



# Chapter 28

## Nested Designs, Subsampling, and Partially Nested Designs

We look at another design, the *nested design* and the closely related *subsampling* and *partially nested* designs.

### 28.1 Distinction between Nested and Crossed Factors

*Nested* designs arise when every level of one factor (B, say) is combined with (“nested inside”) only one level of another factor (A, say) in the design; this compares to the previously discussed *crossed* design (or complete factorial designs) where *all* of the levels of factor A are combined with *all* of the levels of factor B.

#### Exercise 28.1 (Distinction between Nested and Crossed Factors)

1. *Different arrangements of same data*

The following two arrangements of test scores for course/instructor data sets,

	factor A, course →	1	2	3
factor B,	1	69, 75, 46	88, 90, 91	51, 65, 57
instructor	2	71, 80, 77	92, 85, 95	49, 71, 63

course, $i$ :	1	2	3
instructors, $j$ :	1 2	1 2	1 2
tests, $k = 1$ :	69 71	88 92	51 49
$k = 2$ :	75 80	90 85	65 71
$k = 3$ :	46 77	91 95	57 63

are (choose one) **the same** / **different** data sets.

2. *Two-factor crossed design (complete factorial design).*

Consider the following test scores for course/instructor set of data.

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	1	2	1	2
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

The instructors are (choose one) **nested inside** / **crossed with** courses because the same two instructors (1 and 2) are used in all three courses.

3. *Two-factor nested design.*

Consider the following test scores for course/instructor set of data.

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	3	4	5	6
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

The instructors are (choose one) **the same** / **different** for different courses. That is, the instructors are *nested* inside courses.

4. *Another way to look at crossed versus nested*

**True / False**

Whereas the complete factorial crossed design is,

	factor A, course $\rightarrow$	1	2	3
factor B,	1	69, 75, 46	88, 90, 91	51, 65, 57
instructor	2	71, 80, 77	92, 85, 95	49, 71, 63

the *incomplete* factorial nested design is,

	factor A, course $\rightarrow$	1	2	3
factor B,	1	69, 75, 46		
instructor	2	71, 80, 77		
	3		88, 90, 91	
	4		92, 85, 95	
	5			51, 65, 57
	6			49, 71, 63

5. *A word about notation.*

**True / False** Consider the following set of data.

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	1	2	1	2
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

Even though the instructors are designated as “1” and “2” for each course which seems to indicate that this is crossed data, it is sometimes the case that, in fact, *different* instructor have been used for all three courses and so this would then be an example of two-factor *nested* data.

6. *Same data, different designs.*

**True / False**

The data sets for crossed and nested designs are arranged in the same way but treated differently and so, in particular, inference for the two designs is different.

7. *Nested designs are incomplete factorial designs.*

**True / False**

The nested design is an *incomplete* factorial design because *all* of the levels of one factor are *not* combined with *all* of the levels of any other factor.

8. *Two-factor nested design*

Interpreting the following test scores for course/instructor set of data,

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	1	2	1	2
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

as a two-factor nested design, the two factors are (choose *two!*)

**course / instructor / test<sup>1</sup>**

## 28.2 Two-Factor Nested Designs

The nested two (fixed) factor design model is

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk},$$

---

<sup>1</sup>The three tests in each treatment in this two factor nested design are considered to be all *different*; indeed, all eighteen of the tests are considered to be different from one another.

where, notice, the nested two factor design does not (cannot) have any interaction terms. The  $\beta_{j(i)}$  is called the *specific effect* of the  $j$ th level of factor B nested within the  $i$ th level of factor A<sup>2</sup>. Either  $\alpha_i$  or  $\beta_{j(i)}$  or both can be random factors.

**Exercise 28.2 (Two–Factor Nested Designs: test scores)**

Consider the following test scores for course/instructor set of data,

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	1	2	1	2
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

The nested two (fixed) factor design model is

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$$

Match the model with the test scores data.

model	test scores data
(a) $Y_{ijk}$	(a) error
(b) $\mu_{..}$	(b) instructor (nested inside course) effect
(c) $\alpha_i$	(c) course effect
(d) $\beta_{j(i)}$	(d) (grand) average of all test scores
(e) $\varepsilon_{ijk}$	(e) individual test scores
	(f) course (nested inside instructor) effect

model	(a)	(b)	(c)	(d)	(e)
test scores					

### 28.3 Analysis of Variance for Two–Factor Nested Designs

SAS program: att10-28-3-tests-nested-inference

The following ANOVA table is used to make inferences on this model,

Source	$df$	$SS$	$MS$	$E\{MS\}$ both fixed	$E\{MS\}$ A fixed, B(A) random	$E\{MS\}$ both random
Factor A	$a - 1$	$SSA$	$MSA$	$\sigma^2 + bn \frac{\sum \alpha_i^2}{a-1}$	$\sigma^2 + bn \frac{\sum \alpha_i^2}{a-1} + n\sigma_\beta^2$	$\sigma^2 + bn\sigma_\alpha^2 + n\sigma_\beta^2$
Factor B (within A)	$a(b - 1)$	$SSB(A)$	$MSB(A)$	$\sigma^2 + n \frac{\sum \sum \beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
Error	$ab(n - 1)$	$SSE$	$MSE$	$\sigma^2$	$\sigma^2$	$\sigma^2$
Total	$abn - 1$	$SSTO$				

<sup>2</sup>Notice that the notation is the “reverse” of how to interpret a function,  $f(x)$ , say, where  $x$  is considered to be “inside” the function  $f$ .

and so the appropriate test statistics are

Test for	A, B(A) fixed	A fixed, B(A) random	A, B(A) random
Factor A	$MSA/MSE$	$MSA/MSB(A)$	$MSA/MSB(A)$
Factor B (within A)	$MSB(A)/MSE$	$MSB(A)/MSE$	$MS(A)/MSE$

**Exercise 28.3 (ANOVA for Two-Factor Nested Designs: test scores)**

Consider the student test scores for a number of different courses and instructors.

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	1	2	1	2
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

Test both factors at  $\alpha = 0.05$ .

1. ANOVA Table

**True / False** The ANOVA table is given by

Source	$df$	$SS$	$MS$
Course, A	2	2955.44	1477.72
Instructors, within Courses, B(A)	3	258.83	86.28
Error	12	914.67	76.22
Total	17	4128.94	

2. Test courses, both course/instructor fixed

(a) Statement

The statement of the test is (check none, one or more):

- i.  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$  versus  $H_a : \text{at least one } \alpha_i \neq 0, i = 1, 2, 3.$
- ii.  $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$  versus  $H_a : \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2.$
- iii.  $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$  versus  $H_a : \text{at least one } \tau_i \neq 0, i = 1, 2, 3, 4.$
- iv.  $H_0 : \mu_{1.} = \mu_{2.} = \mu_{3.} = \mu_{4.}$  versus  $H_a : \text{at least one } \mu_i \neq \mu_{j.}, i, j = 1, 2, 3, 4.$

(b) Test.

Since the test statistic is

$$F = \frac{MSA}{MSE} = \frac{1477.72}{76.22} = 19.39,$$

the p-value, with (2, 12) degrees of freedom, is given by

$$\text{p-value} = P(F \geq 19.39)$$

which equals (circle one) **0.00** / **0.34** / **0.43**.

The level of significance is 0.05.

(c) *Conclusion*

Since the p-value, 0.0, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the average test scores for the three different courses are the same.

3. *Test instructors (nested in courses), both course/instructor fixed*(a) *Statement*

The statement of the test is (check none, one or more):

- i.  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$  versus  $H_a : \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3$ .
- ii.  $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$  versus  $H_a : \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2$ .
- iii.  $H_0 : \text{all } \beta_{j(i)} = 0$  versus  
 $H_a : \text{at least one } \beta_{j(i)} \neq 0, i = 1, 2, 3; j = 1, 2$ .
- iv.  $H_0 : \mu_{11} = \mu_{12}; \mu_{21} = \mu_{22}; \mu_{31} = \mu_{32}$  versus  
 $H_a : \text{at least one } \mu_{ij} \neq \mu_{ij'}, i = 1, 2, 3; j = 1, 2$ .

(b) *Test*

Since the test statistic is

$$F = \frac{MSB(A)}{MSE} = \frac{86.28}{76.22} = 1.13,$$

the p-value, with (3,12) degrees of freedom, is given by

$$\text{p-value} = P(F \geq 1.13)$$

which equals (circle one) **0.00** / **0.38** / **0.43**.

The level of significance is 0.05.

(c) *Conclusion*

Since the p-value, 0.38, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the average scores for different instructors in different courses are the same.

4. *ANOVA table, broken into three courses*

**True / False** The ANOVA table, where the course sums of squares has been broken down into the sums of squares associated with the three individual courses, is given by

Source	<i>df</i>	<i>SS</i>	<i>MS</i>
Course, A	2	2955.44	1477.72
Instructors, within Courses, B(A)	3	258.83	86.28
Instructors, within Course 1, B(1)	1	240.67	240.67
Instructors, within Course 2, B(2)	1	1.5	1.5
Instructors, within Course 3, B(3)	1	16.67	16.67
Error	12	914.67	76.22
Total	17	4128.94	

Section 3. Analysis of Variance for Two-Factor Nested Designs (ATTENDANCE 10)235

5. Test instructors for each course separately, both course/instructor fixed

Instructors using (within) course 1,

$H_0 : \beta_{j(1)} = 0$  versus

$H_a : \text{at least one } \beta_{j(1)} \neq 0, j = 1, 2.$

Since

$$F = \frac{MSB(1)}{MSE} = \frac{240.67}{76.22} = 3.16$$

and p-value  $P(F > 3.16; 1, 12) =$

(choose one) **0.00** / **0.10** / **0.43**

and  $\alpha = 0.05$

since p-value = 0.10 >  $\alpha = 0.05$

accept null; that is, instructor effects within course 1 are not significant  
(mean test scores same for different instructors in course 1)

Instructors using (within) course 2,

$H_0 : \beta_{j(2)} = 0$  versus

$H_a : \text{at least one } \beta_{j(2)} \neq 0, j = 1, 2.$

since p-value  $P(F > \frac{1.5}{76.22}; 1, 412) = 0.89 > \alpha = 0.05$

accept null; that is, instructor effects within course 2 are not significant  
(mean test scores same for different instructors in course 2)

Instructors using (within) course 3,

$H_0 : \beta_{j(3)} = 0$  versus

$H_a : \text{at least one } \beta_{j(3)} \neq 0, j = 1, 2.$

since p-value  $P(F > \frac{16.67}{76.22}; 1, 12) = 0.65 > \alpha = 0.05$

accept null; that is, instructor effects within course 3 are not significant  
(mean test scores same for different instructors in course 2)

6. Test courses, both course/instructor random

(a) Statement

The statement of the test is (check none, one or more):

i.  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$  versus  $H_a : \alpha_1 \neq \alpha_2, \alpha_1 = \alpha_3.$

ii.  $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$  versus  $H_a : \alpha_1 \neq \alpha_3, \alpha_1 \neq \alpha_2.$

iii.  $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$  versus  
 $H_a : \text{at least one } \tau_i \neq 0, i = 1, 2, 3, 4.$

iv.  $H_0 : \sigma_\alpha^2 = 0$  versus  
 $H_a : \sigma_\alpha^2 > 0.$

(b) Test

Since the test statistic<sup>3</sup> is

$$F = \frac{MSA}{MSB(A)} = \frac{1477.72}{86.28} = 17.13$$

the p-value, with (2,3) degrees of freedom, is given by

$$\text{p-value} = P(F \geq 17.13)$$

which equals (circle one) **0.02** / **0.38** / **0.43**.

The level of significance is 0.05.

(c) *Conclusion*

Since the p-value, 0.02, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the average scores for different courses are the same.

7. Variance,  $\sigma_\beta^2$ ,  $\sigma_\alpha^2$ , both course/instructor random

$$s_\alpha^2 = \frac{MSA - MSB(A)}{bn} \approx \frac{147.72 - 86.28}{2(3)} = (\text{choose one}) \mathbf{9.23} / \mathbf{10.24} / \mathbf{11.43}$$

$$\text{and } s_\beta^2 = \frac{MSB(A) - MSE}{n} \approx \frac{86.28 - 76.22}{3} = (\text{choose one}) \mathbf{3.35} / \mathbf{6.24} / \mathbf{8.43}$$

Both of these formulas are derived<sup>4</sup> from the formulas given in the table above.

8. 95% confidence interval for  $\sigma_\beta^2$ , using MLS method.

Since

$$c_1 = 1/n = 1/3 = 0.333$$

$$c_2 = -1/n = -1/3 = -0.333$$

$$df_1 = 3$$

$$df_2 = 12$$

$$MS_1 = MSB(A) = 86.28$$

$$MS_2 = MSE = 76.22$$

<sup>3</sup>Look at the table above which gives different  $F$  statistics for different fixed/random models.

<sup>4</sup>For example, to derive the formula for  $s_\alpha^2$ , notice

$$E(MASA) = \sigma^2 + bn\sigma_\alpha^2 + n\sigma_\beta^2$$

$$E(MASB(A)) = \sigma^2 + n\sigma_\beta^2$$

and so

$$\frac{E(MASA) - E(MASB(A))}{bn} = \sigma_\alpha^2$$

and

$$\begin{aligned}
 F_1(1 - \alpha/2; df_1, \infty) &= F(1 - 0.05/2; 3, \infty) = 3.12 \\
 F_2(1 - \alpha/2; df_2, \infty) &= F(1 - 0.05/2; 12, \infty) = 1.94 \\
 F_3(1 - \alpha/2; \infty, df_1) &= F(1 - 0.05/2; \infty, 3) = 13.90 \\
 F_4(1 - \alpha/2; \infty, df_2) &= F(1 - 0.05/2; \infty, 12) = 2.73 \\
 F_5(1 - \alpha/2; df_1, df_1) &= F(1 - 0.05/2; 3, 12) = 4.47 \\
 F_6(1 - \alpha/2; df_2, df_1) &= F(1 - 0.05/2; 12, 3) = 14.34
 \end{aligned}$$

and

$$\begin{aligned}
 G_1 &= 1 - 1/F_1 = 1 - 1/3.12 = 0.67949 \\
 G_2 &= 1 - 1/F_2 = 1 - 1/1.94 = 0.48454 \\
 G_3 &= \frac{(F_5 - 1)^2 - (G_1 F_5)^2 - (F_4 - 1)^2}{F_5} = -0.03967 \\
 G_4 &= F_6 \left( \left( \frac{F_6 - 1}{F_6} \right)^2 - \left( \frac{F_3 - 1}{F_6} \right)^2 - G_2^2 \right) = -2.562
 \end{aligned}$$

and

$$\begin{aligned}
 H_L &= \sqrt{(G_1 c_1 MS_1)^2 + ((F_4 - 1) c_2 MS_2)^2 - G_3 c_1 c_2 MS_1 MS_2} = 339.21 \\
 H_U &= \sqrt{((F_3 - 1) c_1 MS_1)^2 + (G_2 c_2 MS_2)^2 - G_4 c_1 c_2 MS_1 MS_2} = 6357.07
 \end{aligned}$$

and since  $s_\beta^2 = 3.35$ , the approximate confidence interval for  $\sigma_\beta^2$  is<sup>5</sup>

$$3.35 - 339.21 \leq \sigma_\beta^2 \leq 3.35 + 6357.07 =$$

(choose one) **(0, 6360.42)** / **(-335.86, 6360.42)** / **(0, 3.35)**

## 28.4 Evaluation of Appropriateness of Nested Design Model

SAS program: att10-28-4-tests-nested-residuals

### Exercise 28.4 (Evaluation of Appropriateness of Nested Design Model)

Consider the student test scores for a number of different courses and instructors.

<sup>5</sup>Remember, the variance can *never* be negative and so if a negative is calculated, it must be set to zero—remember, this confidence interval is a *approximation* after all.

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	1	2	1	2
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

1. *Residuals*

The residual  $e_{ijk} = e_{211} =$

(circle one) **-1.6667** / **-1.38** / **0.43**.

2. *Residual plot versus predicted*

The residual plot indicates

(choose one) **constant** / **nonconstant** variance.

3. *Vertical bar plot of residuals*

The vertical bar plot of residuals indicates

(choose one) **constant** / **nonconstant** variance for the different courses because the range of residuals is *different* for the different courses; (-15 to 9 for course 1; -3 to 3 for course 2; -9 to 9 for course 3)

4. *Normal probability plot of residuals.*

The normal probability plot indicates

(choose one) **normal** / **non-normal** residuals.

## 28.5 Analysis of Factor Effects in Two-Factor Nested Designs

SAS program: att10-28-5-tests-nested-effects

Detailed analysis of treatment effects, using, for example, Tukey, Scheffe and Bonferroni, is possible, just like it was for previous designs.

### Exercise 28.5 (Analysis of Factor Effects in Two-Factor Nested Designs)

Consider the student test scores for a number of different courses and instructors.

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	1	2	1	2
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

1. *Pairwise 95% confidence interval, courses,  $L = \mu_1. - \mu_2.$ ,  
Tukey, both A, B fixed.*

Assume this is one of three  $g = 3$  pairwise comparisons and  $\alpha = 0.05$

From the SAS output,

$$\bar{Y}_{1..} = 69.66667$$

$$\bar{Y}_{2..} = 90.16667$$

$$\hat{L} = \bar{Y}_{1..} - \bar{Y}_{2..} = (\text{choose one}) \mathbf{-2.33} / \mathbf{-13.45} / \mathbf{-20.5}$$

$$s\{\hat{L}\} = \sqrt{MSE \left(\frac{1}{n} + \frac{1}{n}\right)} = \sqrt{76.22 \left(\frac{1}{6} + \frac{1}{6}\right)} \approx 5.04$$

( $n = 6$  number of instructors in each course)

$$T = \frac{1}{\sqrt{2}}q(1 - \alpha; df) = \frac{1}{\sqrt{2}}q(1 - 0.05; 3, 12) = \frac{1}{\sqrt{2}}q(0.95; 3, 12) = \frac{3.77}{\sqrt{2}} \approx 2.67$$

(table B.9 and  $df = (3, 12)$  because 3 contrasts and 12 error  $df$ )

and so the CI is  $-20.5 \pm 2.67(5.04) =$

(choose one)  $(\mathbf{-30.43}, \mathbf{-10.57}) / (\mathbf{-33.96}, \mathbf{-7.04}) / (\mathbf{-21.83}, \mathbf{19.83})$

2. *Pairwise 95% confidence interval, courses,  $L = \mu_1. - \mu_2.$ ,  
Tukey, mixed: A fixed, B random.*

Assume this is one of three  $g = 3$  pairwise comparisons and  $\alpha = 0.05$

$$\bar{Y}_{1..} = 69.66667$$

$$\bar{Y}_{2..} = 90.16667$$

$$\hat{L} = \bar{Y}_{1..} - \bar{Y}_{2..} = (\text{choose one}) \mathbf{-2.33} / \mathbf{-13.45} / \mathbf{-20.5}$$

$$s\{\hat{L}\} = \sqrt{MSB(A) \left(\frac{1}{n} + \frac{1}{n}\right)} = \sqrt{86.28 \left(\frac{1}{6} + \frac{1}{6}\right)} \approx 5.36$$

(notice  $s\{\hat{L}\}$  is different in mixed and fixed cases)

$$T = \frac{1}{\sqrt{2}}q(1 - \alpha; df) = \frac{1}{\sqrt{2}}q(1 - 0.05; 3, 12) = \frac{1}{\sqrt{2}}q(0.95; 3, 12) = \frac{3.77}{\sqrt{2}} \approx 2.67$$

and so the CI is  $-20.5 \pm 2.67(5.36) =$

(choose one)  $(\mathbf{-30.43}, \mathbf{-10.57}) / (\mathbf{-15.12}, \mathbf{32.78}) / (\mathbf{-34.81}, \mathbf{-6.19})$

3. *95% confidence interval for overall mean<sup>6</sup>  
both A, B random*

From the SAS output,

$$\bar{Y}_{...} = 73.06$$

$$s\{\bar{Y}_{...}\} = \sqrt{\frac{MSA}{abn}} = \sqrt{\frac{1477.72}{3(2)(3)}} \approx 82.1$$

$$t(1 - \frac{\alpha}{2}; a - 1) = t(1 - \frac{0.05}{2}; 2) = t(0.975; 2) \approx 4.303$$

and so the CI is  $73.06 \pm 4.303(82.1) =$

(choose one)  $(\mathbf{-20.83}, \mathbf{18.83}) / (\mathbf{-15.12}, \mathbf{32.78}) / (\mathbf{-280.22}, \mathbf{426.34})$

## 28.6 Unbalanced Nested Two-Factor Designs

SAS program: att10-28-6-tests-nested-regression

<sup>6</sup>The  $MSA$  in  $s\{\bar{Y}_{...}\} = \sqrt{\frac{MSA}{abn}}$  becomes  $MSB(A)$  if A is fixed and B is random.

The regression form of the nested model is used to deal with unbalanced nested two-factor data sets.

### Exercise 28.6 (Unbalanced Nested Two-Factor Designs: Test scores)

Consider the student test scores for a number of different courses and instructors.

course, $i$ :	1		2		3		
instructors, $j$ :	1	2	1	2	1	2	3
tests, $k = 1$ :	69	71	88	92	51	49	43
$k = 2$ :	75	80	90		65	71	78
$k = 3$ :	46		91		57	63	

Notice that there are unequal<sup>7</sup> nestings and *missing* observations in this data set.

#### 1. Regression Model

**True / False** The regression model is given by

$$Y_{ijk} = \mu_{..} + \alpha_1 X_{ij1} + \alpha_2 X_{ij2} + \beta_{1(1)} X_{ij3} + \beta_{1(2)} X_{ij4} + \beta_{1(3)} X_{ij5} + \beta_{2(3)} X_{ij6} + \varepsilon_{ijk}$$

where

$$X_{ij1} = \begin{cases} 1, & \text{if test from course 1} \\ -1, & \text{if test from course 3} \\ 0, & \text{otherwise,} \end{cases}$$

$$X_{ij2} = \begin{cases} 1, & \text{if test from course 2} \\ -1, & \text{if test from course 3} \\ 0, & \text{otherwise,} \end{cases}$$

$$X_{ij3} = \begin{cases} 1, & \text{if test from instructor 1 from course 1} \\ -1, & \text{if test from instructor 2 from course 1} \\ 0, & \text{otherwise,} \end{cases}$$

$$X_{ij4} = \begin{cases} 1, & \text{if test from instructor 1 from course 2} \\ -1, & \text{if test from instructor 2 from course 2} \\ 0, & \text{otherwise,} \end{cases}$$

$$X_{ij5} = \begin{cases} 1, & \text{if test from instructor 1 from course 3} \\ -1, & \text{if test from instructor 3 from course 3} \\ 0, & \text{otherwise.} \end{cases}$$

$$X_{ij6} = \begin{cases} 1, & \text{if test from instructor 2 from course 3} \\ -1, & \text{if test from instructor 3 from course 3} \\ 0, & \text{otherwise.} \end{cases}$$

---

<sup>7</sup>In particular, the third course now has three (3) instructors instead of the previously specified two (2) instructors.

2. *Fitted Models*

Using SAS, match the various models with their estimations,

(a) Model A

$$\hat{Y} = 73.32 - 3.91X_1 + 17.51X_2 - 6.08X_3 - 1.17X_4 - 2.06X_5 + 1.28X_6$$

(b) Model B

$$\hat{Y} = 68.74 - 5.95X_3 + 9.88X_4 - 3.34X_5 - 0.01X_6$$

(c) Model C

$$\hat{Y} = 72.69 - 4.49X_1 + 17.56X_2$$

Estimated Models	Fitted Models
(a) Model A	(a) full model
(b) Model B	(b) reduced model, to test courses, $\alpha_i = 0$
(c) Model C	(c) reduced model, to test instructors, $\beta_j(i) = 0$

Estimated Models	(a)	(b)	(c)
Fitted Models			

3. *Test of course effect*

Using SAS,

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0 \text{ versus}$$

$$H_a : \text{at least one } \alpha_i \neq 0, i = 1, 2, 3.$$

$$\frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} = \frac{3566.87 - 1473}{12 - 10} \div \frac{1473}{10} = 7.11$$

and so p-value is  $P(F > 7.11; 2, 10) \approx$  (choose one) **0.012** / **0.70** / **0.75**

since p-value  $< \alpha = 0.05$

reject null; that is, average scores for different courses are different

4. *Test of instructor (nested in course) effect*

Using the SAS output,

$$H_0 : \beta_{1(1)} = \beta_{1(2)} = \beta_{1(3)} = \beta_{2(3)} = 0 \text{ versus}$$

$$H_a : \text{at least one } \beta_{j(i)} \neq 0.$$

$$\frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} = \frac{1673.43 - 1473}{14 - 10} \div \frac{1473}{10} = 0.34$$

and so p-value is  $P(F > 0.34; 4, 10) \approx$  (choose one) **0.34** / **0.62** / **0.85**

since p-value  $> \alpha = 0.05$

accept null; that is, average scores for different instructors are the same

## 28.7 Subsampling in Single-Factor Study with Completely Randomized Design

SAS program: att10-28-7-tests-subsample-inference

*Subsampling* in single-factor studies occur when more than one (set of) observation(s) is made on an experimental unit. The model for subsampling in *single*-factor studies is found to be exactly equivalent to the model for nested *two*-factor designs.

### Exercise 28.7 (Subsampling in Single-Factor Study with CRD)

Consider the following test scores for course/instructor set of data.

course, $i$ :	1		2		3	
instructors, $j$ :	1f	1s	1f	1s	1f	1s
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

Assume the *same* instructor is used in both in the spring and the fall for all of the three courses; for example “1f” means instructor 1 who taught in the *fall* and “1s” means (the same) instructor 1 who taught in the *spring*.

1. *Single-factor subsampling design.*

Instead of a two-factor (nested or crossed) study, we now have a *one*-factor study where the one factor is (choose one) **course** / **instructor** / **test** and where we have “subsamped” from the scores of the tests of instructor 1 *twice* (once in the fall and once in the spring) for each of the three courses.

2. *A word about notation.*

**True / False** Consider the following set of data.

course, $i$ :	1		2		3	
instructors, $j$ :	1	2	1	2	1	2
tests, $k = 1$ :	69	71	88	92	51	49
$k = 2$ :	75	80	90	85	65	71
$k = 3$ :	46	77	91	95	57	63

Even though the instructors are designated as “1” and “2” for each course which seems to indicate that this is crossed data, it is sometimes the case that, in fact, the *same* instructor has been used for all three courses and so this would then be an example of single-factor *subsampled* data.

3. *Same data, different designs.*

**True / False**

Even though the data sets for two-factor mixed (A fixed, B random) nested and one-factor subsampled designs are arranged in the same way but treated (interpreted) differently, they are analyzed in exactly the *same* way!

4. *One-factor subsampled designs are complete factorial designs.*

**True / False**

The subsampled design is a *complete* factorial design because of the (abundance) of observations for each level of the one factor in the study.

5. *One-factor subsampling model*

The one-factor subsampling design model<sup>8</sup> is

$$Y_{ijk} = \mu_{..} + \tau_i + \varepsilon_{j(i)} + \eta_{ijk}$$

Match the model with the test scores data.

model	test scores data
(a) $Y_{ijk}$	(a) observational error
(b) $\mu_{..}$	(b) experimental (instructor) error
(c) $\tau_i$	(c) course effect
(d) $\varepsilon_{j(i)}$	(d) (grand) average of all test scores
(e) $\eta_{ijk}$	(e) observed test scores
	(f) instructor (nested in course) effect

model	(a)	(b)	(c)	(d)	(e)
test scores					

6. *ANOVA table for subsampling model*

**True / False** The ANOVA table for the subsampled data is given by

Source	df	SS	MS
Course, A	2	2955.44	1477.72
Experimental error (instructors)	3	258.83	86.28
Observational Error	12	914.67	76.22
Total	17	4128.94	

and is identical to the previous ANOVA table for the data treated as nested data except we now have two types of error<sup>9</sup>

<sup>8</sup>Recall that the nested design is given by

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$$

<sup>9</sup>Notice that SAS does not indicate this, but uses the nested design designations—at some point, I will figure out how to get SAS to label the ANOVA table properly!

- experimental error,  
which, for the nested design, was the nested factor  $B(A)$ ,
- observational error,  
to distinguish it from the experimental error.

7. *Test course*
 $H_0 : \tau_i = 0$  versus

 $H_a : \text{at least one } \tau_i \neq 0, i = 1, 2, 3.$ 

 since p-value  $P(F > \frac{MSA}{MSEE} = \frac{1477.72}{86.28} = 17.13; 2, 3) = 0.02 < \alpha = 0.05$ 

 reject null; that is, the course effect (choose one) **is / is not** significant  
(mean tests scores different for different courses)
8. *Test instructors*
 $H_0 : \sigma^2 = 0$  versus

 $H_a : \sigma^2 > 0.$ 

 since p-value  $P(F > \frac{MSEE}{MSOE} = \frac{86.28}{76.22} = 1.13; 3, 12) = 0.36 > \alpha = 0.05$ 

 accept null; that is, the experimental (instructor) error (choose one) **is / is not** significant

(mean tests scores same for different instructors)

## 28.8 Pure Subsampling in Three Stages

The model for pure subsampling in three stages in a balanced single-factor study with completely randomized design is

$$Y_{ijk} = \mu_{..} + \tau_i + \varepsilon_{j(i)} + \eta_{ijk},$$

where *both*  $\tau_i$  and  $\varepsilon_{j(i)}$  are random. This design is then identical to a nested design where *both* factors are random.

## 28.9 Three-Factor Partially Nested Designs

It is possible to create designs which are partly nested and partly crossed. These are called *partially nested*, *partially hierarchical* or *cross-nested* designs. This material is not covered.