Topic 25 - One-Way Random Effects Models

STAT 525 - Fall 2013

Outline

- One-way Random effects
  - Model
  - Variance component estimation
  - Confidence intervals

Random Effects vs Fixed Effects

- Consider factor with numerous levels
- Want to draw inference **on population of levels**, not specifically concerned with comparing observed levels
- Clear example of difference (1=fixed, 2=random)
  1. Compare reading ability of 10 2nd grade classes in Indiana

     Go to each of the $r = 10$ specific classes of interest and randomly choose n students from each classroom.
  2. Compare variability **among all** 2nd grade classes in Indiana

     Randomly choose $r = 10$ classes from population of 2nd grade classes. Then randomly choose n students from each classroom.
- Inference broader in random effects case
- Levels chosen randomly $\rightarrow$ inference on pop of levels

Data for One-way Random Effects Model

- Exact same data framework as fixed effects case.....
- Y is the response variable
- Factor with levels $i = 1, 2, ..., r$
- Y_{ij} is the j^{th} observation from cell i
- Consider $j = 1, 2, ..., n_i$
- but different statistical model.

Example Page 1036

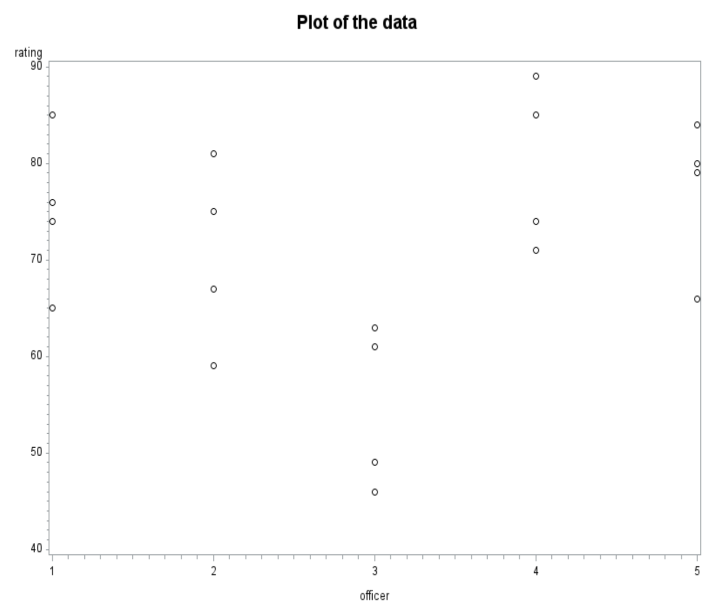
- Interested in studying the variability in the rating of job applicants
- Two sources of variability
  - Variability among applicants
  - Variability among personnel officers
- Y is the job applicant rating
- Factor: officer/interviewer ($r = 5$)
- Interviewers selected at random from population of personnel officers (assume population large)
- Twenty applicants randomly and equally assigned ($n = 4$) to each personnel officer

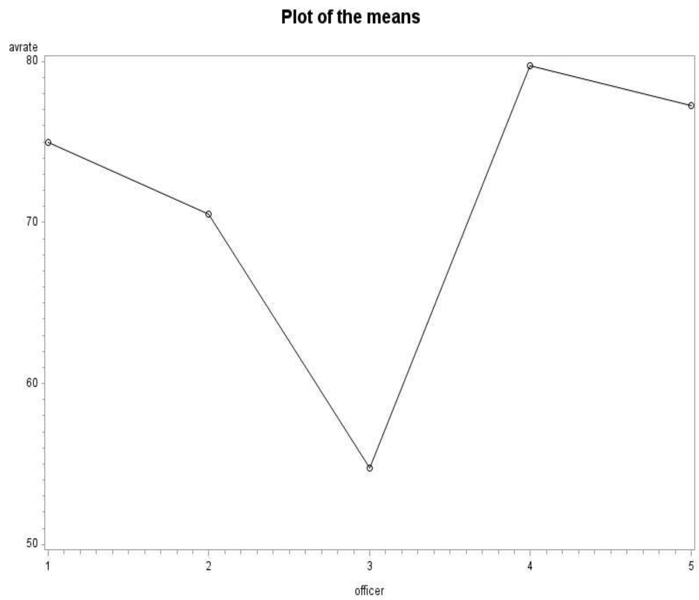
SAS Commands

```
options nocenter;
data a1; infile 'u:\.www\datasets525\CH25TA01.txt';
   input rating officer;
proc print data=a1; run;

title1 'Plot of the data';
symbol1 v=circle i=none c=black;      ***Scatterplot;
proc gplot data=a1;
   plot rating*officer/frame;

proc means data=a1; output out=a2 mean=avrate;
   var rating; by officer;

title1 'Plot of the means';
symbol1 v=circle i=join c=black;
proc gplot data=a2;                  ***Means plot;
   plot avrate*officer/frame;
run;
```


Output

Obs	rating	officer	Obs	rating	officer
1	76	1	13	74	4
2	65	1	14	71	4
3	85	1	15	85	4
4	74	1	16	89	4
5	59	2	17	66	5
6	75	2	18	84	5
7	81	2	19	80	5
8	67	2	20	79	5
9	49	3			
10	63	3			
11	61	3			
12	46	3			





Random Effects Model

- Expressed numerically

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

$$\mu_i \sim N(\mu, \sigma_\mu^2)$$

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

$$\mu_i \text{ and } \varepsilon_{ij} \text{ independent}$$

- Implies $Y_{ij} \sim N(\mu, \sigma_\mu^2 + \sigma^2)$
- $\text{Cov}(Y_{ij}, Y_{ik})$=$\sigma_\mu^2$ — Correlation between some obs
- Also called *Model II*

Random Factor Effects Model

- Statistical model is

$$Y_{ij} \quad = \quad \mu + \tau_i + \varepsilon_{ij}$$

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

$$\tau_i \sim N(0, \sigma_\mu^2)$$

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

- There are TWO parameters/variances in each model
- Cell means are random variables, not parameters

Quantities of Interest

- Often interested in the percent of total variability due to factor

$$\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2} = \frac{\sigma_\mu^2}{\sigma_Y^2}$$

- Is also called the intraclass correlation coefficient because it describes the correlation between two observations from the same factor level

$$\rho_{IC} = \frac{\text{Cov}(Y_{ij}, Y_{ik})}{\sqrt{\text{Var}(Y_{ij})\text{Var}(Y_{ik})}} = \frac{\sigma_\mu^2}{\sigma_Y^2}$$

- Depending of example, may want this value small or large

Least Squares Approach: ANOVA Table and EMS

- Terms and layout of ANOVA table the <u>same</u> as that in the fixed effects case
- The expected means squares (EMS) are <u>different</u> because of the different model assumptions
- This also means the hypotheses being tested are different

Likelihood Approach: General Mixed Effect Model

- Consider expressing the model

$$Y = X\beta + Z\delta + \varepsilon$$

  β is a vector of fixed-effect parameters

  δ is a vector of random-effect parameters

  ε is the error vector

- δ and ε assumed uncorrelated
  - means 0
  - covariance matrices G and R

Likelihood Approach: General Mixed Effect Model

- $\text{Cov}(Y) = ZGZ' + R$
- If $R = \sigma^2 I$ and $Z = 0$, back to standard linear model
- SAS Proc Mixed allows one to specify G and R
- G through RANDOM, R through REPEATED

Likelihood Approach: General Mixed Effect Model

- For known G and R,

$$\hat{\beta} = \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}Y$$
$$\hat{\delta} = GZ'\Sigma^{-1}(Y - X\hat{\beta})$$

- For unknown G and R, their REML estimates can be substituted into these expressions
- REML uses likelihood to take into account loss of DF

$$-2\log L = (n-p)\log(2\pi) + \log(|\Sigma|) + r'\Sigma^{-1}r + \log(|X'\Sigma^{-1}X|)$$
$$\text{where} \quad r = Y - X\hat{\beta}$$

Random Effects Model

- The hypotheses are:

$$H_0 : \sigma_\mu^2 = 0$$
$$H_a : \sigma_\mu^2 > 0$$

- Same breakdown of Total SS but
  E(MSE)=σ^2
  E(MSTR)=$\sigma^2 + n\sigma_\mu^2$
- Under H_0, $F_0 \sim F_{\alpha,r-1,n_T-r}$
- <u>Same F test as before</u>
- Conclusion pertains to entire population of levels

Model Estimates

- Typically interested in estimating variances and functions of these variances
- Under ANOVA/least squares approach, use mean squares
  $\hat{\sigma}^2 = \text{MSE}$
  $\hat{\sigma}_\mu^2 = (\text{MSTR} - \text{MSE})/\text{n}$
  If unbalanced, replace n with
  $n_0 = ((\sum n_i)^2 - \sum n_i^2)/((r-1)\sum n_i)$
- Under this approach, estimate of σ_μ^2 can be negative
  - Supports H_0 so use zero as estimate?
  - If σ_μ^2 small, chance variation can result in negative estimate
  - Bayesian approach (nonnegative prior)
  - Residual maximum likelihood (nonnegative restriction)

SAS Commands

```
proc glm data=a1;
   class officer;
   model rating=officer;
   random officer;
run;

proc mixed data=a1 cl;
   class officer;
   model rating=;
   random officer/vcorr;
run;
```


Output

```
                           Sum of
Source          DF        Squares  Mean Square  F Value  Pr > F
Model            4    1579.700000   394.925000     5.39  0.0068
Error           15    1099.250000    73.283333
Corrected Total 19    2678.950000

R-Square     Coeff Var      Root MSE    rating Mean
0.589671      11.98120      8.560569       71.45000

Source          DF      Type I SS  Mean Square  F Value  Pr > F
officer          4    1579.700000   394.925000     5.39  0.0068

Source          DF    Type III SS  Mean Square  F Value  Pr > F
officer          4    1579.700000   394.925000     5.39  0.0068

Source                 Type III Expected Mean Square
officer                Var(Error) + 4 Var(officer)
```

Output

```
Estimated V Correlation Matrix for Subject 1
 Row      Col1      Col2      Col3      Col4
   1    1.0000    0.5232    0.5232    0.5232
   2    0.5232    1.0000    0.5232    0.5232
   3    0.5232    0.5232    1.0000    0.5232
   4    0.5232    0.5232    0.5232    1.0000

             Covariance Parameter Estimates
Cov Parm      Estimate     Alpha        Lower       Upper
officer        80.4104      0.05      24.4572     1498.97
Residual       73.2833      0.05      39.9896      175.54

           Fit Statistics
-2 Res Log Likelihood          145.2
AIC (smaller is better)        149.2
AICC (smaller is better)       150.0
BIC (smaller is better)        148.5
```


Confidence intervals

- σ^2: Page 1041

$$\frac{r(n-1)\text{MSE}}{\sigma^2} \sim \chi^2_{r(n-1)}$$

$$\frac{r(n-1)\text{MSE}}{\chi^2_{\alpha/2,r(n-1)}} \leq \sigma^2 \leq \frac{r(n-1)\text{MSE}}{\chi^2_{1-\alpha/2,r(n-1)}}$$

- σ^2_μ: Page 1043

$$\frac{(r-1)\text{MSTR}}{\sigma^2 + n\sigma^2_\mu} \sim \chi^2_{r-1}$$

so

$$f(\sigma^2_\mu) = \frac{\sigma^2 + n\sigma^2_\tau}{n(r-1)}\chi^2_{r-1} - \frac{\sigma^2}{nr(n-1)}\chi^2_{r(n-1)}$$

No closed form expression for this distribution

Satterthwaite Procedure page 1043 (Proc Mixed)

MLS Procedure page 1045

Confidence intervals

- Intraclass Correlation Coefficient : Page 1040

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

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

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

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

- Population mean -μ : Page 1038-1039

$$\overline{Y}_{..} = \frac{1}{r}(\overline{Y}_{1.} + \overline{Y}_{2.} + ... + \overline{Y}_{r.})$$

$$\overline{Y}_{i.} \sim N\left(\mu, \sigma^2_\mu + \frac{\sigma^2}{n}\right)$$

Background Reading

- KNNL Section 25.1
- knnl1036.sas
- KNNL Sections 25.2-25.6