Lecture 10: Experiments with Random Effects

Montgomery, Chapter 12 or 13

Example 1

A textile company weaves a fabric on a large number of looms. It would like the looms to be homogeneous so that it obtains a fabric of uniform strength. A process engineer suspects that, in addition to the usual variation in strength within samples of fabric from the same loom, there may also be significant variations in strength between looms. To investigate this, she selects four looms at random and makes four strength determinations on the fabric manufactured on each loom. The layout and data are given in the following.

	Observations			
looms	1	2	3	4
1	98	97	99	96
2	91	90	93	92
3	96	95	97	95
4	95	96	99	98

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

     Select $a = 10$ specific classes of interest. Randomly choose n students from each classroom. Want to compare τ_i (class-specific effects).
  2. Study the 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 σ_τ^2 (class to class variability).
- Inference broader in random effects case
- Levels chosen randomly $\longrightarrow$ inference on population

Random Effects Model (CRD)

- Similar model (as in the fixed case) with different assumptions

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

μ - grand mean

τ_i - ith treatment effect (random)

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

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

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

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

- Var(y_{ij}) = $\sigma_\tau^2 + \sigma^2$

  σ_τ^2 and σ^2 are called variance components

Statistical Analysis

- The basic hypotheses are:

$$H_0 : \sigma_\tau^2 = 0 \text{ vs. } H_1 : \sigma_\tau^2 > 0$$

- Same ANOVA table (as before)

Source	SS	DF	MS	F_0
Between	SS_{tr}	$a-1$	MS_{tr}	$F_0 = \frac{\mathrm{MS}_{tr}}{\mathrm{MS}_E}$
Within	SS_E	$N-a$	MS_E	
Total	$\mathrm{SS}_T = \mathrm{SS}_{tr} + \mathrm{SS}_E$	$N-1$		

  - E($\mathrm{MS_E}$)=σ^2
  - E($\mathrm{MS_{tr}}$)=$\sigma^2 + n\sigma_\tau^2$

- Under H_0, $F_0 \sim F_{a-1,N-a}$
- Same test as before
- Conclusions, however, pertain to entire population

Estimation

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

  $$\hat{\sigma}^2 = \mathrm{MS_E}$$

  $$\hat{\sigma}_\tau^2 = (\mathrm{MS_{tr}} - \mathrm{MS_E})/n$$

  If unbalanced, replace n with

  $$n_0 = \frac{1}{a-1}\left(\sum_{i=1}^a n_i - \frac{\sum_{i=1}^a n_i^2}{\sum_{i=1}^a n_i}\right)$$

- Estimate of σ_τ^2 can be negative
  - Supports H_0? Use zero as estimate?
  - Validity of model? Nonlinear?
  - Other approaches (MLE, Bayesian with nonnegative prior)

Confidence intervals

- σ^2: Same as fixed case

$$\frac{(N-a)\mathrm{MS_E}}{\sigma^2} \sim \chi^2_{N-a}$$

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

- σ^2_τ: Linear combination of χ^2

$$\frac{(a-1)\mathrm{MS_{tr}}}{\sigma^2 + n\sigma^2_\tau} \sim \chi^2_{a-1}$$

$$\hat{\sigma}^2_\tau = (MS_{tr} - MS_E)/n, \text{ so}$$

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

No closed form expression for this distribution. Can use approximation

Recall

$$\frac{MS_{tr}/(\sigma^2 + n\sigma_\tau^2)}{MS_E/\sigma^2} \sim F_{a-1,N-a}$$

- For σ_τ^2/σ^2:

$$L \leq \frac{\sigma_\tau^2}{\sigma^2} \leq U$$

$\mathrm{L} = \left(\frac{\mathrm{MS_{tr}}}{\mathrm{MS_E} F_{\alpha/2,a-1,N-a}} - 1\right)/n$ and $\mathrm{U} = \left(\frac{\mathrm{MS_{tr}}}{\mathrm{MS_E} F_{1-\alpha/2,a-1,N-a}} - 1\right)/n$

- For $\sigma_\tau^2/(\sigma^2 + \sigma_\tau^2)$:

$$\frac{\mathrm{L}}{\mathrm{L}+1} \leq \frac{\sigma_\tau^2}{\sigma^2 + \sigma_\tau^2} \leq \frac{\mathrm{U}}{\mathrm{U}+1},$$ or

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

Loom Experiment (continued)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between	89.19	3	29.73	15.68
Within	22.75	12	1.90	
Total	111.94	15		

Highly significant result ($F_{.05,3,12} = 3.49$)

$\hat{\sigma}_\tau^2 = (29.73 - 1.80)/4 = 6.98$

78.6% (=6.98/(6.98+1.90)) is attributable to loom differences

Time to improve consistency of the looms

Loom Experiment

Confidence Intervals

- 95% CI for σ^2

$$\frac{\mathrm{SS_E}}{\chi^2_{.025,12}} \leq \sigma^2 \leq \frac{\mathrm{SS_E}}{\chi^2_{.975,12}} \quad = \quad c(22.75/23.34, 22.75/4.40)$$

$$= \quad (0.97, 5.17)$$

- 95% CI for $\sigma^2_\tau/(\sigma^2_\tau + \sigma^2)$

$$\left(\frac{15.68 - 4.47}{15.68 + (4-1)4.47}, \frac{15.68 - (1/14.34)}{15.68 + (4-1)(1/14.34)}\right) = (0.385, 0.982)$$

$F_{0.025,3,12} = 4.47,\ F_{.975,3,12} = 1/14.34$

using property that

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

- the variability between looms is not negligible.

Using SAS

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

data example;
  input batch percent;
  datalines;
   1 98
   1 97
   1 99
   1 96
   .
   4 98
 ;
proc glm;
 class loom;
 model strength=loom;
 random loom;
 output out=diag r=res p=pred;

 proc plot;
```

```
 plot res*pred;

proc varcomp method = type1;
 class loom;
 model strength = loom;

proc mixed cl;
 class loom;
 model strength = ;
 random loom;
run;
```

SAS Output

Dependent Variable: strength

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	89.1875000	29.7291667	15.68	0.0002
Error	12	22.7500000	1.8958333		
Corrected Total	15	111.9375000			

Source	DF	Type I SS	Mean Square	F Value
loom	3	89.18750000	29.72916667	15.68

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

Variance Components Estimation Procedure

Dependent Variable: strength

Source	DF	Sum of Squares	Mean Square
loom	3	89.187500	29.729167

```
Error                     12       22.750000        1.895833
Corrected Total           15      111.937500               .

Source                 Expected Mean Square
loom                   Var(Error) + 4 Var(loom)
Error                  Var(Error)

Variance Component        Estimate
Var(loom)                  6.95833
Var(Error)                 1.89583
```

SAS Output (Continued)

The Mixed Procedure

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	75.48910190	
1	1	63.19303249	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
loom	6.9583	0.05	2.1157	129.97
Residual	1.8958	0.05	0.9749	5.1660

Fit Statistics

-2 Res Log Likelihood	63.2
AIC (smaller is better)	67.2
AICC (smaller is better)	68.2
BIC (smaller is better)	66.0

A Measurement Systems Capability Study

A typical gauge R&R experiment is shown below. An instrument or gauge is used to measure a critical dimension of certain part. Twenty parts have been selected from the production process, and three randomly selected operators measure each part twice with this gauge. The order in which the measurements are made is completely randomized, so this is a two-factor factorial experiment with design factors parts and operators, with two replications. Both parts and operators are random factors.

Parts	Operator 1		Operator 2		Operator 3	
1	21	20	20	20	19	21
2	24	23	24	24	23	24
3	20	21	19	21	20	22
.	.	.	.	.	.	.
19	25	26	25	24	25	25
20	19	19	18	17	19	17

Variance components equation: $\sigma_y^2 = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2$

Total variability=Parts + Operators + Interaction + Experimental Error

=Parts + Reproducibility + Repeatability

Statistical Model with Two Random Factors

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases}$$

$$\tau_i \sim \mathrm{N}(0, \sigma_\tau^2) \qquad \beta_j \sim \mathrm{N}(0, \sigma_\beta^2) \qquad (\tau\beta)_{ij} \sim \mathrm{N}(0, \sigma_{\tau\beta}^2)$$

- Var(y_{ijk}) = $\sigma^2 + \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2$
- Expected MS's similar to one-factor random model

  E(MS$_\mathrm{E}$)=σ^2; E(MS$_\mathrm{A}$) = $\sigma^2 + bn\sigma_\tau^2 + n\sigma_{\tau\beta}^2$

  E(MS$_\mathrm{B}$) = $\sigma^2 + an\sigma_\beta^2 + n\sigma_{\tau\beta}^2$; E(MS$_\mathrm{AB}$) = $\sigma^2 + n\sigma_{\tau\beta}^2$

- EMS determine what MS to use in denominator

  $$H_0 : \sigma_\tau^2 = 0 \rightarrow \mathrm{MS_A}/\mathrm{MS_{AB}}$$

  $$H_0 : \sigma_\beta^2 = 0 \rightarrow \mathrm{MS_B}/\mathrm{MS_{AB}}$$

  $$H_0 : \sigma_{\tau\beta}^2 = 0 \rightarrow \mathrm{MS_{AB}}/\mathrm{MS_E}$$

Estimating Variance Components

- Using ANOVA method

$$\hat{\sigma}^2 = \mathrm{MS_E}$$

$$\hat{\sigma}^2_\tau = (\mathrm{MS_A} - \mathrm{MS_{AB}})/bn$$

$$\hat{\sigma}^2_\beta = (\mathrm{MS_B} - \mathrm{MS_{AB}})/an$$

$$\hat{\sigma}^2_{\tau\beta} = (\mathrm{MS_{AB}} - \mathrm{MS_E})/n$$

- Sometimes results in negative estimates
- Proc Varcomp and Proc Mixed compute estimates
- Can use different estimation procedures

  ANOVA method - Method = type1

  RMLE method - Method = reml (default)

- Proc Mixed

  Variance component estimates

  Hypothesis tests and confidence intervals

Gauge Capability Example in Text 12-2

```
options nocenter ls=75;

data randr;
 input part operator resp @@;
 cards;
1 1 21 1 1 20 1 2 20 1 2 20 1 3 19 1 3 21
2 1 24 2 1 23 2 2 24 2 2 24 2 3 23 2 3 24
3 1 20 3 1 21 3 2 19 3 2 21 3 3 20 3 3 22
4 1 27 4 1 27 4 2 28 4 2 26 4 3 27 4 3 28
..........................................
20 1 19 20 1 19 20 2 18 20 2 17 20 3 19 20 3 17
;

proc glm;
 class operator part;
 model resp=operator|part;
 random operator part operator*part / test;
 test H=operator E=operator*part;
 test H=part E=operator*part;
```

```
proc mixed cl maxiter=20 covtest method=type1;
 class operator part;
 model resp = ;
 random operator part operator*part;


proc mixed cl maxiter=20 covtest;
 class operator part;
 model resp = ;
 random operator part operator*part;
run;
quit;
```

Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	59	1215.091667	20.594774	20.77	<.0001
Error	60	59.500000	0.991667		
CorreTotal	119	1274.591667			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
operator	2	2.616667	1.308333	1.32	0.2750
part	19	1185.425000	62.390789	62.92	<.0001
operator*part	38	27.050000	0.711842	0.72	0.8614

Source	Type III Expected Mean Square
operator	Var(Error) + 2 Var(operator*part) + 40 Var(operator)
part	Var(Error) + 2 Var(operator*part) + 6 Var(part)
operator*part	Var(Error) + 2 Var(operator*part)

Tests of Hypotheses Using the Type III
MS for operator*part as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F

operator	2	2.616667	1.308333	1.84	0.1730
part	19	1185.425000	62.390789	87.65	<.0001

Tests of Hypotheses for Random Model Analysis of Variance

Dependent Variable: resp

Source	DF	Type III SS	Mean Square	F Value	Pr > F
operator	2	2.616667	1.308333	1.84	0.1730
part	19	1185.425000	62.390789	87.65	<.0001
Error	38	27.050000	0.711842		

Error: MS(operator*part)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
operator*part	38	27.050000	0.711842	0.72	0.8614
Error: MS(Error)	60	59.500000	0.991667		

The Mixed Procedure

Type 1 Analysis of Variance

Source	DF	Sum of Squares	Mean Square
operator	2	2.616667	1.308333
part	19	1185.425000	62.390789
operator*part	38	27.050000	0.711842
Residual	60	59.500000	0.991667

Type 1 Analysis of Variance

Source	Expected Mean Square	Error Term	Error DF
operator	Var(Residual) + 2 Var(operator*part) + 40 Var(operator)	MS(operator*part)	38
part	Var(Residual) + 2 Var(operator*part) + 6 Var(part)	MS(operator*part)	38
operator*part	Var(Residual) + 2 Var(operator*part)	MS(Residual)	60
Residual	Var(Residual)	.	

Source	F Value	Pr > F
operator	1.84	0.1730
part	87.65	<.0001
operator*part	0.72	0.8614

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
operator	0.0149	0.0330	0.45	0.6510	0.05	-0.0497	0.0795
part	10.2798	3.3738	3.05	0.0023	0.05	3.6673	16.8924
operator*part	-0.1399	0.1219	-1.15	0.2511	0.05	-0.3789	0.0990
Residual	0.9917	0.1811	5.48	<.0001	0.05	0.7143	1.4698

The Mixed Procedure

Estimation Method REML

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	624.67452320	
1	3	409.39453674	0.00003340
2	1	409.39128078	0.00000004
3	1	409.39127700	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
operator	0.0106	0.03286	0.32	0.3732	0.05	0.001103	3.7E12
part	10.2513	3.3738	3.04	0.0012	0.05	5.8888	22.1549
operator*part	0	.	.	.	.	.	.
Residual	0.8832	0.1262	7.00	<.0001	0.05	0.6800	1.1938

Confidence Intervals for Variance Components

- Can use asymptotic variance estimates to form CI
- Known as Wald's approximate CI
- Mixed: option CL=WALD or METHOD=TYPE1

  Use standard normal $\longrightarrow$ 95% CI uses 1.96

  $\hat{\sigma}^2_\beta \pm 1.96(.0330) = (-0.05, 0.08)$

  $\hat{\sigma}^2_\tau \pm 1.96(3.3738) = (3.67, 16.89)$

- In general Proc Mixed uses Satterthwaite CI

  Default method - REML

  Versions $<$ 6.12 computed Wald CI

  Current uses Satterthwaite's Approximation

  Will discuss this CI construction later on.

Rules For Expected Mean Squares

- In models so far, EMS fairly straightforward
- Could calculate EMS using brute force method
- For mixed models, good to have formal procedure
- Montgomery describes procedure for **restricted** model

0 Write the error term in the model as $\epsilon_{(ij..)m}$, where m represents the replication subscript

1 Write each variable term in model as a row heading in a two-way table

2 Write the subscripts in the model as column headings. Over each subscript write F if factor fixed and R if random. Over this, write down the levels of each subscript

3 For each row, copy the number of observations under each subscript, providing the subscript does not appear in the row variable term

4 For any bracketed subscripts in the model, place a 1 under those subscripts that are inside the brackets

5 Fill in remaining cells with a 0 (if subscript represents a fixed factor) or a 1 (if random factor).

6 To find the expected mean square of any term (row), cover the entries in the columns that contain non-bracketed subscript letters in this term in the model. For those rows with at least the same subscripts, multiply the remaining numbers to get coefficient for corresponding term in the model.

Two-Factor Fixed Model:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

	F	F	R	
	a	b	n	
term	i	j	k	EMS
τ_i	0	b	n	$\sigma^2 + \frac{bn\Sigma\tau_i^2}{a-1}$
β_j	a	0	n	$\sigma^2 + \frac{an\Sigma\beta_j^2}{b-1}$
$(\tau\beta)_{ij}$	0	0	n	$\sigma^2 + \frac{n\Sigma\Sigma(\tau\beta)_{ij}^2}{(a-1)(b-1)}$
$\epsilon_{(ij)k}$	1	1	1	σ^2

Two-Factor Random Model:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

	R	R	R	
	a	b	n	
term	i	j	k	EMS
τ_i	1	b	n	$\sigma^2 + n\sigma^2_{\tau\beta} + bn\sigma^2_\tau$
β_j	a	1	n	$\sigma^2 + n\sigma^2_{\tau\beta} + an\sigma^2_\beta$
$(\tau\beta)_{ij}$	1	1	n	$\sigma^2 + n\sigma^2_{\tau\beta}$
$\epsilon_{(ij)k}$	1	1	1	σ^2

Two-Factor Mixed Model (A Fixed):

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

	F	R	R	
	a	b	n	
term	i	j	k	EMS
τ_i	0	b	n	$\sigma^2 + n\sigma^2_{\tau\beta} + \frac{bn\Sigma\tau_i^2}{a-1}$
β_j	a	1	n	$\sigma^2 + an\sigma^2_{\beta}$
$(\tau\beta)_{ij}$	0	1	n	$\sigma^2 + n\sigma^2_{\tau\beta}$
$\epsilon_{(ij)k}$	1	1	1	σ^2

Three-Factor Mixed Model (A Fixed):

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

	F	R	R	R	
	a	b	c	n	
term	i	j	k	l	EMS
τ_i	0	b	c	n	$\sigma^2 + cn\sigma^2_{\tau\beta} + bn\sigma^2_{\tau\gamma} + n\sigma^2_{\tau\beta\gamma} + \frac{bcn\Sigma\tau_i^2}{a-1}$
β_j	a	1	c	n	$\sigma^2 + an\sigma^2_{\beta\gamma} + acn\sigma^2_{\beta}$
γ_k	a	b	1	n	$\sigma^2 + an\sigma^2_{\beta\gamma} + abn\sigma^2_{\gamma}$
$(\tau\beta)_{ij}$	0	1	c	n	$\sigma^2 + n\sigma^2_{\tau\beta\gamma} + cn\sigma^2_{\tau\beta}$
$(\tau\gamma)_{ik}$	0	b	1	n	$\sigma^2 + n\sigma^2_{\tau\beta\gamma} + bn\sigma^2_{\tau\gamma}$
$(\beta\gamma)_{jk}$	a	1	1	n	$\sigma^2 + an\sigma^2_{\beta\gamma}$
$(\tau\beta\gamma)_{ijk}$	0	1	1	n	$\sigma^2 + n\sigma^2_{\tau\beta\gamma}$
ϵ_{ijkl}	1	1	1	1	σ^2

Two-Factor Mixed Effects Model

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

- Assume A fixed and B random

  1 $\sum \tau_i = 0$ and $\beta \sim \mathrm{N}(0, \sigma^2_\beta)$ usual assumptions

  2 $(\tau\beta)_{ij} \sim \mathrm{N}(0, (a-1)\sigma^2_{\tau\beta}/a)$ $(a-1)/a$ simplifies EMS

  3 $\sum (\tau\beta)_{ij} = 0$ for β level j added restriction

- Due to added restriction
  - Not all $(\tau\beta)_{ij}$ indep, Cov$((\tau\beta)_{ij}, (\tau\beta)_{i'j}) = -\frac{1}{a}\sigma^2_{\tau\beta}$
  - Cov($y_{ijk}, y_{i'jk'}) = \sigma^2_\beta - \frac{1}{a}\sigma^2_{\tau\beta}, i \neq i'$.

- Known as **restricted** mixed effects model

- This model coincides with EMS algorithm

  E(MS$_{\mathrm{E}}$)=σ^2

  E(MS$_{\mathrm{A}}$) = $\sigma^2 + bn\sum \tau_i^2/(a-1) + n\sigma^2_{\tau\beta}$

  E(MS$_{\mathrm{B}}$) = $\sigma^2 + an\sigma^2_\beta$

  E(MS$_{\mathrm{AB}}$) = $\sigma^2 + n\sigma^2_{\tau\beta}$

Hypotheses Testing and Diagnostics

- Testing hypotheses:

$$H_0 : \tau_1 = \tau_2 = \ldots = 0 \rightarrow \mathrm{MS_A}/\mathrm{MS_{AB}}$$

$$H_0 : \sigma^2_\beta = 0 \rightarrow \mathrm{MS_B}/\mathrm{MS_E}$$

$$H_0 : \sigma^2_{\tau\beta} = 0 \rightarrow \mathrm{MS_{AB}}/\mathrm{MS_E}$$

- Variance Estimates (Using ANOVA method)

$$\hat{\sigma}^2 = \mathrm{MS_E}$$

$$\hat{\sigma}^2_\beta = (\mathrm{MS_B} - \mathrm{MS_E})/an$$

$$\hat{\sigma}^2_{\tau\beta} = (\mathrm{MS_{AB}} - \mathrm{MS_E})/n$$

- Diagnostics
  - Histogram or QQplot

    Normality or Unusual Observations
  - Residual Plots

    Constant variance or Unusual Observations

Multiple Comparisons for Fixed Effects

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

$$\overline{y}_{i..} = \mu + \tau_i + \overline{\beta}_{.} + \overline{(\tau\beta)}_{i.} + \overline{\epsilon}_{i..}$$

$$\text{Var}(\overline{y}_{i..}) = \sigma^2_\beta/b + (a-1)\sigma^2_{\tau\beta}/ab + \sigma^2/bn$$

$$\overline{y}_{i..} - \overline{y}_{i'..} = \tau_i - \tau_{i'} + \overline{(\tau\beta)}_{i.} - \overline{(\tau\beta)}_{i'.} + \overline{\epsilon}_{i..} - \overline{\epsilon}_{i'..}$$

$$\begin{aligned} \text{Var}(\overline{y}_{i..} - \overline{y}_{i'..}) &= 2\sigma^2_{\tau\beta}/b + 2\sigma^2/bn \\ &= 2(n\sigma^2_{\tau\beta} + \sigma^2)/bn \end{aligned}$$

- Need to plug in variance estimates to compute Var($\overline{y}_{i..}$)
- What are the DF?
- For pairwise comparisons, use estimate $2\text{MS}_{AB}/bn$
- Use df$_{AB}$ for t-statistic or Tukey's method.

```
                    Gauge Capability Example in Text 12-3

options nocenter ls=75;

data randr;
 input part operator resp @@;
 cards;
1 1 21 1 1 20 1 2 20 1 2 20 1 3 19 1 3 21
2 1 24 2 1 23 2 2 24 2 2 24 2 3 23 2 3 24
3 1 20 3 1 21 3 2 19 3 2 21 3 3 20 3 3 22
4 1 27 4 1 27 4 2 28 4 2 26 4 3 27 4 3 28
.........................................
20 1 19 20 1 19 20 2 18 20 2 17 20 3 19 20 3 17
;
proc glm;
 class operator part;
 model resp=operator|part;
run;
==================================================
Dependent Variable: resp
                              Sum of
```

Source	DF	Squares	Mean Square	F Value	Pr > F
Model	59	1215.091667	20.594774	20.77	<.0001
Error	60	59.500000	0.991667		
CorrTotal	119	1274.591667			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
operator	2	2.616667	1.308333	1.32	0.2750
part	19	1185.425000	62.390789	62.92	<.0001
operator*part	38	27.050000	0.711842	0.72	0.8614

Gauge Capability Example

- $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$:

$$F_0 = \frac{MS_A}{MS_{AB}} = \frac{1.308}{0.712} = 1.84$$

P-value based on $F_{2,38}$: 0.173.

- $H_0 : \sigma_\beta^2 = 0$:

$$F_0 = \frac{MS_B}{MS_E} = \frac{62.391}{0.992} = 62.89$$

P-value based on $F_{19,60}$: 0.000

- H_0: $\sigma_{\tau\beta}^2 = 0$:

$$F_0 = \frac{MS_{AB}}{MS_E} = \frac{0.712}{0.992} = 0.72$$

P-value based on $F_{38,60}$: 0.86

- Variance components estimates:

$$\hat{\sigma}_{\beta}^{2} = \frac{62.39 - 0.99}{(3)(2)} = 10.23$$

$$\hat{\sigma}_{\tau\beta}^{2} = \frac{0.71 - 0.99}{2} = -.14(\approx 0)$$

$$\hat{\sigma}^{2} = 0.99$$

- Pairwise comparison for τ_1, τ_2 and τ_3.

Unrestricted Mixed Model

- SAS uses **unrestricted mixed model** in analysis
- $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$

  $\sum \tau_i = 0$ and $\beta_j \sim \mathrm{N}(0, \sigma^2_\beta)$

  $(\tau\beta)_{ij} \sim \mathrm{N}(0, \sigma^2_{\tau\beta})$

- Expected mean squares:

  E(MS$_{\mathrm{E}}$)=σ^2

  E(MS$_{\mathrm{A}}$) = $\sigma^2 + bn\sum\tau_i^2/(a-1) + n\sigma^2_{\tau\beta}$

  E(MS$_{\mathrm{B}}$) = $\sigma^2 + an\sigma^2_\beta + n\sigma^2_{\tau\beta}$

  E(MS$_{\mathrm{AB}}$) = $\sigma^2 + n\sigma^2_{\tau\beta}$

- random statement in SAS also gives these results
- Differences
  - E(MS$_{\mathrm{B}}$)
  - Test $H_0 : \sigma^2_\beta = 0$ using MS$_{\mathrm{AB}}$ in denominator
  - Cov($y_{ijk}, y_{i'jk'}$)=σ^2_β, $i \neq i'$.

- Connection

$$\bar{(\tau\beta)}_{.j} = (\sum_i (\tau\beta)_{ij})/a$$

$$y_{ijk} = \mu + \tau + (\beta_j + \bar{(\tau\beta)}_{.j}) + ((\tau\beta)_{ij} - \bar{(\tau\beta)}_{.j}) + \epsilon_{ijk}$$

  Check the model above satisfies the conditions of restricted mixed model

- Restricted model is slightly more general.

Gauge Capability Example (Unrestricted Model)

```
options nocenter ls=75;

data randr;
 input part operator resp @@;
 cards;
1 1 21 1 1 20 1 2 20 1 2 20 1 3 19 1 3 21
2 1 24 2 1 23 2 2 24 2 2 24 2 3 23 2 3 24
3 1 20 3 1 21 3 2 19 3 2 21 3 3 20 3 3 22
4 1 27 4 1 27 4 2 28 4 2 26 4 3 27 4 3 28
.
.
;

proc glm;
 class operator part;
 model resp=operator|part;
 random part operator*part / test;
 means operator / tukey lines E=operator*part;
 lsmeans operator / adjust=tukey E=operator*part tdiff stderr;
```

```
proc mixed alpha=.05 cl covtest;
 class operator part;
 model resp=operator / ddfm=kr;
 random part operator*part;
 lsmeans operator / alpha=.05 cl diff adjust=tukey;
run;
quit;
```

Approximate F Tests

- For some models, no exact F-test exists
- Recall 3 Factor Mixed Model (A - fixed)
- No exact test for A based on EMS

Assume $a = 3, b = 2, c = 3, n = 2$ and following MS were obtained

Source	DF	MS	EMS	F	P
A	2	0.7866	$12\phi_A + 6\sigma^2_{AB} + 4\sigma^2_{AC}$ $+2\sigma^2_{ABC} + \sigma^2$	?	?
B	1	0.0010	$18\sigma^2_B + 6\sigma^2_{BC} + \sigma^2$	0.33	.622
AB	2	0.0056	$6\sigma^2_{AB} + 2\sigma^2_{ABC} + \sigma^2$	2.24	.222
C	2	0.0560	$12\sigma^2_C + 6\sigma^2_{BC} + \sigma^2$	18.87	.051
AC	4	0.0107	$4\sigma^2_{AC} + 2\sigma^2_{ABC} + \sigma^2$	4.28	.094
BC	2	0.0030	$6\sigma^2_{BC} + \sigma^2$	10.00	.001
ABC	4	0.0025	$2\sigma^2_{ABC} + \sigma^2$	8.33	.001
Error	18	0.0003	σ^2		

- Possible approaches:
  - Could assume some variances are negligible, not recommended without "conclusive" evidence
  - Pool (insignificant) means squares with error, also risky, not recommended when df for error is already big.

Satterthwaite's Approximate F-test

- H_0 : effect $= 0$, e.g., $H_0 : \tau_1 = \cdots = \tau_a = 0$ or equivalently $H_0 : \sum \tau_i^2 = 0$.
  No exact test exists.

- Get two linear combinations of mean squares

  $$MS' = \mathrm{MS}_r \pm \ldots \pm \mathrm{MS}_s$$

  $$MS'' = \mathrm{MS}_u \pm \ldots \pm \mathrm{MS}_v$$

  such that 1) MS' and MS'' do not share common mean squares; 2) $E(MS') - E(MS'')$ is a multiple of the effect.

  - approximate test statistic F: $F = \frac{MS'}{MS''} = \frac{\mathrm{MS}_r \pm \ldots \pm \mathrm{MS}_s}{\mathrm{MS}_u \pm \ldots \pm \mathrm{MS}_v} \approx F_{p,q}$
    where $p = \frac{(\mathrm{MS}_r \pm \ldots \pm \mathrm{MS}_s)^2}{\mathrm{MS}_r^2/f_r + \ldots + \mathrm{MS}_s^2/f_s}$ and $q = \frac{(\mathrm{MS}_u \pm \ldots \pm \mathrm{MS}_v)^2}{\mathrm{MS}_u^2/f_u + \ldots + \mathrm{MS}_v^2/f_v}$
  - f_i is the degrees of freedom associated with MS_i
  - p and q may not be integers, interpolation is needed. SAS can handle noninteger dfs.
  - Caution when subtraction is used

Example: 3-Factor Mixed Model (A Fixed)

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$

$$MS' = \mathrm{MS}_A$$

$$MS'' = \mathrm{MS}_{AB} + \mathrm{MS}_{AC} - \mathrm{MS}_{ABC}$$

$$E(MS' - MS'') = 12\phi_A = 12\tfrac{\Sigma\tau_i^2}{3-1}$$

$$F = \frac{\mathrm{MS}_A}{\mathrm{MS}_{AB} + \mathrm{MS}_{AC} - \mathrm{MS}_{ABC}} = \frac{.7866}{.0107 + .0056 - .0025} = 57.0$$

$$p = 2 \qquad q = \frac{.0138^2}{.0107^2/4 + .0056^2/2 + .0025^2/4} = 4.15$$

- Interpolation needed

$$\mathrm{P}(F_{2,4} > 57) = .0011 \qquad \mathrm{P}(F_{2,5} > 57) = .0004$$

$$\mathrm{P} = .85(.0011) + .15(.0004) = .001$$

- SAS can be used to compute P-values and quantile values for F and χ^2 values with noninteger degrees of freedom.

  Upper Tail Probability: probf(x,df1,df2) and probchi(x,df)

  Quantiles : finv(p,df1,df2) and cinv(p,df)

```
data one;
  p=1-probf(57,2.0,4.15);
  f=finv(.95,2.0,4.15);
  c1=cinv(.025,18.57);
  c2=cinv(.975,18.57);
proc print data=one;
```

```
OBS        P           F           C1          C2
 1     .00096     6.7156      8.61485    32.2833
```

Another Approach to Testing $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$

$MS' = \mathrm{MS}_A + \mathrm{MS}_{ABC}$

$MS'' = \mathrm{MS}_{AB} + \mathrm{MS}_{AC}$

$E(MS' - MS'') = ?$

$$F = \frac{\mathrm{MS}_A + \mathrm{MS}_{ABC}}{\mathrm{MS}_{AB} + \mathrm{MS}_{AC}} = \frac{.7866 + .0025}{.0107 + .0056} = 48.41$$

$$p = \frac{.7891^2}{.7866^2/2 + .0025^2/4} = 2.01 \qquad q = \frac{.0163^2}{.0107^2/4 + .0056^2/2} = 6.00$$

P-value= $P(F > 48.41) = 0.002$

- This is again found significant
- Avoid subtraction,summation should be preferred.

Approximate Confidence Intervals

Suppose we are interested in σ_x^2.

- Case 1: there exists a mean square MS_x with df$_x$ such that $E(MS_x) = \sigma_x^2$. Then $\hat{\sigma}_x^2 = MS_x$, and

$$\frac{\text{df}_x MS_x}{\sigma_x^2} \sim \chi^2(\text{df}_x)$$

Exact 100(1-α)% CI: $\frac{\mathrm{df}_x MS_x}{\chi^2_{\alpha/2,\mathrm{df}_x}} \leq \sigma_x^2 \leq \frac{\mathrm{df}_x MS_x}{\chi^2_{1-\alpha/2,\mathrm{df}_x}}$

- Case 2: there exist

$MS' = \mathrm{MS}_r + \ldots + \mathrm{MS}_s$ and, $MS'' = \mathrm{MS}_u + \ldots + \mathrm{MS}_v$

such that $E(MS' - MS'') = k\sigma_x^2$. Then

$$\hat{\sigma}_x^2 = \frac{MS' - MS''}{k}, \text{ and } \frac{\text{df}_x \hat{\sigma}_x^2}{\sigma_x^2} \approx \chi^2(\text{df}_x)$$

where

$$\text{df}_x = \frac{(\hat{\sigma}_x^2)^2}{\sum \frac{MS_i}{k^2 f_i}} = \frac{(\mathrm{MS}_r + \cdot + \mathrm{MS}_s - \mathrm{MS}_u - \cdot - \mathrm{MS}_v)^2}{\mathrm{MS}_r^2/f_r + \cdot + \mathrm{MS}_s^2/f_s + \mathrm{MS}_u^2/f_u + \cdot + \mathrm{MS}_v^2/f_v}$$

Approximate 100(1-α)% CI:

$$\frac{\mathrm{df}_x \hat{\sigma}_x^2}{\chi^2_{\alpha/2,\mathrm{df}_x}} \leq \sigma_x^2 \leq \frac{\mathrm{df}_x \hat{\sigma}_x^2}{\chi^2_{1-\alpha/2,\mathrm{df}_x}}$$

Gauge Capability Example (Both Factors are Random)

Dependent Variable: RESP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	59	1215.09166667	20.594774	20.77	0.0001
Error	60	59.50000000	0.991667		
CorrecTotal	119	1274.59166667			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
OPERATOR	2	2.61666667	1.308333	1.32	0.2750
PART	19	1185.42500000	62.390789	62.92	0.0001
OPERATOR*PART	38	27.05000000	0.711842	0.72	0.8614

$\hat{\sigma}^2_\tau = \frac{MS_A - MS_{AB}}{bn} = (62.39 - 0.71)/6 = 10.28$

$$\text{df} = \frac{(62.39 - 0.71)^2}{62.39^2/19 + 0.71^2/38} = 18.57$$

CI: $(18.57(10.28)/32.28, 18.57(10.28)/8.61) = (5.91, 22.17)$

$$\hat{\sigma}_\beta^2 = \frac{MS_B - MS_{AB}}{an} = (1.31 - 0.71)/40 = 0.015$$

$$\text{df} = \frac{(1.31 - 0.71)^2}{1.31^2/2 + 0.71^2/38} = .413$$

CI: $(.413(.015)/3.079, .413(.015)/2.29 \times 10^{-8}) = (.002, 270781)$

General Mixed Effect Model

- In terms of linear model

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

  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 (allows correlation)
- 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
- Unrestricted linear mixed model is default