9.1

Nested Designs, Two Factors
(sec. 14.1)

~~example~~ : Handout 9, Example 1 (Let's start with an example to motivate the idea of a nested factor.)

motivating application

A manufacturer is studying the dimensional variability of a component that is produced on three machines.

factor A = machine $\quad$ $\boxed{m\ 1}$ $\boxed{m\ 2}$ $\boxed{m\ 3}$

response = dimension

Each machine has $b = 2$ spindles. Each spindle is specific to a machine.

machine 1		machine 2		machine 3	
sp 1(1)	sp 2(1)	sp 1(2)	sp 2(2)	sp 1(3)	sp 2(3)

(Note that the number of spindles in the experiment is actually $ab = 3 \cdot 2 = 6$

Here, we say that factor B is nested within factor A. Let's think about how that compares to designs we have previously considered.

Def: Factors A, B are crossed in an experimental design if the levels of B are the same at each level of A.

see HW *

example:

		student 1	2	---	b
subject	1				
	2				
	⋮				
	a				

Each student takes a test in each subject.

(9.2)

HW 1(a) *

Factor B is <u>nested</u> within A if the levels of B are different for each of the levels of A.

(levels of B are specific to a level of A)

example:

subject 1	....	subject a
student 1(1)	....	student 1(a)
⋮		⋮
student b(1)		student b(a)

Each student only takes the test in their subject.
Note that student j(1) is different from student j(a).

data structure: (hierarchical form)

$$\begin{array}{ccccccc} & i=1 & & \cdots & & i=a & \\ j=1(1) & \cdots & j=b(1) & & j=1(a) & \cdots & j=b(a) \\ \begin{bmatrix} Y_{111} \\ \vdots \\ Y_{11n} \end{bmatrix} & \cdots & \begin{bmatrix} Y_{1b1} \\ \vdots \\ Y_{1bn} \end{bmatrix} & \cdots & \begin{bmatrix} Y_{a11} \\ \vdots \\ Y_{a1n} \end{bmatrix} & \cdots & \begin{bmatrix} Y_{ab1} \\ \vdots \\ Y_{abn} \end{bmatrix} \end{array}$$

model: (A, B are fixed factors)

HW *

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk} \quad \begin{cases} i = 1, \dots, a \\ j = 1, \dots, b \\ k = 1, \dots, n \end{cases}$$

fixed effects:

$$\tau_1, \dots, \tau_a \quad \left(\sum_1^a \tau_i = 0 \right) \quad df = a-1$$

$$\begin{matrix} \beta_{1(1)}, \dots, \beta_{b(1)} \\ \vdots \\ \beta_{1(a)}, \dots, \beta_{b(a)} \end{matrix} \quad \left(\sum_{j=1}^{b} \beta_{j(i)} = 0 \text{ for each } i \right)$$

[1.]

$$df = ab - a = a(b-1)$$

df = number of free parameters

9.3

HW 1(b)

*

estimates: $\hat{\tau}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}$, $\hat{\beta}_{j(i)} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot}$

levels of factor A are compared $(\hat{\tau}_1, \ldots, \hat{\tau}_a)$

levels of B are compared only with the same level of A $(\hat{\beta}_{1(i)}, \ldots, \hat{\beta}_{b(i)})$

mean squares: $$MSA = \frac{bn \sum_{i=1}^{a} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2}{a-1}$$

HW

*

$$MSB(A) = \frac{n \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2}{a(b-1)}$$

$$MSE = \frac{\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij\cdot})^2}{ab(n-1)}$$

test statistics

$$F_A = \frac{MS_A}{MS_E}, \quad F_{B(A)} = \frac{MS_{B(A)}}{MS_E}$$

(remember that B is a fixed factor.)

see Routput

Back to Example 9.1

```
nested.mod = aov(dimension ~ machine/spindle)
summary(nested.mod)
```

$F_A = 18.934$, $F_{B(A)} = 9.906$

interpretation: The experiment finds that machine has an effect on dimension. Also, the experiment finds that spindle~~s~~ ~~within~~ within machine~~s~~ ~~have~~ has an effect on dimension.

Next: investigate differences between machines, and differences between spindles within machines.

9.4 parameter estimates:

$\hat{\tau}_1 = -2.125$, $\hat{\tau}_2 = 0.750$, $(\hat{\tau}_3 = 1.375)$

$\hat{\beta}_{1(1)} = 1.125$, $\hat{\beta}_{1(2)} = 1.250$, $\hat{\beta}_{1(3)} = -1.625$

$(\hat{\beta}_{2(1)} = -1.125$, $\hat{\beta}_{2(2)} = -1.250$, $\hat{\beta}_{2(3)} = 1.625)$

grouping information

See R output: $\boxed{3,2}$ $\boxed{1}$ machine

larger to smaller dimension

(Factor B levels)

HW *

We cannot directly compare spindles across machines since we only have data for spindle performance with respect to a particular machine

(Factor A levels) (B levels) (A level)

We can compare spindles within the same machine, or compare spindle * machine combinations.

grouping information

see R output: 1,1 − 1,2 (+) p = .0173 — m 1: $\boxed{\text{sp } 1}$ $\boxed{\text{sp } 2}$

2,1 − 2,2 (+) p = .0093 — m 2: $\boxed{\text{sp } 1}$ $\boxed{\text{sp } 2}$

3,1 − 3,2 (−) p = .0013 — m 3: $\boxed{\text{sp } 2}$ $\boxed{\text{sp } 1}$

Remember: there are 6 different spindles, we are only comparing those spindles within the same machine.

(9.5) Nested Design with a random factor:

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{ijk}$$

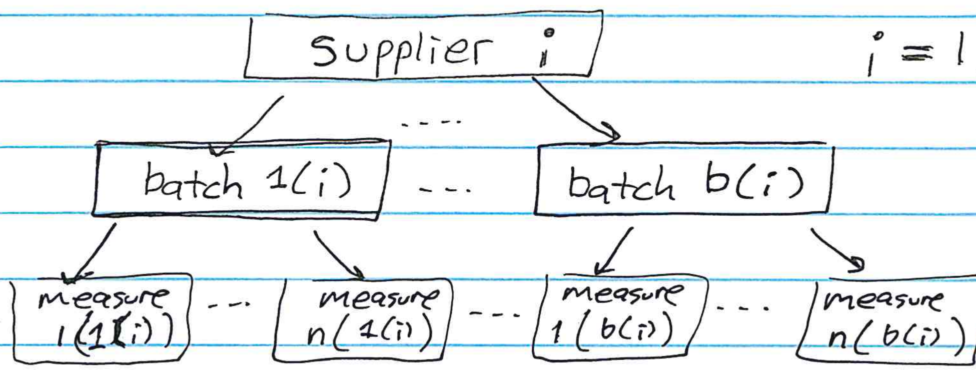
A Fixed: $\tau_1, \dots, \tau_a \quad (\sum_1^a \tau_i = 0)$

B(A) random: σ_β^2 , residual: σ^2

motivating application: Handout 9, Example 2

Company purchases raw material from $a = 3$ suppliers (fixed factor A). Batches of material are sampled from each supplier (random factor B(A)), and multiple determinations are made on each batch.

two-stage (hierarchical) sampling



$i = 1, 2, 3$

3.

σ_β^2 is the batch variance (between batch)

σ^2 is the measurement variance (within batch)

As we saw with the mixed effects model in Section 8, B(A) as a random factor leads to a different test statistic.

The motivations for this change in analysis repeat those we gave last Section.

(9.6) expected mean squares (B(A) random)

see HW 2(a) *

$$E(MS_A) = \sigma^2 + n\sigma_\beta^2 + \frac{bn}{a-1}\sum_i^a \tau_i^2$$

$$E(MS_{B(A)}) = \sigma^2 + n\sigma_\beta^2 \quad , \quad E(MSE) = \sigma^2$$

test for factor A effect: $H_0: \tau_1 = \dots = \tau_a = 0$

$$F_A = \frac{MSA}{MS_{B(A)}}$$

* error term is determined from expected mean squares

see R output

Back to Example 9.2

nested.test(supplier, batch, purity) ← user created function

$F_A = 0.969$, $p = .416$

(Note that $MSE = 2.63$, while $MS_{B(A)} = 7.77$. Thus, $F_A < \frac{MSA}{MSE}$

interpretation: The experiment finds that supplier does not have an effect on purity.

The evidence in favor of a supplier effect will be overstated when the incorrect error term is used.

Comments:

HW (*)

1. Think of batch as the experimental unit. The ~~appropriate~~ appropriate error term is then a measure of batch variance.

2. Summarize repeat measurements by taking a sample mean. The one-way ANOVA model for factor A using sample means results in test statistic 4.

$$F_A = \frac{b\sum_i^a(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2/(a-1)}{\sum_i^a\sum_j^b(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2/a(b-1)} = \frac{MSA}{MS_{B(A)}}$$

9.7

comment

Taking repeat measurements on each random factor level does not increase the pertinent sample size

variance estimates (see E(MS))

$$\hat{\sigma}^2 = MS_E \quad , \quad \hat{\sigma}^2_\beta = \frac{MS_{B(A)} - MS_E}{n}$$

Note: $$E(\hat{\sigma}^2_\beta) = \frac{(\sigma^2 + n\sigma^2_\beta) - \sigma^2}{n} = \sigma^2_\beta$$

see R output:

Back to Example 9.2

```
random.mod = lmer(y ~ A + (1|B:A))
summary(random.mod)
```

$$\hat{\sigma}^2_\beta = 1.710 \quad , \quad \hat{\sigma}^2 = 2.639$$

batch variance (between batches for a given supplier)

measurement variance (within batches for a given supplier)

End Notes, Section 9

1. There may be scenarios where it is not clear whether factors are crossed or nested. Consider the following.

$$A = \text{drug type (drug 1, drug 2)}$$
$$B = \text{dosage (low,high)}$$

It is not clear immediately whether to consider the levels of B as the same or different across the levels of A. The distinction between nested and crossed, then, may be viewed by which comparisons are of interest. Let's recall how contrasts are used in partitioning the sum of squares into the effects of interest.

Below is a table of contrasts for the above example analyzed as a two factor ANOVA (i.e., A,B are crossed).

	drug effect	dose effect	drug*dose interaction
drug1 low	-1	-1	+1
drug 1 high	-1	+1	-1
drug 2 low	+1	-1	-1
drug 2 high	+1	+1	+1

Below is a table of contrasts for the above example analyzed as a Nested Design.

	drug effect	dose effect for drug 1	dose effect for drug 2
drug1 low	-1	-1	0
drug 1 high	-1	+1	0
drug 2 low	+1	0	-1
drug 2 high	+1	0	+1

Both sets of contrasts are orthogonal, and thus both will partition the sum of squares. The design we choose then is based on whether we want to estimate dose effect within drug type or across drug type.

2. In the machine * spindle example, there are 6 unique treatment combinations. Thus, there are $\binom{6}{2} = 15$ pairwise comparisons listed in the R output. In general, comparisons between treatment combinations will be more difficult to summarize than comparisons between B levels within A levels.

3. The hierarchical sampling scheme can be extended. For example, suppose we are running an experiment where the experimental unit is plants of a specific variety. (Perhaps we are investigating the effect of pesticides.) Consider taking a random sample of plants, then taking a random sample of branches from each selected plant, then taking a random sample of leaves from each selected branch, then taking multiple measurements from each selected leaf. We then have a design with multiple nested random factors.

4. If we summarize repeat measurements using a sample mean, the data layout can be displayed as

exp units		B_1	B_2	$\cdots$	B_b
	A_1	$\overline{y}_{11}$	$\overline{y}_{12}$	$\cdots$	$\overline{y}_{1b}$
factor levels	A_2	$\overline{y}_{21}$	$\overline{y}_{22}$	$\cdots$	$\overline{y}_{2b}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	A_a	$\overline{y}_{a1}$	$\overline{y}_{a1}$	$\cdots$	$\overline{y}_{ab}$

The rows constitute random samples of experimental units at each of the a factor A levels. Thus, we have a one way ANOVA design in the sample means. The F statistic for testing a factor A effect from the above data layout equals $F_A = MSA/MSB\,(A)$ from a nested design.