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 τ_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 σ_τ^2 (class to class variability).

- Inference broader in random effects case
- Levels chosen randomly → inference on population

Random Effects Model (CRD)

- Same model as in the fixed case

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

μ - grand mean

τ_i - ith treatment effect

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

But view number of treatment levels as infinite

- Instead of $\sum \tau_i = 0$, assume

  $\tau_i \sim \mathrm{N}(0, \sigma_\tau^2)$

  $\{\tau_i\}$ and $\{\epsilon_{ij}\}$ independent

- $\mathrm{Var}(y_{ij}) = \sigma_\tau^2 + \sigma^2$ (HW #1 Problem 4)

Random Effects Model

- The hypotheses are:

$$\begin{aligned} H_0 &: \sigma_\tau^2 = 0 \\ H_1 &: \sigma_\tau^2 > 0 \end{aligned}$$

- Partitioning of Total Sum of Squares identical

  $\mathrm{E}(\mathrm{MS_E}) = \sigma^2$

  $\mathrm{E}(\mathrm{MS_{Treatment}}) = \sigma^2 + n\sigma_\tau^2$

- Under H_0, $F_0 \sim F_{\alpha, a-1, N-a}$

- 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)

  $\hat{\sigma}^2 = \text{MS}_\text{E}$

  $\hat{\sigma}^2_\tau = (\text{MS}_\text{Treatment} - \text{MS}_\text{E})/n$

  If unbalanced, replace n with

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

- Estimate of σ^2_τ can be negative
  - Supports H_0? Use zero as estimate?
  - Validity of model? Nonlinear?
  - Bayesian approach (nonnegative prior)

Confidence intervals

- σ^2: Same as fixed case

$$\frac{(N-a)\text{MS}_\text{E}}{\sigma^2} \sim \chi^2_{N-a}$$

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

- σ^2_τ: Linear combination of χ^2

$$\frac{(a-1)\text{MS}_\text{Trt}}{\sigma^2 + n\sigma^2_\tau} \sim \chi^2_{a-1}$$

so

$$f(\sigma^2_\tau) = \frac{\sigma^2 + n\sigma^2_\tau}{n(a-1)}\chi^2_{a-1} - \frac{\sigma^2}{n(N-a)}\chi^2_{N-a}$$

No closed form expression for this distribution

Approximations available (Section 12-7)

- Proportion of σ^2_τ in Var(y_{ij})

  Common estimate if goal is to reduce variance

  Uses ratio of two χ^2 distributions (i.e., F dist)

  Instead of

$$\frac{\text{L}}{\text{L}+1} \leq \frac{\sigma^2_\tau}{\sigma^2 + \sigma^2_\tau} \leq \frac{\text{U}}{\text{U}+1}$$

$$\text{L} = \frac{1}{n}\left(\frac{\text{MS}_\text{Trt}}{\text{MS}_\text{E}\ F_{\alpha/2,a-1,N-a}} - 1\right)$$

$$\text{U} = \frac{1}{n}\left(\frac{\text{MS}_\text{Trt}}{\text{MS}_\text{E}\ F_{1-\alpha/2,a-1,N-a}} - 1\right)$$

  Consider

$$\frac{F_0 - F_{\alpha/2,a-1,N-a}}{F_0 + (n-1)F_{\alpha/2,a-1,N-a}} \leq \frac{\sigma^2_\tau}{\sigma^2 + \sigma^2_\tau} \leq \frac{F_0 - F_{1-\alpha/2,a-1,N-a}}{F_0 + (n-1)F_{1-\alpha/2,a-1,N-a}}$$

- Grand mean -μ

  Example: Average reading ability of 2nd grade class. Unit is the class, selection of students is subsampling.

$$\overline{y}_{..} = \frac{1}{a}(\overline{y}_{1.} + \overline{y}_{2.} + ... + \overline{y}_{a.})$$

  $\overline{y}_{i.}$ iid Normal but what is variance?

  CI for μ : $\overline{y}_{..} \pm t\sqrt{\text{Var}(\overline{y}_{..})}$

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.

Batch 1	2	3	4	5
74	68	75	72	79
76	71	77	74	81
75	72	77	73	79

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between	147.73	4	36.93	20.5
Within	18.00	10	1.80	
Total	165.73	14		

Highly significant result ($F_{.05,4,10} = 3.48$)

$\hat{\sigma}^2_\tau = (36.93 - 1.80)/3 = 11.71$

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

Time to improve consistency of the batches

Example

Confidence Intervals

- 95% CI for σ^2

$$\frac{\mathsf{SS_E}}{\chi^2_{.025,10}} \le \sigma^2 \le \frac{\mathsf{SS_E}}{\chi^2_{.975,10}} = (18.00/20.48, 18.00/3.25)$$
$$= (0.879, 5.538)$$

- 95% CI for Intraclass Correlation

$$\left(\frac{20.52-4.47}{20.52+(3-1)4.47}, \frac{20.52-(1/8.84)}{20.52+(3-1)(1/8.84)}\right)$$

$$(0.545, 0.984)$$

using property that

$$F_{1-\alpha/2,v_1,v_2} = 1/F_{\alpha/2,v_2,v_1}$$

Using SAS

```
options nocenter ps=35 ls=72;

data example;
  input batch percent;
  cards;
   1 74
   1 76
   1 75
   2 68
   .
   5 79
 ;

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 = ;
 random batch;
run;
```



```
Dependent Variable: PERCENT
                               Sum of          Mean
Source                 DF     Squares        Square  F Value    Pr > F
Model                   4   147.73333      36.93333    20.52    0.0001
Error                  10    18.00000       1.80000
Corrected Total        14   165.73333

Source                 DF   Type I SS   Mean Square  F Value    Pr > F
BATCH                   4   147.73333      36.93333    20.52    0.0001

Source       Type III Expected Mean Square
BATCH        Var(Error) + 3 Var(BATCH)

------------------------------------------------------------------------

Variance Components Estimation Procedure

Dependent Variable: PERCENT
Source                   DF         Type I SS         Type I MS
BATCH                     4      147.73333333       36.93333333
Error                    10       18.00000000        1.80000000
Corrected Total          14      165.73333333

Source                   Expected Mean Square
BATCH                    Var(Error) + 3 Var(BATCH)
Error                    Var(Error)

Variance Component                   Estimate
Var(BATCH)                        11.71111111
Var(Error)                         1.80000000
```



```
                     The MIXED Procedure

                 Class     Levels  Values
                 BATCH          5  1 2 3 4 5

              REML Estimation Iteration History

           Iteration  Evaluations     Objective     Criterion
                   0            1   51.30656858
                   1            1   37.02237479    0.00000000
Convergence criteria met.

             Covariance Parameter Estimates (REML)

           Cov Parm       Estimate  Alpha     Lower      Upper
           BATCH       11.71111111   0.05    4.0450   114.2090
           Residual     1.80000000   0.05    0.8788     5.5436

             Model Fitting Information for PERCENT

                Description                        Value
                Observations                     15.0000
                Res Log Likelihood              -31.3763
                Akaike's Information Criterion  -33.3763
                Schwarz's Bayesian Criterion    -34.0154
                -2 Res Log Likelihood            62.7527
```

Negative σ_τ^2 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
Source                  DF    Sum of Squares  F Value   Pr > F
Model                    2       10.11154667     0.60   0.5557
Error                   27      227.34895000
Corrected Total         29      237.46049667

                  R-Square              C.V.        SCORE Mean
                  0.042582          3.966909        73.1496667


Source                  DF         Type I SS  F Value   Pr > F
CLASS                    2       10.11154667     0.60   0.5557

Source                  DF       Type III SS  F Value   Pr > F
CLASS                    2       10.11154667     0.60   0.5557

General Linear Models Procedure

Source      Type III Expected Mean Square
CLASS       Var(Error) + 10 Var(CLASS)

Tests of Hypotheses for Random Model Analysis of Variance

Source: CLASS
Error: MS(Error)
                         Denominator    Denominator
 DF    Type III MS            DF             MS     F Value  Pr > F
 2   5.0557733333           27   8.4203314815      0.6004   0.5557
```



```
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
Source                    DF         Type I SS         Type I MS
CLASS                      2       10.11154667        5.05577333
Error                     27      227.34895000        8.42033148
Corrected Total           29      237.46049667

Source                    Expected Mean Square
CLASS                     Var(Error) + 10 Var(CLASS)
Error                     Var(Error)

Variance Component                    Estimate
Var(CLASS)                         -0.33645581
Var(Error)                          8.42033148

---------------------------------------------------------------
                        REML Procedure
Dependent Variable: SCORE

Iteration        Objective       Var(CLASS)        Var(Error)
    0          60.97845805                0        8.18829299
    1          60.97845805                0        8.18829299

Convergence criteria met.

Asymptotic Covariance Matrix of Estimates

                  Var(CLASS)        Var(Error)
Var(CLASS)                 0                 0
Var(Error)                 0      4.6240097976
```



```
                    The MIXED Procedure

                  Class Level Information
                  Class     Levels  Values
                  CLASS          3  1 2 3

           REML Estimation Iteration History

       Iteration  Evaluations     Objective     Criterion
               0            1   93.37965543
               1            1   93.37965543    0.00000000
Convergence criteria met.


         Covariance Parameter Estimates (REML)

       Cov Parm       Estimate  Alpha     Lower     Upper
       CLASS        0.00000000      .         .         .
       Residual     8.18829299   0.05    5.1935   14.7977


          Model Fitting Information for SCORE

            Description                        Value
            Observations                     30.0000
            Res Log Likelihood              -73.3390
            Akaike's Information Criterion  -75.3390
            Schwarz's Bayesian Criterion    -76.7063
            -2 Res Log Likelihood           146.6781
```