28. Two-stage nested designs

- Recall Lecture 22 - partial confounding - where there were two replicates of an experiment. In each there were two blocks, but in one pair of blocks ABC was confounded, in the other set AB was confounded. Thus 'Block1' and 'Block2' meant different things within Replicate 1 than within Replicate 2 - the blocks were 'nested within replicates'.

- Another example - the surface finish of metal parts made on four machines is being studied. Different operators are used on each machine. Each machine is run by three different operators, and two specimens from each operator are tested.

Operators nested within machines											
Machine 1			Machine 2			Machine 3			Machine 4		
1	2	3	1	2	3	1	2	3	1	2	3
79	94	46	92	85	76	88	53	46	36	40	62
62	74	57	99	79	68	75	56	57	53	56	46

- Here 'Operator 1' makes sense only within the context of the machine on which this operator works - it refers to something different within Machine 1 than within Machine 2, etc.. When the levels of factor B (operators) make sense only within the levels of factor A, we say that B is 'nested within' A, and that this is a 'nested design'.

- Model and ANOVA. The effects model is

$$\begin{aligned} y_{ijk} &= \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}, \\ k &= 1, ..., n, \; j = 1, ..., b, \; i = 1, .., a. \end{aligned}$$

'Interaction' makes no sense here. The SS and df can be decomposed (how?) as

$$\begin{aligned} \sum_{i,j.k} \left(y_{ijk} - \bar{y}_{...} \right)^2 &= bn \sum_i \left(\bar{y}_{i..} - \bar{y}_{...} \right)^2 \\ &\quad + n \sum_{i,j} \left(\bar{y}_{ij.} - \bar{y}_{i..} \right)^2 + \sum_{i,j.k} \left(y_{ijk} - \bar{y}_{ij.} \right)^2, \\ SS_T &= SS_A + SS_{B(A)} + SS_E, \\ abn - 1 &= (a-1) + a(b-1) + ab(n-1). \end{aligned}$$

```
> g <- lm(y~machine + operator %in% machine)
> anova(g)
Analysis of Variance Table

Response: y
                 Df Sum Sq Mean Sq F value   Pr(>F)
machine           3 3617.7  1205.9 14.2709 0.000291
machine:operator  8 2817.7   352.2  4.1681 0.013408
Residuals        12 1014.0    84.5
```

- The F's and p-values in the preceding ANOVA are for the fixed effects model (when would this example have both factors fixed?). Both F's have MS_E in their denominators. We conclude that variation between the machines is very significant, and that within one or more machines, variation between operators is quite significant.

- In this 'two-stage' nested design we might have

  - both factors fixed (with $\sum_i \tau_i = 0$, $\sum_j \beta_{j(i)} = 0$ for each i),

  - A fixed and B random ($\sum_i \tau_i = 0$, each $\beta_{j(i)} \sim N(0, \sigma_\beta^2)$), or

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

The appropriate F-ratios are determined by the expected mean squares in each case.

$E(MS)$	A fixed B fixed	A fixed B random	A random B random
MS_A	$\sigma^2 + \frac{bn\sum_i \tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2$ $+\frac{bn\sum_i \tau_i^2}{a-1}$	$\sigma^2 + n\sigma_\beta^2$ $+bn\sigma_\tau^2$
$MS_{B(A)}$	$\sigma^2 + \frac{n\sum_{i,j} \beta_{j(i)}^2}{a(b-1)}$	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
MS_E	σ^2	σ^2	σ^2

- A fixed, B random
  - F to test for effect of A is $F_0 =$?
  - F to test for effect of B(A) is $F_0 =$?
  - $\hat{\sigma}_\beta^2 =$?
- A random, B random
  - F to test for effect of A is $F_0 =$?
  - F to test for effect of B(A) is $F_0 =$?
  - $\hat{\sigma}_\tau^2 =$? $\hat{\sigma}_\beta^2 =$?

If the machines were randomly chosen:

```
> cat("F to test effect of machines is",
1205.9/352.2, "and p-value is",
 1-pf(1205.9/352.2, 3, 8), "\n")
F to test effect of machines is 3.423907
and p-value is 0.07279175
```

If machines are random so are operator effects (since the operators are operating randomly chosen machines):

```
> cat("Estimate of operator within machines
variance is", (352.2-84.5)/2, "\n")
Estimate of operator within machines
variance is 133.85
```

29. Nested and crossed factors

- Extending the analysis of nested designs to the case where there are three factors A, B, C with B nested in A and C in B is straightforward.
  - In R: `lm(y ~ A + B%inA% + C%in%B)`
  - Consult or derive the expected mean squares in order to form appropriate F-ratios, and estimates of variance components.

- A design might have some factorial factors and some nested factors. Again, the analysis uses the same basic principles as above.

- Example: Printed circuit boards (used in electronic equipment - stereos, TVs etc.) have electronic components inserted on them by hand. There are three types of equipment (the 'fixtures') and 2 workplace layouts to be investigated. These factors are crossed (i.e. factorials), and fixed. In layout 1, four operators are (randomly) chosen to insert the components (2 replicates for each fixture). In layout 2, which is in a different location, this is done with a different 4 operators. So *operators is a random factor nested within locations*. A fixture/operator interaction (in each location) makes sense here. Response variable is $y =$ time to assemble.

	Layout 1				Layout 2			
Oper:	1	2	3	4	1	2	3	4
Fix. 1	22	23	28	25	26	27	28	24
	24	24	29	23	28	25	25	23
Fix. 2	30	29	30	27	29	30	24	28
	27	28	32	25	28	27	23	30
Fix. 3	25	24	27	26	27	26	24	28
	21	22	25	23	25	24	27	27

- Effects model:

$$\begin{aligned} & \text{Obs'n using fixture i, layout j,} \\ & \text{operator k, replicate l} \\ = \; & y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_{k(j)} \\ & + (\tau\beta)_{ij} + (\tau\gamma)_{ik(j)} + \varepsilon_{(ijk)l}, \\ & i \le a = 3,\ j \le b = 2,\ k \le c = 4,\ l \le n = 2. \end{aligned}$$

```
> g <- lm(y~(fixture + layout)^2 +
(operator + fixture*operator)%in%layout)
> anova(g)
Analysis of Variance Table

Response: time
                      Df Sum Sq Mean Sq F value
A - fixture            2 82.792  41.396 17.7411
B - layout             1  4.083   4.083  1.7500
AB - fix:lay           2 19.042   9.521  4.0804
C(B)-lay:oper          6 71.917  11.986  5.1369
AC(B) - fix:lay:oper  12 65.833   5.486  2.3512
Residuals             24 56.000   2.333
```

- Expected mean squares for this model:

$$
\begin{aligned}
E\left[MS_A\right] &= \sigma^2 + 2\sigma^2_{\tau\gamma} + 8\sum_i \tau_i^2, \\
E\left[MS_B\right] &= \sigma^2 + 6\sigma^2_{\gamma} + 24\sum_j \beta_j^2, \\
E\left[MS_{AB}\right] &= \sigma^2 + 2\sigma^2_{\tau\gamma} + 4\sum_{i,j} (\tau\beta)^2_{ij}, \\
E\left[MS_{C(B)}\right] &= \sigma^2 + 6\sigma^2_{\gamma}, \\
E\left[MS_{AC(B)}\right] &= \sigma^2 + 2\sigma^2_{\tau\gamma}, \\
E\left[MS_E\right] &= \sigma^2.
\end{aligned}
$$

- What are the F-ratios?

R programme on web site gives:

```
                        F value p-value
A - fixture              7.546  .0076
B - layout               0.341  .5807
AB - fix:lay             1.736  .2178
C(B)-lay:oper            5.138  .0016
AC(B) - fix:lay:oper     2.351  .0360
```

- How are the variance components estimated?

R programme on web site gives:

$$\begin{aligned}\hat{\sigma}^2_{\tau\gamma} &= 1.577\\ \hat{\sigma}^2_{\gamma} &= 1.609\end{aligned}$$

- Tukey confidence intervals on differences of the τ_i. In general, for mixed models these are $(\bar{y}_{i\ldots} - \bar{y}_{i'\ldots})\pm \frac{q_\alpha}{\sqrt{2}}\cdot\sqrt{\frac{2}{r}MS}$, where MS is the mean square in the denominator of the F to test the effect, and r is the number of observations used in each treatment mean. In this case we have ($\alpha = .05$)

$$\begin{aligned}\frac{q_\alpha}{\sqrt{2}}\cdot\sqrt{\frac{2}{r}MS} &= \frac{\texttt{qtukey}(.95, 3, 12)}{\sqrt{2}}\cdot\sqrt{\frac{2}{16}MS_{AC(B)}}\\ &= 2.21.\end{aligned}$$

```
The three fixture means are
  25.25 , 27.9375 , 25.0625.
```

- Conclusions:
  - Fixtures are significantly different; fixtures 1 and 3 result in smaller mean assembly times than fixture 2
  - Operators differ significantly within at least one of the layouts
  - The fixture $\times$ operator interaction is significant within at least one of the layouts (so some operators are quicker than others, using the same fixtures)
  - Layout does not have a significant effect on assembly time

- Recommendations:
  - Use only fixtures 1 and 3
  - Retrain the slower operators