Lecture 12: Nested and Split-Plot Designs

Montgomery, Chapter 14

Crossed vs Nested Factors

- Factors A (a levels)and B (b levels) are considered crossed if

  Every combinations of A and B (ab of them) occurs.

  An example:

	Factor A			
Factor B	1	2	3	4
1	xx	xx	xx	xx
2	xx	xx	xx	xx
3	xx	xx	xx	xx

A	1			2			3			4		
B	1	2	3	1	2	3	1	2	3	1	2	3
	x	x	x	x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	x	x	x	x	x

- Factor B is considered nested under A (a levels) if
  1. under each fixed level (i) of A, B has b_i levels.
  2. the levels of B under the same level of A are comparable.
  3. under a level of A, the levels of B can be arbitrarily numbered.

A	1			2			3			4		
B	1	2	3	4	5	6	7	8	9	10	11	12
	x	x	x	x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	x	x	x	x	x

Material Purity Experiment

Consider a company that buys raw material in batches from three different suppliers. The purity of this raw material varies considerably, which causes problems in manufacturing the finished product. We wish to determine if the variability in purity is attributable to difference between the suppliers. Four batches of raw material are selected at random from each supplier, three determinations of purity are made on each batch. The data, after coding by subtracting 93 are given below.

	Supplier 1				Supplier 2				Supplier 3			
Batches	1	2	3	4	1	2	3	4	1	2	3	4
	1	-2	-2	1	1	0	-1	0	2	-2	1	3
	-1	-3	0	4	-2	4	0	3	4	0	-1	2
	0	-4	1	0	-3	2	-2	2	0	2	2	1
$y_{ij.}$	0	-9	-1	5	-4	6	-3	5	6	0	2	6
$y_{i..}$	-5				4				14			

Other Examples for Nested Factors

1 Drug company interested in stability of product

- Two manufacturing sites
- Three batches from each site
- Ten tablets from each batch

2 Stratified random sampling procedure

- Randomly sample five states
- Randomly select three counties
- Randomly select two towns
- Randomly select five households

Statistical Model

- Two factor nested model

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

- Bracket notation represents nested factor
- Cannot include interaction
- Factors may be random or fixed
- Can use EMS algorithm to derive tests

Sum of Squares Decomposition

$$y_{ijk} = \bar{y}_{\ldots} + (\bar{y}_{i\ldots} - \bar{y}_{\ldots}) + (\bar{y}_{ij.} - \bar{y}_{i\ldots}) + (y_{ijk} - \bar{y}_{ij.}).$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{\ldots})^2 = bn\sum_{i=1}^{a}(\bar{y}_{i\ldots} - \bar{y}_{\ldots})^2 + n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i\ldots})^2$$

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2$$

$$\mathrm{SS_T} = \mathrm{SS_A} + \mathrm{SS_{B(A)}} + \mathrm{SS_E}$$

Analysis of Variance Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A	SS_{A}	$a-1$	MS_{A}	
B(A)	$\text{SS}_{\text{B(A)}}$	$a(b-1)$	$\text{MS}_{\text{B(A)}}$	
Error	SS_{E}	$ab(n-1)$	MS_{E}	
Total	SS_{T}	$abn-1$		

$$\text{SS}_{\text{T}} = \sum\sum\sum y_{ijk}^2 - y_{...}^2/abn$$

$$\text{SS}_{\text{A}} = \frac{1}{bn}\sum y_{i..}^2 - y_{...}^2/abn$$

$$\text{SS}_{\text{B(A)}} = \frac{1}{n}\sum\sum y_{ij.}^2 - \frac{1}{bn}\sum y_{i..}^2$$

$$\text{SS}_{\text{E}} = \sum\sum\sum y_{ijk}^2 - \frac{1}{n}\sum\sum y_{ij.}^2$$

- Use EMS to define proper tests

Two-Factor Nested Model with Fixed Effects:

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

where (1) $\sum_{i=1}^{a} \tau_i = 0$, (2) $\sum_{j=1}^{b} \beta_{j(i)} = 0$ for each i.

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

- Estimates: $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$; $\hat{\beta}_{j(i)} = \bar{y}_{ij.} - \bar{y}_{i..}$.
- Tests: $\mathrm{MS_A}/\mathrm{MS_E}$ for $\tau_i = 0$; $\mathrm{MS_{B(A)}}/\mathrm{MS_E}$ for $\beta_{j(i)} = 0$.

Two-Factor Nested Model with Random Effects:

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

where $\tau_i \sim N(0, \sigma_\tau^2)$ and $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$.

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

- Estimates: $\hat{\sigma}_\tau^2 = (\mathrm{MS_A} - \mathrm{MS_{B(A)}})/nb$; $\hat{\sigma}_\beta^2 = (\mathrm{MS_{B(A)}} - \mathrm{MS_E})/n$.
- tests: $\mathrm{MS_A}/\mathrm{MS_{B(A)}}$ for $\sigma_\tau^2 = 0$; $\mathrm{MS_{B(A)}}/\mathrm{MS_E}$ for $\sigma_\beta^2 = 0$.

Two-Factor Nested Model with Mixed Effects:

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

where $\sum_{i=1}^{a} \tau_i = 0$, and $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$.

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

- Estimates: $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$; $\hat{\sigma}_\beta^2 = (\mathrm{MS}_{\mathrm{B(A)}} - \mathrm{MS}_{\mathrm{E}})/n$.
- Tests: $\mathrm{MS}_{\mathrm{A}}/\mathrm{MS}_{\mathrm{B(A)}}$ for $\tau_i = 0$; $\mathrm{MS}_{\mathrm{B(A)}}/\mathrm{MS}_{\mathrm{E}}$ for $\sigma_\beta^2 = 0$.

SAS Code for Purity Experiment

```
option nocenter ps=40 ls=72;
data purity;
input supp batch resp@@;
datalines;
1 1 1 1 1 -1 1 1 0
1 2 -2 1 2 -3 1 2 -4
1 3 -2 1 3 0 1 3 1
1 4 1 1 4 4 1 4 0
2 1 1 2 1 -2 2 1 -3
2 2 0 2 2 4 2 2 2
2 3 -1 2 3 0 2 3 -2
2 4 0 2 4 3 2 4 2
3 1 2 3 1 4 3 1 0
3 2 -2 3 2 0 3 2 2
3 3 1 3 3 -1 3 3 2
3 4 3 3 4 2 3 4 1
```

```
;

proc mixed method=type1;
class supp batch;
model resp=;
random supp batch(supp);
run;

proc mixed method=type1;
class supp batch;
model resp=supp;
random batch(supp);
run;
quit;
```

Both suppliers and batches are random effects

Source	DF	Sum of Squares	Mean Square
supp	2	15.055556	7.527778
batch(supp)	9	69.916667	7.768519
Residual	24	63.333333	2.638889

Source	Expected Mean Square	Error Term
supp	Var(Residual) + 3 Var(batch(supp)) + 12 Var(supp)	MS(batch(supp))
batch(supp)	Var(Residual) + 3 Var(batch(supp))	MS(Residual)
Residual	Var(Residual)	.

Source	DF	F Value	Pr > F
supp	9	0.97	0.4158
batch(supp)	24	2.94	0.0167
Residual	.	.	.

Covariance Parameter Estimates

Cov Parm	Estimate
supp	-0.02006
batch(supp)	1.7099
Residual	2.6389

Suppliers are fixed effects and batches are random

Source	DF	Sum of Squares	Mean Square
supp	2	15.055556	7.527778
batch(supp)	9	69.916667	7.768519
Residual	24	63.333333	2.638889

Source	Expected Mean Square	Error Term
supp	Var(Residual) + 3 Var(batch(supp)) + Q(supp)	MS(batch(supp))
batch(supp)	Var(Residual) + 3 Var(batch(supp))	MS(Residual)
Residual	Var(Residual)	

Source	DF	F Value	Pr > F
supp	9	0.97	0.4158
batch(supp)	24	2.94	0.0167
Residual	.	.	.

```
Covariance Parameter Estimates

Cov Parm          Estimate
batch(supp)         1.7099
Residual            2.6389
```

Results summary when suppliers are fixed effects

- Estimates:

$$\hat{\tau}_1 = \bar{y}_{1..} - \bar{y}_{...} = -28/36$$

$$\hat{\tau}_2 = \bar{y}_{2..} - \bar{y}_{...} = -1/36$$

$$\hat{\tau}_3 = \bar{y}_{3..} - \bar{y}_{...} = -29/36$$

$$\hat{\sigma}^2 = MS_E = 2.64$$

$$\hat{\sigma}^2_\beta = \frac{MS_{B(A)} - MS_E}{n} = \frac{7.77 - 2.64}{3} = 1.71$$

- Hypothesis test

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

$$F_0 = .97, \text{P-value} = 0.4158, \text{Accept } H_0$$

$H_0 : \sigma^2_\beta = 0$:

$$F_0 = 2.94, \text{P-value} = 0.0167, \text{Reject } H_0$$

- Suppliers are not different, variability due to batches.

Other Scenarios for Nested Factors

- Staggered Nested Designs
- General m-Stage Nested Designs

$$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \epsilon_{l(ijk)}$$

- Designs with Both Nested and Factorial Factors

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_{k(j)} + (\tau\beta)_{ij} + (\tau\gamma)_{ik(j)} + \epsilon_{l(ijk)}$$

- Sections 14.2, 14.3 in Montgomery.

Split-Plot Designs

- Example 1: Study six corn varieties and four fertilizers and yield is the response. Three replicates are needed.

  Method 1: completely randomized full factorial design, 24 level combinations of variety and fertilizer are applied to 24*3=72 pieces of land (each to three).

  Method 2: Select three fields of large area. Each field is divided into four areas (four whole-plots), four fertilizers are randomly assigned to the four whole-plots. Each area is further divided into six subareas (sub-plots), and the six varieties are randomly planted in these sub-plots.

  This leads to a split-plot design:

- whole-plot (treatment) factor: fertilizer
- sub-plot (treatment) factor: corn variety

- Example 2: A paper manufacturer is investigating three different pulp preparation methods and four different cooking temperatures for the pulp and study their effect on the tensile strength of the paper. Three replicates are needed.
- Because the pilot plant is only capable of making 12 runs per day, so the experimenter decides to run one replicate on each of the three days and to consider the days as blocks.
- On any day, a batch of pulp is produced by one of the the three methods (a whole-plot). Then the batch is divided into four samples (four sub-plots), and each sample is cooked at one of the four temperatures. Then a second batch of pulp is made up using another of the three methods. This second batch is also divided into four samples that are tested at the four temperatures. The process is then repeated for the third method. The data is given below.
- whole-plot factor: preparation method
- sub-plot factor: cooking temperature

	Day 1			Day 2			Day 3		
Method	1	2	3	1	2	3	1	2	3
Temp									
200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	27	40	34
250	27	38	33	40	42	32	41	39	39
275	36	42	36	41	40	40	40	44	45

Split-Plot Structure

- factors are crossed (different than nested)
- randomization restriction (different than completely randomized)
- Information on factor effects from two levels (or strata).
- split-plot can be considered as two superimposed blocked designs:
  - A: whole-plot factor(a); B: sub-plot factor (b), r replicates
  - RCBD_A: number of trt: a, number of blk: r.
  - RCBD_B: number of trt: b, number of blk: ra.

  for whole-plots, subdivision to smaller sub-plots are ignored. For sub-plots, whole-plots considered blocks.
- More power for main subplot effect and interaction
- Should use design only for practical reasons
- Randomized factorial design more powerful if feasible

A typical Data Layout

	Block 1			Block 2			Block 3		
WP-Factor A	1	2	3	1	2	3	1	2	3
SP-Factor B									
1	y_{111}	y_{121}	.	.	.	.	.	.	y_{331}
2	y_{112}	y_{122}	.	.	.	.	.	.	y_{332}
3	y_{113}	y_{123}	.	.	.	.	.	.	y_{333}
4	y_{114}	y_{124}	.	.	.	.	.	.	y_{334}

In general:

y_{ijk} where i denotes Block i, j denotes the jth level of the whole-plot factor A, and k denotes the kth level of the sub-plot factor B.

Statistical Model I

- $$y_{ijk} = \mu + r_i + \alpha_j + (r\alpha)_{ij} + \beta_k + (r\beta)_{ik} + (\alpha\beta)_{jk} + (r\alpha\beta)_{ijk} + \epsilon_{ijk}$$

  $$i = 1, 2, \ldots, r, j = 1, 2, \ldots, a, k = 1, 2, \ldots, b$$

  - r_i: block effects (random) $\sim N(0, \sigma_r^2)$
  - α_j: whole-plot factor (A) main effects (fixed)
  - $(r\alpha)_{ij}$: whole-plot error (random) $\sim$ normal with $\sigma^2_{r\alpha}$.
  - β_k: sub-plot factor (B) main effects (fixed)
  - $(r\beta)_{ik}$: block-B interaction (random) $\sim$ normal with $\sigma^2_{r\beta}$.
  - $(\alpha\beta)_{jk}$ Interaction between A and B (fixed)
  - $(r\alpha\beta)_{ijk}$: sub-plot error (random) $\sim$ normal with $\sigma^2_{r\alpha\beta}$
  - ϵ_{ijk}: random error $\sim N(0, \sigma^2)$

Sum of Squares

- $\mathrm{SS_r} = ab\sum_i(\bar{y}_{i..} - \bar{y}_{...})^2$, df=r-1.
- $\mathrm{SS_A} = rb\sum_j(\bar{y}_{.j.} - \bar{y}_{...})^2$, df=a-1.
- $\mathrm{SS_{rA}} = b\sum_{i,j}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$, df=(r-1)(a-1)
- $\mathrm{SS_B} = ar\sum_k(\bar{y}_{..k} - \bar{y}_{...})^2$, df=(b-1)
- $\mathrm{SS_{rB}} = a\sum_{i,k}(\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...})^2$ df=(r-1)(b-1)
- $\mathrm{SS_{AB}} = r\sum_{j,k}(\bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...})^2$ df=(a-1)(b-1)
- $\mathrm{SS_{rAB}} = \sum_{i,j,k}(y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...})^2$, df=(r-1)(a-1)(b-1).
- $\mathrm{SS_E} =$?

Expected mean squares (restricted)

		r	a	b	1	
		R	F	F	R	
	term	i	j	k	h	$E(MS)$
	r_i	1	a	b	1	$\sigma^2 + ab\sigma_r^2$
whole plot	α_j	r	0	b	1	$\sigma^2 + b\sigma_{r\alpha}^2 + \frac{rb\Sigma\alpha_j^2}{a-1}$
	$(r\alpha)_{ij}$	1	0	b	1	$\sigma^2 + b\sigma_{r\alpha}^2$
	β_k	r	a	0	1	$\sigma^2 + a\sigma_{r\beta}^2 + \frac{ra\Sigma\beta_k^2}{b-1}$
	$(r\beta)_{ik}$	1	a	0	1	$\sigma^2 + a\sigma_{r\beta}^2$
subplot	$(\alpha\beta)_{jk}$	r	0	0	1	$\sigma^2 + \sigma_{r\alpha\beta}^2 + \frac{r\Sigma\Sigma(\alpha\beta)_{jk}^2}{(a-1)(b-1)}$
	$(r\alpha\beta)_{ijk}$	1	0	0	1	$\sigma^2 + \sigma_{r\alpha\beta}^2$
	ϵ_{ijk}	1	1	1	1	σ^2 (not estimable)

Estimates and tests of fixed effects

- $\hat{\alpha}_j = \bar{y}_{.j.} - \bar{y}_{...}$ for $j = 1, 2, \ldots, a$
- $\hat{\beta}_k = \bar{y}_{..k} - \bar{y}_{...}$ for $k = 1, 2, \ldots, b$
- $(\hat{\alpha\beta})_{jk} = \bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...}$.
- Test $\alpha_j = 0$, $F_0 = \mathrm{MS_A}/\mathrm{MS_{rA}}$
- Test $\beta_k = 0$, $F_0 = \mathrm{MS_B}/\mathrm{MS_{rB}}$
- Test $(\alpha\beta)_{jk} = 0$, $F_0 = \mathrm{MS_{AB}}/\mathrm{MS_{rAB}}$.

SAS Code

```
data paper;
input block method temp resp@@;
datalines;
1 1 1 30 1 1 2 35 1 1 3 37 1 1 4 36
1 2 1 34 1 2 2 41 1 2 3 38 1 2 4 42
1 3 1 29 1 3 2 26 1 3 3 33 1 3 4 36
2 1 1 28 2 1 2 32 2 1 3 40 2 1 4 41
2 2 1 31 2 2 2 36 2 2 3 42 2 2 4 40
2 3 1 31 2 3 2 30 2 3 3 32 2 3 4 40
3 1 1 31 3 1 2 37 3 1 3 41 3 1 4 40
3 2 1 35 3 2 2 40 3 2 3 39 3 2 4 44
3 3 1 32 3 3 2 34 3 3 3 39 3 3 4 45
;
proc glm data=paper;
class block method temp;
```

```
model resp=block method block*method temp block*temp
 method*temp block*method*temp;
random block block*method block*temp block*method*temp;
test h=method e=block*method;
test h=temp e=block*temp;
test h=method*temp e=block*method*temp;
run;
```

SAS Output

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	35	822.9722222	23.5134921	.	.
Error	0	0.0000000	.		
CoTotal	35	822.9722222			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	2	77.5555556	38.7777778	.	.
method	2	128.3888889	64.1944444	.	.
block*method	4	36.2777778	9.0694444	.	.
temp	3	434.0833333	144.6944444	.	.
block*temp	6	20.6666667	3.4444444	.	.
method*temp	6	75.1666667	12.5277778	.	.
blo*meth*tmp	12	50.8333333	4.2361111	.	.

Tests Using the Type III MS for block*method as Error Ter

Source	DF	Type III SS	Mean Square	F Value	Pr >
method	2	128.3888889	64.1944444	7.08	0.0485

Tests Using the Type III MS for block*temp as Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr >
temp	3	434.0833333	144.6944444	42.01	0.0002

Tests Using the Type III MS for block*method*temp as E.Te

Source	DF	Type III SS	Mean Square	F Value	Pr
method*temp	6	75.16666667	12.52777778	2.96	0.05

Statistical Model II

- $$y_{ijk} = \mu + r_i + \alpha_j + (r\alpha)_{ij} + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk}$$
  - r_i: block effects (random) $\sim N(0, \sigma_r^2)$
  - α_j: whole-plot factor (A) main effects (fixed)
  - $(r\alpha)_{ij}$: whole plot error $\sim$ normal with $\sigma_{r\alpha}^2$
  - β_k: sub-plot factor (B) main effects (fixed)
  - $(\alpha\beta)_{jk}$: A and B interaction (fixed)
  - ϵ_{ijk}: sub-plot error $N(0, \sigma_\epsilon^2)$.
- Expected mean square

Term	$E(MS)$
r_i	$\sigma_\epsilon^2 + ab\sigma_r^2$
α_j (A)	$\sigma_\epsilon^2 + b\sigma_{r\alpha}^2 + \frac{rb\Sigma\alpha_j^2}{a-1}$
$(r\alpha)_{ij}$	$\sigma_\epsilon^2 + b\sigma_{r\alpha}^2$ (whole plot error)
β_k (B)	$\sigma_\epsilon^2 + \frac{ra\Sigma\alpha_j^2}{b-1}$
$(\alpha\beta)_{jk}$ (AB)	$\sigma_\epsilon^2 + \frac{r\Sigma\Sigma(\alpha\beta)_{jk}^2}{(a-1)(b-1)}$
ϵ_{ijk}	σ_ϵ^2 (subplot error)

General Split-Plot Designs

- Can have $>$ one whole-plot factor and $>$ one subplot factor with various blocking schemes.
- split-plot design consists of two superimposed blocked design

  Whole Plot

  - CRD, RCBD, Factorial D, BIBD, etc.

  Subplot

  - RCBD, BIBD, Factorial Design, etc.
- Analysis of Covariance
  - Covariate linear with response in subplot and whole plot

Other Variations

- Split-split-plot design

  1. randomization restriction can occur at any number of levels within the experiment

  2. two-level: split-split-plot design

- Strip-split-plot design (or Criss cross design, or Split-block design)

  Example: we want to compare the yield of a certain crop under different systems of soil preparation ($A : a_1, a_2, a_3, a_4$) and different density of seeding (B: b_1, b_2, b_3, b_4, b_5). Both operations (tilling and seeding) are done mechanically and it is impossible to perform both on small pieces of land. The arrangement shown below (strip-split-plot design) is then replicated r times, each time using different randomizations for A and B.

a_4b_1	a_4b_4	a_4b_2	a_4b_3	a_4b_5
a_1b_1	a_1b_4	a_1b_2	a_1b_3	a_1b_5
a_2b_1	a_2b_4	a_2b_2	a_2b_3	a_2b_5
a_3b_1	a_3b_4	a_3b_2	a_3b_3	a_3b_5

- For statistical models and analyses, refer to other books.