Random and Mixed Factorial Designs II

Design of Experiments - Montgomery
Chapter 12

A. Two-Factor Mixed Effects Model

- Same model but different parameter restrictions
- Assume A fixed and B random
  1. $\sum \tau_i = 0$ and $\beta \sim$ N$(0, \sigma^2_\beta)$ usual assumptions
  2. $(\tau\beta)_{ij} \sim$ 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$((\tau\beta)_{ij}, (\tau\beta)_{i'j}) = -\frac{1}{a}\sigma^2_{\tau\beta}$
- Known as **restricted** mixed effects model
- This model coincides with EMS algorithm

  E(MS$_E$)$=\sigma^2$

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

  E(MS$_B$) $= \sigma^2 + an\sigma^2_\beta$

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

NOTE: If $X_i \sim N(0, \sigma^2)$ then $\begin{cases} X_i - \overline{X} \sim N(0, \frac{n-1}{n}\sigma^2) \\ \text{Cov}(X_i - \overline{X}, X_j - \overline{X}) = -\frac{1}{n}\sigma^2 \end{cases}$

Hypothesis Tests and Diagnostics

- Hypothesis tests

  $H_0: \tau_1 = \tau_2 = \ldots = 0 \rightarrow$ MS$_A$/MS$_{AB}$

  $H_0: \sigma^2_\beta = 0 \rightarrow$ MS$_B$/MS$_E$

  $H_0: \sigma^2_{\tau\beta} = 0 \rightarrow$ MS$_{AB}$/MS$_E$
- Variance Estimates (Using ANOVA method)

  $\hat{\sigma}^2 =$ MS$_E$

  $\hat{\sigma}^2_\beta =$ (MS$_B$ − MS$_E$)$/an$

  $\hat{\sigma}^2_{\tau\beta} =$ (MS$_{AB}$ − MS$_E$)$/n$
- Diagnostics
  - Histogram or QQplot

    Normality or Unusual Observations
  - Residual Plots

    Constant variance or Unusual Observations

Multiple Comparisons

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

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

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

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

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

$$= 2(n\sigma^2_{\tau\beta} + \sigma^2)/bn$$

- Need to plug in variance estimates to compute Var$(\overline{y}_{i..})$
- What are the DF?
- For pairwise comparisons, use estimate 2MS$_{AB}/bn$
- Use df$_{AB}$ for t-statistic

Sample Size Calculations

Use Charts V and VI

Random Effects Model

Factor	λ	DF_{Num}	DF_{Den}
A	$\sqrt{1+\frac{bn\sigma_\tau^2}{\sigma^2+n\sigma_{\tau\beta}^2}}$	$a-1$	$(a-1)(b-1)$
B	$\sqrt{1+\frac{an\sigma_\beta^2}{\sigma^2+n\sigma_{\tau\beta}^2}}$	$b-1$	$(a-1)(b-1)$
AB	$\sqrt{1+\frac{n\sigma_{\tau\beta}^2}{\sigma^2}}$	$(a-1)(b-1)$	$ab(n-1)$

Mixed Effects Model

Factor	λ or Φ	DF_{Num}	DF_{Den}
A	$\sqrt{\frac{bn\sum\tau_i^2}{a(\sigma^2+n\sigma_{\tau\beta}^2)}}$	$a-1$	$(a-1)(b-1)$
B	$\sqrt{1+\frac{an\sigma_\beta^2}{\sigma^2}}$	$b-1$	$ab(n-1)$
AB	$\sqrt{1+\frac{n\sigma_{\tau\beta}^2}{\sigma^2}}$	$(a-1)(b-1)$	$ab(n-1)$

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
.
.
;

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;
```



```
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
Corrected Total         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) + Q(operator)
part                    Var(Error) + 2 Var(operator*part) + 6 Var(part)
operator*part           Var(Error) + 2 Var(operator*part)

Tests of Hypotheses for Mixed Model Analysis of Variance

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
```



```
Tukey's Studentized Range (HSD) Test for resp

Alpha                                     0.05
Error Degrees of Freedom                    38
Error Mean Square                     0.711842
Critical Value of Studentized Range  3.44902
Minimum Significant Difference        0.4601

          Mean      N    operator
A       22.6000     40    3
A
A       22.3000     40    1
A
A       22.2750     40    2

Standard Errors and Probabilities Calculated Using the Type III MS for
operator*part as an Error Term
                              Standard                    LSMEAN
operator     resp LSMEAN         Error    Pr > |t|        Number
1             22.3000000     0.1334018      <.0001             1
2             22.2750000     0.1334018      <.0001             2
3             22.6000000     0.1334018      <.0001             3

   Least Squares Means for Effect operator
   t for H0: LSMean(i)=LSMean(j) / Pr > |t|
          Dependent Variable: resp

i/j              1             2             3
   1                    0.132514      -1.59017
                          0.9904        0.2622
   2      -0.13251                    -1.72269
            0.9904                      0.2100
   3      1.590173      1.722688
            0.2622        0.2100
```

The Mixed Procedure

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	622.27805725	
1	2	409.45998838	0.00002843
2	1	409.45716449	0.00000003
3	1	409.45716136	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
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

Fit Statistics

-2 Res Log Likelihood	409.5
AIC (smaller is better)	413.5
AICC (smaller is better)	413.6
BIC (smaller is better)	415.4

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F	
operator	2	38	1.48	0.2401	
operator	2	98	1.48	0.2324	*** KR adjustment

Unrestricted Model

$\text{Var}(\overline{y}_{1..}) = (\sigma^2 + n\sigma^2_{\tau\beta} + n\sigma^2_\beta)/bn = (.8832 + 2(10.2513))/40$

$\text{Var}(\overline{y}_{1..} - \overline{y}_{2..}) = 2(\sigma^2 + n\sigma^2_{\tau\beta})/bn = .8832/20$

Least Squares Means

Effect	operator	Estimate	Standard Error	DF	t Value	Pr > |t|
operator	1	22.3000	0.7312	20.1	30.50	<.0001
operator	2	22.2750	0.7312	20.1	30.46	<.0001
operator	3	22.6000	0.7312	20.1	30.91	<.0001

Differences of Least Squares Means

Effect	operator	_operator	Estimate	Standard Error	DF	t Value	Pr > |t|
operator	1	2	0.02500	0.2101	38	0.12	0.9059
operator	1	3	-0.3000	0.2101	38	-1.43	0.1616
operator	2	3	-0.3250	0.2101	38	-1.55	0.1302
operator	1	2	0.02500	0.2101	98	0.12	0.9055**
operator	1	3	-0.3000	0.2101	98	-1.43	0.1566**
operator	2	3	-0.3250	0.2101	98	-1.55	0.1252**

Differences of Least Squares Means

Effect	operator	_operator	Adjustment	Adj P	Alpha
operator	1	2	Tukey-Kramer	0.9922	0.05
operator	1	3	Tukey-Kramer	0.3371	0.05
operator	2	3	Tukey-Kramer	0.2811	0.05

Differences of Least Squares Means

Effect	operator	_operator	Lower	Upper	Adj Lower	Adj Upper
operator	1	2	-0.4004	0.4504	-0.4875	0.5375
operator	1	3	-0.7254	0.1254	-0.8125	0.2125
operator	2	3	-0.7504	0.1004	-0.8375	0.1875
operator	1	2	-0.3920	0.4420	-0.4751	0.5251**
operator	1	3	-0.7170	0.1170	-0.8001	0.2001**
operator	2	3	-0.7420	0.09201	-0.8251	0.1751**

Other Mixed Models

- SAS uses different mixed model in analysis
- Reduce parameter restrictions

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

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

- Known as **unrestricted** mixed model
- For two-factor model

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

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

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

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

- Random statement in SAS also gives these results
- Differences
  - Test $H_0 : \sigma^2_\beta = 0$ using MS_AB in denominator
  - Often more conservative test, $\hat{\sigma}^2_\beta = (\text{MS}_\text{B} - \text{MS}_\text{AB})/an$

To decide which is appropriate, suppose you ran experiment again and sampled the same random effects. Should this mean you also have the same set of interaction effects? Yes: Restricted No: Unrestricted

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)
- $\text{Cov}(Y) = ZGZ' + R$
- 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
- Unrestricted linear mixed model is default

Example

A corporation wants to compare two different sunscreens for protecting the skin of adults age 20-25 from burning/tanning. A random sample of 10 subjects ages 20-25 were chosen for the study. With each person, four squares on the back were marked and each sunscreen was randomly applied to two of the squares. The color of skin was noted prior to treatment and then after a two hour period of sun bathing. The difference was recorded. A large positive difference means less protection.

- What are the factors in the model?

- Which are random and which are fixed?

Results

Subject	Sunscreen 1		Sunscreen 2	
1	8.2	7.6	6.1	6.8
2	3.6	3.5	4.3	4.7
3	10.7	10.3	9.6	9.2
4	3.9	4.4	2.3	2.5
5	12.9	12.1	12.4	12.8
6	5.5	5.9	4.8	4.0
7	9.1	9.7	8.3	8.6
8	13.7	13.2	12.9	13.6
9	8.1	8.7	8.0	7.5
10	2.5	2.8	2.1	2.5

Which appears to be better?

```
/* sunscreen.sas */
options nocenter ls=75 ps=60;

data new;
infile "sunscreen.dat";
input subject lotion resp;

proc mixed covtest cl maxiter=20;
 class subject lotion;
 model resp=lotion / ddfm=kr;
 random subject subject*lotion;
 lsmeans lotion / diff cl;
run;
quit;
```



```
                    Covariance Parameter Estimates
                         Standard       Z
Cov Parm        Estimate    Error   Value     Pr Z  Alpha    Lower    Upper
subject          14.2086   6.7767    2.10   0.0180   0.05   6.6748  48.2352
subject*lotion    0.2660   0.1579    1.68   0.0460   0.05   0.1084   1.3723
Residual          0.1320  0.04174    3.16   0.0008   0.05  0.07726   0.2753

        Type 3 Tests of Fixed Effects
              Num     Den
Effect         DF      DF    F Value    Pr > F
lotion          1       9       6.76    0.0287

    1. Significant source of variation due to combination
          - one lotion may not be best for all subjects
    2. Significant subject-to-subject variability
    3. Lotion 2 ‘‘on average’’ offers more protection
    4. Is difference practically significant?

                              Least Squares Means
                                Standard
Effect    lotion    Estimate       Error      DF    t Value    Pr > |t|     Alpha
lotion    1           7.8200      1.2058    9.21       6.49      0.0001      0.05
lotion    2           7.1500      1.2058    9.21       5.93      0.0002      0.05

                    Differences of Least Squares Means
                                       Standard
Effect    lotion    _lotion    Estimate       Error      DF    t Value    Pr > |t|
lotion    1         2            0.6700      0.2577       9       2.60      0.0287
```

For restricted model

Var($\overline{y}_{i..}$) $= (\sigma^2 + (a-1)n\sigma^2_{\tau\beta}/a + n\sigma^2_\beta)/bn = (.1320 + 0.2660 + 2(14.2086))/20 = 1.44$ which in slightly smaller than the unrestricted model.

B. Approximate F Tests and CI

- 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		

1 Could assume some variances negligible

   – Not recommended without "conclusive" evidence

Examples

- If you assume σ^2_{ABC} and σ^2_{AB} equals 0

Source	DF	MS	EMS	F	P
A	2	0.7866	$12\phi_A + 4\sigma^2_{AC} + \sigma^2$	73.54	.001
B	1	0.0010	$18\sigma^2_B + 6\sigma^2_{BC} + \sigma^2$	0.33	.622
C	2	0.0560	$12\sigma^2_C + 6\sigma^2_{BC} + \sigma^2$	18.87	.051
AC	4	0.0107	$4\sigma^2_{AC} + \sigma^2$	11.89	.001
BC	2	0.0030	$6\sigma^2_{BC} + \sigma^2$	3.33	.053
Error	24	0.0009	σ^2		

- If you assume σ^2_{AC} and σ^2_{AB} equals 0

Source	DF	MS	EMS	F	P
A	2	0.7866	$12\phi_A + 2\sigma^2_{ABC} + \sigma^2$	314.64	.001
B	1	0.0010	$18\sigma^2_B + 6\sigma^2_{BC} + \sigma^2$	0.33	.622
C	2	0.0560	$12\sigma^2_C + 6\sigma^2_{BC} + \sigma^2$	18.87	.051
BC	2	0.0030	$6\sigma^2_{BC} + \sigma^2$	1.2	.319
ABC	4	0.0025	$2\sigma^2_{ABC} + \sigma^2$	1.0	.427
Error	24	0.0025	σ^2		

2 Could test interactions and then possibly remove

- Based on first table (18-17),

  AC and AB found insignificant

  Test A over ABC → F=314.64 and P $<$ .001
- Both Type I and Type II errors possible
- What level to test insignificance?

3 **Pooling** Mean Squares with Error

- Variation of second approach
- Works well when df for error is small ($<$ 6)
- Often test significance at $\alpha = .25$
- May pool together something that is different from zero
- Use high α to protect against that

Pooling Procedure

1 Test highest order interaction vs error

2 If ABC found insignificant, pool together mean squares

$$\text{MS}'_E = \frac{\text{SS}_\text{E} + \text{SS}_{ABC}}{\text{df}_\text{E} + \text{df}_{ABC}}$$

3 Continue by testing AB, AC, BC over new error and pool accordingly

4 In SAS, pooling accomplished by simply dropping term from model

- Procedure primarily used for error, not interactions
- If higher order interaction found significant - stop
- Pooling procedure of no benefit in example

Satterthwaite's Approximate F-test

4 Use linear combination of mean squares

- To test certain factor, choose numerator and denominator such that the difference in MS is a multiple of the effect of interest
- Ratio approximately F where

$$\text{F}_{p,q} = \frac{\text{MS}_r \pm \ldots \pm \text{MS}_s}{\text{MS}_u \pm \ldots \pm \text{MS}_v}$$

$$p = \frac{(\text{MS}_r \pm \ldots \pm \text{MS}_s)^2}{\text{MS}_r^2/f_r + \ldots + \text{MS}_s^2/f_s}$$

$$q = \frac{(\text{MS}_u \pm \ldots \pm \text{MS}_v)^2}{\text{MS}_u^2/f_u + \ldots + \text{MS}_v^2/f_v}$$

- f_i is the degrees of freedom associated with MS_i
- No MS in both num and denom (indep)
- Caution when subtraction is used

Example

For the 3 factor model,

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

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

- Interpolation needed

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

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

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

P-values : probf(x,df1,df2) and probchi(x,df)

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

```
data pvalue;
  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;
```

```
OBS        P           F          C1          C2

 1      .00096      6.7156      8.61485     32.2833
```


Example

For the 3 factor model (avoiding subtraction),

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

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

- This is again found significant

Confidence Intervals

- Use Satterthwaite's Pseudo F-tests to create CI
- Recall $\text{df}_\text{E}\text{MSE}/\sigma^2 \sim \chi^2$

$$\frac{\text{df}_E\text{MSE}}{\chi^2_{\alpha/2,\text{df}_E}} \le \sigma^2 \le \frac{\text{df}_E\text{MSE}}{\chi^2_{1-\alpha/2,\text{df}_E}}$$

- Using Pseudo F-tests, $\hat{\sigma}^2 = \text{MS}' - \text{MS}''$
- Both MS are indep and have similar χ^2 distribution
- Assume linear combination of χ^2 is χ^2 with df

$$\frac{(\text{MS}_r+\ldots+\text{MS}_s-\text{MS}_u-\ldots-\text{MS}_v)^2}{\text{MS}_r^2/f_r+\ldots+\text{MS}_s^2/f_s+\text{MS}_u^2/f_u+\ldots+\text{MS}_v^2/f_v}$$

- Use same CI formula as above

Random Effects Example 12-2

```
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
Corrected Total        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
```

$$\hat{\sigma}^2_\tau = (62.39-0.71)/6 = 10.28$$

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

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

$$\hat{\sigma}^2_\beta = (1.31-0.71)/40 = 0.015$$

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

CI: $(.413(.015)/3.079, .413(.015)/2.29E{-}8) = (.002, 270781)$