STAT 5200 Handout #16: Nested Design (Ch. 12)

Main effects and interactions make sense when factors A and B are <u>crossed</u> (i.e., when treatments are defined by the combination of A and B factor levels). But what if different levels of B are used for each level of A?

This is called nesting

Example: An experiment is carried out to evaluate three fertilizer sprays for trees. Twelve trees are used in the experiment, with four trees randomly assigned to each of the sprays. At the end of the experiment 6 leaves are randomly selected from each tree and the nitrogen content of each leaf evaluated once. Thus, there are a total of 3 × 4 ×6 = 72 measurements.

(Tree) (Spray)

We say that Factor A is nested in Factor B if the levels of Factor A change with the levels of Factor B. In this experiment Trees are nested in Sprays (and Leaves are nested in Trees). Factors that are not nested are said to be crossed. If Factor A is nested within Factor B we write A(B). Almost always a nested factor will be a random factor.

or 5,6,7,8 or 9,10,11,12

Source	DF	Levels											
Spray	3-1=2	1				2				3			
Tree(Spray)	3×(4-1)=9	1	2	3	4	1	2	3	4	1	2	3	4
Leaf(Tree×Spray)	3×4×(6-1)=60	1..6	1..6	1..6	1..6	1..6	1..6	1..6	1..6	1..6	1..6	1..6	1..6
	72-1=71												

7...12 13...18 . . .

M_1
|
$Spray_2$
|
$(Tree)_9$
|
$(Error)_{60}$

Some points here:

1. Nesting indicated in the Hasse Diagram using vertical lines.
2. Nested factors are almost always random effects.
3. The error term (E) in this case could be considered the Leaf(Tree×Spray) term.

DF for nested terms are calculated for each "parent" level and then summed.

What is measurement unit here?
Leaf

What is experimental unit here?
Tree

How are these reconciled here?

We have multiple exp. units (Trees) per treatment (Spray) [replicates]
and we have multiple msmnt. units (Leaves) per exp. unit (Tree). Two approaches:

equivalent { 1- ave of leaves within tree
2- account for nesting

Model and Notation:

or $B(A)_{j(i)}$

Tree

$$Y_{ijk} = \mu + A_i + B_{j(i)} + \varepsilon_{k(ij)}$$

Nitrogen content

Spray $i=1\ldots3$ $j=1\ldots4$ $k=1\ldots6$

- Tree is nested within Spray – different Trees for each Spray

Notation: Tree (Spray) or $j(i)$

- Note that error is also nested – and always has been in previous examples

So sometimes $\varepsilon_{k(ij)}$ or ε_{ijk}

As before, use Hasse Diagram and EMS calculation (next page) to build test statistics; Example: test Spray effect:

F stat = MS_{Term} / $MS_{(\text{next random term below in Hasse})}$

$$F = \frac{MS_{Spray}}{MS_{Tree}} \sim F_{2,9}$$

Note that with nesting, A×B interaction is meaningless:

Tree & Spray not crossed (every level of Spray doesn't "see" every level of Tree)

Does the effect of Spray depend on Tree? (interaction)

→ We can't answer this because each Tree only received one Spray

With nesting, the averaging over measurement units (within exp. unit) is done automatically (see pp. 6-7 of this Handout #16)

& gives equivalent results to accounting for nesting.

Ignoring nesting can lead to incorrect claims of significance (see p. 8 of this Handout #16, where Leaf is treated as experimental unit) because:

1. true variance is underestimated i.e., MSE is a biased estimate of σ^2

2. sample size is inflated

↳ should be total # of exp. units

What if we:

- Averaged over measurement units (as discussed at beginning of semester)?

get same result

- Ignored the nesting completely, and treated measurement units as experimental units?

$$F_A^{(right)} = \frac{MS_A}{MS_{B(A)}} = \frac{SS_A/2}{SS_{B(A)}/9}$$

$$F_A^{(wrong)} = \frac{MS_A}{MS_E^{(wrong)}} = \frac{SS_A/2}{(SS_E + SS_{B(A)})/69}$$

```
/* STAT 5200
   fully nested model
   spray data
 */


/* Read in data */
data sprays;
     input Spray Tree Leaf Nitrogen @@;
     cards;
  1 1 1  4.50 1 1 2  7.04 1 1 3  4.98 1 1 4  5.48 1 1 5  6.54
  1 2 1  5.78 1 2 2  7.69 1 2 3  9.68 1 2 4  5.89 1 2 5  4.07
  1 3 1 13.32 1 3 2 15.05 1 3 3 12.67 1 3 4 12.42 1 3 5 10.03
  1 4 1 11.59 1 4 2  8.96 1 4 3 10.95 1 4 4  9.87 1 4 5 10.48
  2 1 1 15.32 2 1 2 14.97 2 1 3 14.81 2 1 4 14.26 2 1 5 15.88
  2 2 1 14.53 2 2 2 14.51 2 2 3 12.61 2 2 4 16.13 2 2 5 13.65
  2 3 1 10.89 2 3 2 10.27 2 3 3 12.21 2 3 4 12.77 2 3 5 10.45
  2 4 1 15.12 2 4 2 13.79 2 4 3 15.32 2 4 4 11.95 2 4 5 12.56
  3 1 1  7.18 3 1 2  7.98 3 1 3  5.51 3 1 4  7.48 3 1 5  7.55
  3 2 1  6.70 3 2 2  8.28 3 2 3  6.99 3 2 4  6.40 3 2 5  4.96
  3 3 1  5.94 3 3 2  5.78 3 3 3  7.59 3 3 4  7.21 3 3 5  6.12
  3 4 1  4.08 3 4 2  5.46 3 4 3  5.40 3 4 4  6.85 3 4 5  7.74
  1 1 6  7.20 1 2 6  4.08 1 3 6 13.50 1 4 6 12.79 2 1 6 16.01
  2 2 6 14.78 2 3 6 11.44 2 4 6 15.31 3 1 6  5.64 3 2 6  7.03
  3 3 6  7.13 3 4 6  6.81
;
run;

/* Get EMS */
proc mixed data=sprays method=type3;
     class Spray Tree Leaf;
     model Nitrogen = Spray;
     random Tree(Spray);
     title1 'Get Expected Mean Squares using TYPE3 option';
run;
```

Type 3 Analysis of Variance

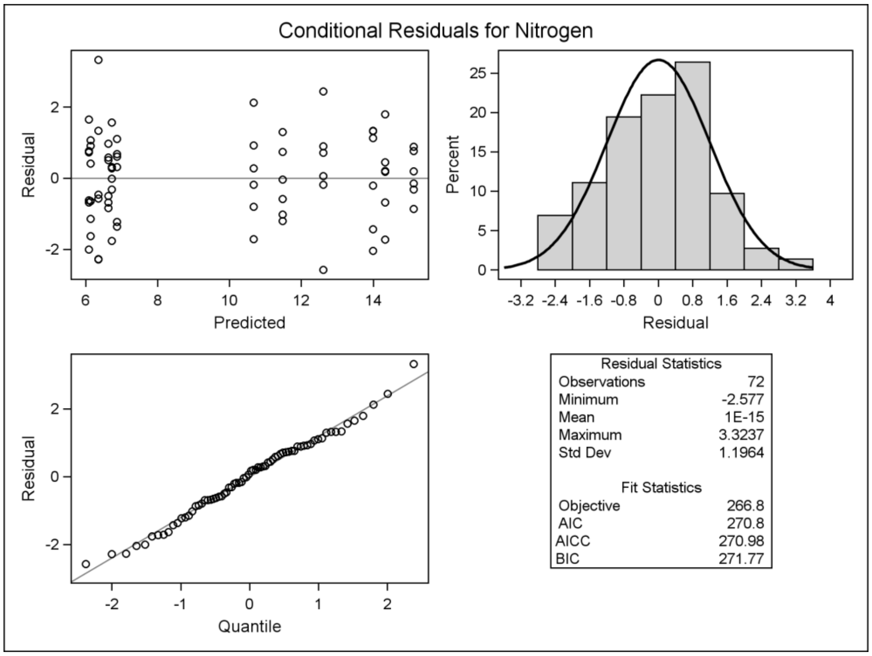
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Spray	2	637.9466	318.9733	Var(Residual) + 6Var(Tree(Spray)) + Q(Spray)	MS(Tree(Spray))	9	10.95	0.0039
Tree(Spray)	9	262.2444	29.1382	Var(Residual) + 6 Var(Tree(Spray))	MS(Residual)	60	17.35	<.0001
Residual	60	100.7525	1.6792	Var(Residual)	.	.	.	.

ratio of EMS's that $=1$ when $H_0: Q_A = 0$

$$\frac{Q_A + 6\sigma^2_{T(S)} + \sigma^2}{6\sigma^2_{T(S)} + \sigma^2} = \frac{EMS_{Spray}}{EMS_{Tree(Spray)}} = 1 \text{ when } H_0 \text{ true}$$

```
/* Get tests */
proc mixed data=sprays covtest plots=residualpanel;
     class Spray Tree Leaf;
     model Nitrogen = Spray / ddfm=satterthwaite;
     random Tree(Spray);
     lsmeans Spray / pdiff=all adjust=Tukey;
     title1 'Now using REML estimation';
run;
```

Now using REML estimation



Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
Tree(Spray)	4.5765	2.2899	2.00	0.0228
Residual	1.6792	0.3066	5.48	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Spray	2	9	10.95	0.0039

Least Squares Means						
Effect	**Spray**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Spray	**1**	8.9400	1.1019	9	8.11	<.0001
Spray	**2**	13.7308	1.1019	9	12.46	<.0001
Spray	**3**	6.5754	1.1019	9	5.97	0.0002

← H_0: $Spray_1 = 0$
$A_1 = 0$

Differences of Least Squares Means *					
Effect	**Spray**	**_Spray**	**Estimate**	**Adjustment**	**Adj P**
Spray	**1**	**2**	-4.7908	Tukey	0.0322
Spray	**1**	**3**	2.3646	Tukey	0.3283
Spray	**2**	**3**	7.1554	Tukey	0.0034

"minus"

← post-hoc mean comparisons (all pairwise comp.)

Tukey groupings, by hand:		
Spray	LSMEAN	Grouping
2	13.7308	A
1	8.9400	B
3	6.5754	B

```
/* What if we just averaged over Leafs
   (within Tree within Spray)? */
proc sort data=sprays;
  by Tree Spray;
proc means data=sprays mean noprint;
  by Tree Spray;
  var Nitrogen;
  output out=out2 mean=Nbar;
proc print data=out2;
   title1 'Averaged Nitrogen data';
   var Tree Spray Nbar;
run;
```

Averaged Nitrogen data

Obs	Tree	Spray	Nbar
1	1	1	5.9567
2	1	2	15.2083
3	1	3	6.8900
…			
12	4	3	6.0567

```
proc glm data=out2;
  class Spray;
  model Nbar = Spray;
  lsmeans Spray / pdiff adjust=tukey;
  title1 'Using Averaged Nitrogen Data';
run;
```

Using Averaged Nitrogen Data

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Spray	2	106.3244347	53.1622174	10.95	0.0039

Least Squares Means for effect Spray Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: Nbar			
i/j	**1**	**2**	**3**
1		0.0322	0.3283
2	0.0322		0.0034
3	0.3283	0.0034	

```
/* What if we completely ignored the nested structure here? */
proc glm data=sprays;
  class Spray;
  model Nitrogen = Spray;
  lsmeans Spray / pdiff adjust=tukey;
  title1 'Ignoring nested structure';
run;
```

Ignoring nested structure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	637.946608	318.973304	60.63	<.0001
Error	69	362.996979	5.260826		
Corrected Total	71	1000.943588			

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Spray	Nitrogen LSMEAN	LSMEAN Number
1	8.9400000	1
2	13.7308333	2
3	6.5754167	3

Least Squares Means for effect Spray
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: Nitrogen

i/j	1	2	3
1		<.0001	0.0019
2	<.0001		<.0001
3	0.0019	<.0001	

Smaller P-value due to larger test statistic (smaller variance of differences results from artificially inflated sample size).