## Nested ANOVA

A nested ANOVA is one that might "appear" to be a two way ANOVA but isn't.

The reason for this is that not all levels of two factors occur with each other.

When one level of one factor occurs with a level of another factor, those two levels are said to be crossed.

In a factorial or regular two way ANOVA, all factors are crossed with each other.

This is not true in a nested design.

Let's give an example before we get confused.

But, we'll use our cats again. But this time we'll set things up differently.

Example: We want to compare the effects of three different anesthetics.

The problem is, we don't have enough anesthetic from supplier 1, so we need to get more from a 2nd and even 3rd supplier.

So we decide to parcel out the anesthetic from the suppliers in such a way so that we might be able to evaluate the effect of supplier. But we can't do blocking (not enough drug).

Supplier 1 gives us 5 vials of A and 5 vials of C

Supplier 2 gives us 5 vials of A, 5 vials of B, and 5 vials of C

Supplier 3 gives us 5 vials B

How do we set up our experiment?

Remember, we are mostly interested in the effect of the anesthetic. Supplier is (hopefully) not that important, but we need to make sure.

We divide our cats into 3 groups of 10 each:

Drug A Drug B Drug C

each group gets 10 cats

Now we subdivide each group to account for our supplier:

Drug A		Drug B		Drug C	
supp 1	supp 2	supp 2	supp 3	supp 1	supp 2
5 cats					

Notice the setup. We subdivide our cats based on both supplier and drug, but this is not a fully crossed design:

supplier 3 only occurs with drug B. supplier 1 only occurs with drug A and C.

For a fully two way ANOVA we'd need every supplier to be crossed with every drug.

This is a "nested" design, since not every "sub" treatment occurs at every full treatment; the subtreatments are "nested" within the treatments.

See also your text on page 308, but notice that in the example there, every "source" is present only once (but still doesn't occur with every drug).

Either design will work.

Analyzing a nested ANOVA:

We need to calculate 5 different sums of squares.

We'll skip the mathematical details (they're on pages 310 - 311, but please ignore the machine formulas) and summarize them in words:

$SS_{total}$	The total sum of squares. We calculate this as usual by subtracting each measurement in turn from the grand mean, squaring, and adding up the total.					
SS <sub>groups</sub>	The group sum of squares. This is for the treatment we're really interested in. We calculate it the same way as we would the SS groups for a one way ANOVA (subtract the grand mean from the group mean, square, and multiply by $n_i$ , and add all of the result for each group). Above we'd have 3 groups.					
$\mathrm{SS}_{\mathrm{among\ subgroups}}$	The among subgroup sum of squares. We pretend that each subgroup is now a group, and calculate the sum of squares (subtract the grand mean from each subgroup mean, square, multiply by $n_{ij}$ , etc.). Above we'd have 6 subgroups.					
$\mathrm{SS}_{\mathrm{subgroups}}$	The sum of squares for the subgroups within a particular group. Think of it as trying to get at the variation between subgroups but within the same group. Think of comparing $\bar{x}_{ij}$ with $\bar{x}_i$ . The easiest way to calculate it is to take advantage of the following:					
	$SS_{subgroups} = SS_{among subgroups}$ - $SS_{groups}$					
SS <sub>error</sub> Since	there are duplicate measurements within each group, this could be					

considered SS within. We calculate it the easy way by doing:

 $SS_{error} = SS_{total}$  -  $SS_{among subgroups}$ 

Once we have all the various sums of squares, we arrange them into an ANOVA table as follows:

Source	SS	df	MS	F
among subgroups groups subgroups	$SS_{among \ subgroups} \ SS_{groups} \ SS_{subgroups}$	ab-1 a-1 a(b-1)	$SS_{groups}/(a-1)$ $SS_{subgroups}/a(b-1)$	MS <sub>groups</sub> /MS <sub>subgroups</sub> MS <sub>subgroups</sub> /MS <sub>error</sub>
error	SS <sub>error</sub>	N-ab	SS <sub>error</sub> /(N-ab)	
total	SS <sub>total</sub>	N-1		

The main thing we're interested is differences in our groups. If this is more variable than differences in subgroups (which is hopefully irrelevant), then we have a significant effect. That's why we divide  $MS_{groups}$  by  $MS_{subgroups}$ .

Let's walk through the example on page 308/309, skipping some of the math details:

Finally, let's see how to do this in R:

Start the usual way:

level <- scan()
102 103 108 109 104 105 104 104 110 108 106 107
drug <- scan()
1 1 2 2 3 3 1 1 2 2 3 3
srce <- scan(what = character())
a q d b l s a q d b l s
drug <- factor(drug)
srce <- factor(srce)</pre>

Now do the following:

chol <- aov(level ~ drug/srce)</pre>

This version of the aov command tells R to keep src inside drug:

summary(chol)

Df Sum Sq Mean Sq F value Pr(>F) drug 2 61.17 30.58 20.389 0.00211 \*\* drug:srce 3 1.50 0.50 0.333 0.80220 Residuals 6 9.00 1.50 ---Signif. codes: 0 `\*\*\*' 0.001 `\*\*' 0.01 `\*' 0.05 `.' 0.1 ` ' 1

And from this we see that the effect of subgroups (our "source" - the second line) is N.S. (source didn't make a difference). However, the first line for our groups (drug) is not quite right.

Unfortunately, R, by default, divides the  $MS_{drug}/MS_{residuals}$ , which is wrong. So we need to do it again and tell R to use the right denominator:

chol1 <- aov(level ~ drug + Error(drug:srce))</pre>

R will complain about the model being singular - just ignore it.

```
summary(chol1)
Error: drug:srce
         Df Sum Sq Mean Sq F value Pr(>F)
             61.17
                     30.58
                           61.17 0.0037 **
          2
druq
Residuals 3 1.50
                     0.50
___
Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 `' 1
Error: Within
         Df Sum Sq Mean Sq F value Pr(>F)
Residuals 6
                 9
                       1.5
```

Our second line now disappeared, but this time R correctly divides  $MS_{groups}$  (notice that this command shoved the  $SS_{subroups}$  into  $SS_{residuals}$ , but it all works out).

(Comment: I haven't tried it, but I suspect you can use the "Error" above to get the correct F value in a mixed model as well).