

14.5: Nested designs

Introduction

Crossed versus nested design

Factors are independent variables whose values we control and wish to study because we believe they have an effect on the dependent variable. While it is logical to think of factors and **levels** within factors as independent variables fully under our control, a moments reflection will come up with examples in which the groups (levels) depend on the factor.

Crossed – each level of a factor is in each level of the other factor. This was illustrated in [Chapter 14.1](#) on the crossed, balanced, fully replicated two-way ANOVA.

Nested – levels of one factor are NOT the same in each of the levels of the other factor. Nested designs are an important experimental design in science, and they have some advantages over the 2-way ANOVA design (for one), but they also have limitations.

Classic examples of nesting: culturing and passage of cell lines in routine cell colony maintenance means that even repeated experiments are done on different experimental units. Cells derived from one vial are different from cells derived from a different vial. Similarly, although mice from an inbred strain are thought to be genetically identical, environments vary across time, so mice from the same strain but born or purchased at different times are necessarily different. These scenarios involving time create a natural **block effect**. Thus, cells are nested by block effect passage number and mice are nested by block effect colony time. We introduced randomized block design in the previous section, [Chapter 14.4](#).

Statistical model

If Factor B is nested within Factor A, then a group or level within Factor B occurs only within a level of Factor A. Like the randomized block model, there will be no way to estimate the interaction in a nested two-way ANOVA. Our statistical model then is

$$Y_{ijk} = \mu + A_i + B_{j(i)} + \epsilon_{ijk}$$

Examples

Example 1. Three different drugs, 6 different sources of the drugs. The researcher obtains three different drugs from 6 different companies and wants to know if one of the drugs is better than another drug (Factor A) in lowering the blood cholesterol in women. There is always the possibility that different companies will be better or worse at making the drug. So the researchers also use the Factor Source (Factor B) to examine this possibility. Unfortunately they can not obtain all drugs from the same sources. This leads to a **Nested ANOVA** — notice that each drug is obtained from a different source.

We CANNOT perform the typical **two-factor ANOVA** because we cannot get a mean of the different drugs by combining the same levels of the Sources: the data is NOT crossed. The Sources of the drugs (Factor B) are NESTED within the type of Drug (Factor A); each source is only found in one of the Drug categories. So, we can't calculate a mean for the Drug levels independent of the SOURCE from which the drug came.

Table 14.5.1. Example of a nested design.

Drug A		Drug B		Drug C	
Source 1	Source 2	Source 3	Source 4	Source 5	Source 6
202.6	189.3	212.3	203.6	189.1	194.7
207.8	198.5	204.4	209.8	219.9	192.8
190.2	208.4	221.6	204.1	196.0	226.5
211.7	205.3	209.2	201.8	205.3	200.9
201.5	210.0	222.1	202.6	204.0	219.7

Scroll to end of this page to get the data set in **stacked worksheet** format, or [click here](#).

Compare Table 14.5.1 to CROSSED data structure (Table 14.5.2) — a typical two-factor ANOVA — which would look like

Table 14.5.2. Contents of Table 14.5.1 presented as crossed design.

Drug A		Drug B		Drug C	
Source 1	Source 2	Source 1	Source 2	Source 1	Source 2
202.6	189.3	?	?	?	?
207.8	198.5	?	?	?	?
190.2	208.4	?	?	?	?
211.7	205.3	?	?	?	?
201.5	210.0	?	?	?	?

We can take a mean of the different drugs by combining the same levels of the Sources. Here's the nested design (Table 14.5.3).

Table 14.5.3. Group means, nested design.

Drug A		Drug B		Drug C	
Source 1	Source 2	Source 3	Source 4	Source 5	Source 6
202.76	202.3	213.92	204.38	202.86	206.92

We can take a mean of the different drugs by combining the same levels of the Sources. Here's the crossed design (Table 14.5.4).

Drug A		Drug B		Drug C	
Source 1	Source 2	Source 1	Source 2	Source 1	Source 2
202.76	202.3	?	?	?	?

Why the “?” in Tables 14.5.2 and 14.5.4? Manufacturing source 1 & 2 do not sell Drug B and Drug C. So, there cannot be a crossed design.

Why can't we just use a One-Way ANOVA? Can't we just ANALYZE the three DRUGS separately, ignoring the source issue (after all, the drugs are not all made by the same manufacturer)? But it is not a one-way ANOVA problem... Here's why.

The researcher suspects that the response of a particular drug might be dependent upon the particular source from which the drug was purchased. So, the type of source from which the drug was purchased is another FACTOR. Thus, drugs from one source might have more (less) affect compared to drugs from another source regardless of the type of drug. However, each drug is NOT available from each source. Thus the research design can NOT be crossed and Drug is NESTED within Source.

We can ask ONLY two questions (hypotheses) from this NESTED ANOVA research design:

H_O : There is no difference in the average effect of the drugs on (tumor size, cholesterol level, blood pressure, etc.)

H_A : There is a difference in the average effect of the drugs on (tumor size, cholesterol level, blood pressure, etc.)

H_O : There is no difference in the average effect of the drugs on (tumor size, cholesterol level, blood pressure, etc.) purchased from different manufacturers.

H_A : There is a difference in the average effect of the drugs on (tumor size, cholesterol level, blood pressure, etc.) purchased from different manufacturers.

Notice that we do NOT examine the effect of the interaction between Drug type and source of the drug. Why not?

Table PageIndex5. Sources of Variation in Nested ANOVA.

Source of Variation	Sum of Squares	DF	Mean Squares
Total	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (X_{ijk} - \mu)^2$	$N - 1$	

Source of Variation	Sum of Squares	DF	Mean Squares
Among all subgroups	$\sum_{i=1}^a \sum_{j=1}^b (X_{ij} - \mu)^2$	$ab - 1$	
Among groups	$\sum_{i=1}^a (X_i - \mu)^2$	$a - 1$	$\frac{\text{among groups } SS}{\text{among groups } DF}$
Among subgroups	$\sum_{i=1}^a \sum_{j=1}^b n_{ij} (X_i - \mu)^2$	$a(b - 1)$	$\frac{\text{among subgroups } SS}{\text{among subgroups } DF}$
Error	Subtract all of the subgroup Sums of Squares from the Total Sums of Squares	$N - ab$	$\frac{\text{error } SS}{\text{error } DF}$

Testing nested ANOVA with one main factor

Perhaps surprisingly given the number of terms above, there are only two hypothesis tests, and, only one of REAL interest to us. There are exceptions (e.g., quantitative genetics provides many examples), but we are generally most interested in the among group test — this is the test of the main factor. In our example, the main factor was DRUG and whether the drugs differed in their effects on cholesterol levels. The second test is important in the sense that we prefer that it contributes little or no variation to the differences in cholesterol levels. But it might.

Table 14.5.6. F statistics for nested ANOVA.

F for the main effect is given as	$F = \frac{\text{Groups } MS}{\text{Subgroups } MS}$
F for the subgroup is given by	$F = \frac{\text{Subgroups } MS}{\text{Error } MS}$
and of course, use the appropriate DF when testing the F values!! The Critical Value $(F_{0.05(2), df \text{ numerator}, df \text{ denominator}})$	

One way to look at this: it would not make sense to conclude that an effect of the main group was significant if the variation in the subgroups was much, much larger. That's in part why we test the main effect with the subgroups MS and not the error MS. If variation due to the nested variable is not significant, then it is an estimate of the error variance, too.

The nested model we are describing is a two-factor ANOVA, but it is incomplete (compared to the balanced, fully crossed 2-way design we've talked about before). We don't have scores in every cell. Instead, each level of nested factor is paired with one and only one level of the other factor. In our example, Source is paired with only one other level of the other factor Drug (e.g., Source 1 goes with Drug 1 only), but the main effect is paired with 2 levels of the nesting factor (e.g., Drug 1 is manufactured at Source 1 and Source 2).

Note:

Nesting is strictly one-way. Drug is not nested within Source, for example.

Some important points about testing the null hypotheses in a nested design. For one, the test of the effect of the nesting factor (Source) is confounded by the interaction between the main factor. We don't actually know if the interaction is present, but we also get no way to test for it because of the incomplete design. We must therefore be cautious in our interpretation of the effect of the nested factor.

Consider our example. We want to interpret the effect of source as the contribution to the response based on variation among the different suppliers of the drugs. It might be good to know that some drug manufacturer is better (or worse) than others. However, differences among the sources for the different drugs are completely contained in the main effect factor (the test of effects of the different drugs themselves on the response). Therefore, the observed differences between sources COULD be entirely due to the effects of the different drugs and have nothing to do with variation among sources!!

Questions

- Identify the response variable and whether the described factor (in all caps) is suitable for crossed design or nested design
 - In a breeding colony of lab mice, BREEDERS are used to generate up to five LITTERS; effects on offspring REPRODUCTIVE SUCCESS.
 - Effects of individual TEACHERS at different SCHOOLS on STUDENT LEARNING in biology.

- c. Lisinopril, an ACE-inhibitor drug prescribed for treatment of high blood pressure, is now a generic drug, meaning a number of COMPANIES can manufacture and distribute the medication. Millions of DOSES of lisinopril are made each year; drug companies are required by the FDA to record when a dose is made and to record these dates by LOT NUMBER.
2. Work the example data set provide in this page. After loading the data set into Rcmdr (R), use linear model. The command to nest requires use of the forward slash, /. For example, if factor b is nested within factor a, then a/b. The linear model formula then,

```
Model <- lm(Obs ~ a/b, data=source)
```

1. Describe the problem, i.e., what is a? What is b? What are the hypotheses?
2. What is the statistical model?
3. Test the model.
4. Conclusions?

Data set used in this page

Drug	Source	Obs
A	s1	202.6
A	s1	207.8
A	s1	190.2
A	s1	211.7
A	s1	201.5
A	s2	189.3
A	s2	198.5
A	s2	208.4
A	s2	205.3
A	s2	210
B	s3	212.3
B	s3	204.4
B	s3	221.6
B	s3	209.2
B	s3	222.1
B	s4	203.6
B	s4	209.8
B	s4	204.1
B	s4	201.8
B	s4	202.6
C	s5	189.1
C	s5	219.9
C	s5	196
C	s5	205.3

C	s5	204
C	s6	194.7
C	s6	192.8
C	s6	226.5
C	s6	200.9
C	s6	219.7

This page titled [14.5: Nested designs](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.