ANOVA method

\[ \hat{\sigma}^2 = 8.42 \]

\[ \hat{\sigma}_\tau^2 = \frac{5.0558 - 8.4203}{10} = -3.3645/10 = -0.37 \]
The MIXED Procedure

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>school</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Residual</td>
<td>8.1883</td>
<td>0.05</td>
<td>5.1935</td>
<td>14.7977</td>
</tr>
</tbody>
</table>

Fit Statistics

-2 Res Log Likelihood 146.7
AIC (smaller is better) 148.7
AICC (smaller is better) 148.8
BIC (smaller is better) 147.8
Word of Caution

- In log file for PROC MIXED analysis,

  NOTE: Convergence criteria met.
  NOTE: Estimated G matrix is not positive definite.

- Caused by variance estimated to be zero

- Suggests possibly removing this term from the model

- Decision often an issue in more complicated models

- Can remove positivity constraint (NOBOUND)

  proc mixed cl nobound;
  class school;
  model score = ;
  random school ;
  run;

- Includes null model LRT to determine if it is necessary to model the covariance structure