

NESTED DESIGNS

Nested designs

In a crossed design, every level of factor A appears with each level of factor B (e.g., type \times method in the cement strength example).

In a nested design, each level of factor A contains different levels of factor B.

In a crossed design, every level of factor A appears with each level of factor B. For example, type \times method in the cement strength example.

In a nested design, each level of factor A contains different levels of factor B.

EXAMPLES

Example 1

Example. Training of mechanics as a function of school and instructor. The experimental unit is class, and the response is class learning score.

Factor A (school)	Factor B (instructor)	
	1	2
Atlanta	78	..
	82	
Chicago	:	
S.F.	:	

If the same two instructors are used in all three schools, we have a crossed design.

In this example, we examine the effects of “school” and “instructor” on the class learning scores.

If the same instructors are used in all three schools, we have “crossed” design.

This table shows the treatment combinations in this crossed design.

If each school uses *different* instructors, we have a nested design (“instructor is nested within school”). The instructors could be re-numbered 1-6:

school	Instructor					
	1	2	3	4	5	6
1	x	x				
2	x	x				
3			x	x		
4			x	x		
5					x	x
6					x	x

However, if each school uses different instructors, we have a “nested” design. We say that the factor instructor is nested within the factor school.

In analyzing the data from a such an experiment, we may as well renumber the instructors 1–6.

This table shows the treatment combinations in this nested design.

Example 2

Effect of three drugs (fixed effect) on the amount of pigment in the skin of rats. There are three levels of random effects:

1. Randomly select five rats for each drug.
2. Take three skin samples at random from the ventral side of each rat.
3. Take two readings of the optical density of each hydrolyzed skin sample.

Total variance = (variance among rats within drugs) + (variance among skin samples within rat) + (variance among readings within skin sample).

Variance estimates can be helpful in designing future studies—the most variable aspect of the experiment needs the greatest replication. For a discussion of “optimal allocation of resources”, see Sokal and Rohlf, *Biometry*, 4th ed., p. 397.

In the next example, we have an experiment where we study the effect of three drugs—considered fixed effect—on the amount of pigment in the skin of rats. The three drugs are three levels of a fixed-effect factor.

There are three levels of random effects:

First, randomly select five rats for each drug.

Second, take three skin samples at random from the ventral side of each rat.

Third, take two readings of the optical density of each hydrolyzed skin sample.

The total variance can thus be partitioned into the sum of (the variance among rats within drugs), (the variance among skin samples within rat), and (the variance among readings within skin sample).

For this experiment, one might be interested in estimating the variance components corresponding to the different sources by partitioning the total variance into parts corresponding to the different random effects mentioned above.

Variance estimates can be helpful in designing future studies—the most variable aspect of the experiment needs the greatest replication.

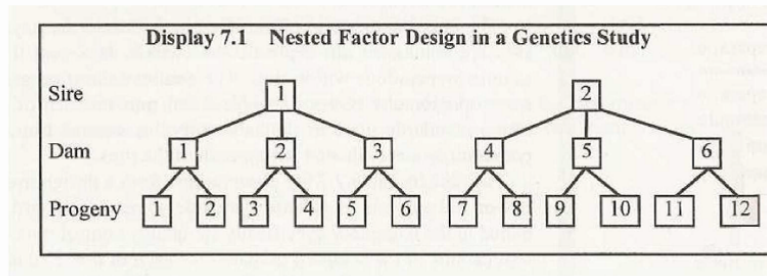
For a discussion of “optimal allocation of resources”, see Sokal and Rohlf’s paper in Biometry.

Example 3 (quantitative genetics)

Variation in blood pH as a function of genetic heritage: two male rats (sires), each mated with three female rats (dams).

Total variance = (variance among sires) + (variance among dams within sire) + (variance among offspring within dam).

Estimates of variance components can be used to formulate or validate models of inheritance of blood pH.



Here is another example from quantitative genetics that uses a nested design. This experiment studies the variation in blood pH among offspring from two male rats mated with three female rats.

Each of two male rats was mated with a different set of three female rats. So the factor “female rat” is nested within the factor “male rat”.

In this experiment, the total variance can be decomposed into the sum of (the variance attributable to male rats), (the variance attributable to the female rats) and (the variance among the offspring within each female rat).

MODEL

Two-factor Nested Design: Random effects

Model:

$$y_{ijk} = \mu + a_i + b_{j(i)} + c_{k(ij)}, (i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, c)$$

where there are

- a levels of factor A; $a_i \sim N(0, \sigma_a^2)$
- b levels of factor B nested under each level of A
 - $b_{j(i)}$ = effect of j th level of B nested under i th level of A
 - $b_{j(i)} \sim N(0, \sigma_{b(a)}^2)$.
- c replicates
 - $c_{k(ij)}$ = error for k th replicate nested under combinations of levels of A and B
 - $c_{k(ij)} \sim N(0, \sigma_{c(b)}^2)$.

Here we have the model equation for a two-factor nested design.

The levels of factor B are nested under each level of factor A.

($c_{k(ij)}$) indexes the replicates.

Note that in this equation, the effects for A, B, C are all random. The parameters of interest are the three variables components: the variances corresponding to each of the factors A, B, C.

With such a design, the research question is often about attributing the total variance to difference sources of variation (the different factors), and correspondingly, the analysis is often focused on partitioning the total variance into the different variance components.

school (A)
instructor
(B within A)
class
(C within B)

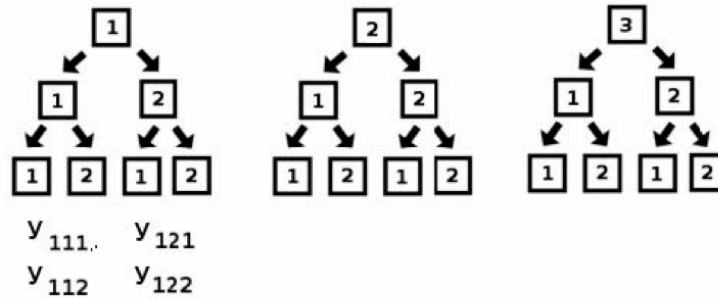


Illustration of a nested model

This figure illustrates how the factor levels are nested in a nested design.

INFERENCE

Sums of squares for the two-factor nested design

One can show that

$$\begin{aligned} & \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \bar{y}_{...})^2 \\ &= bc \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2 + c \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (\bar{y}_{ijk} - \bar{y}_{ij.})^2 \\ & \text{SSTotal} = \text{SSA} + \text{SSB(A)} + \text{SSE}, \end{aligned}$$

where SSB(A) is the sum of squares for B nested within A.

Degrees of freedom:

$$abc - 1 = (a - 1) + a(b - 1) + ab(c - 1).$$

The partitioning of the sum of squares in a nested design looks similar as in an ordinary two-factor design.

The nested design is a special two-factor design: The key is that the levels of factor B are all nested under each level of factor A. The levels of factor B nested under different levels of factor A are all different.

ANOVA Table

Table 7.8 Expected mean squares for the analysis of variance of a nested design with three random factors, *A*, *B*, and *C*

<i>Source of Variation</i>	<i>Degrees of Freedom</i>	<i>Mean Square</i>	<i>Expected Mean Square</i>
Total	$abc - 1$		
<i>A</i>	$a - 1$	MSA	$\sigma_{c(b)}^2 + c\sigma_{b(a)}^2 + bc\sigma_a^2$
<i>B</i> within <i>A</i>	$a(b - 1)$	$MS(B/A)$	$\sigma_{c(b)}^2 + c\sigma_{b(a)}^2$
<i>C</i> within <i>B</i>	$ab(c - 1)$	$MS(C/B)$	$\sigma_{c(b)}^2$

Here is the ANOVA table for a nested design, which summarizes, among other things, summarizes the partitioning of the sum of squares (see the previous page).

Again, pay attention to the column of “Expected Mean Square” and also note that since factors are random, the expected mean square are sums of the variance components.

Hypothesis tests.

To test $H_0: \sigma_a^2 = 0$, compare MSA/MSB(A) to an $F_{a-1, a(b-1)}$ distribution.

If factor A is a fixed effect, the associated F -statistic is the same as above, namely MSA/MSB(A) (see Kuehl, p. 248).

To test $H_0: \sigma_{b(a)}^2 = 0$, compare MSB(A)/MSC(B) to an $F_{a(b-1), ab(c-1)}$ distribution.

When the factors are all random, we should formulate our research questions in terms of the variance components.

For example, the testing null hypothesis (σ_a^2 equals 0) corresponds to asking whether there is any significant between factor A level variation—in other words, whether some of the variation in the data set we observe is attributable to factor A.

If factor is a fixed-effect factor, the corresponding F -test will have the same form, but the research question should then be formulated in terms of the group means: for example, we can ask whether the group means are the same.

Similarly, we can test variation components corresponding to factor B. The numbers of degrees of freedoms are summarized in the ANOVA table on the previous page.

Estimation of variance components

$$\begin{aligned}\hat{\sigma}_{c(b)}^2 &= MS_{C(B)} \\ \hat{\sigma}_{b(a)}^2 &= \frac{MS_{B(A)} - MS_{C(B)}}{c} \\ \hat{\sigma}_a^2 &= \frac{MSA - MSB(A)}{bc}\end{aligned}$$

Again, from the ANOVA table, we should be able to derive the estimators of the variance components.

Recall the basic steps:

We equate the mean squares (in the third column of ANOVA table) to the correspond values of the expected mean squares (the last column in the ANOVA table), add “hats” to the variance components to indicate that we want to solve for the estimators of the variance components, and solve for the estimators of the variance components from the resulting equations.

Here we will need three equations, to solve for estimators of the three variance components.

Note

There is no interaction between A and B, since the levels of B are different for different levels of A.

It can be shown that, if we do not re-code the levels of B to all distinct values,

$$\begin{aligned} & \text{SSB(A) (from fitting a nested design)} \\ &= \text{SSB} + \text{SSAB (from fitting a crossed design),} \end{aligned}$$

so a nested analysis can be done with output from a factorial (crossed) analysis.

Note that there is no interaction term between A and B, since the levels of B are different for different levels of A.

It can be shown that, if we do not re-code the levels of B, SSB(A) from fitting a nested model is the same as SSB plus SSAB from fitting a crossed model, so a nested analysis can be done with output from a factorial (crossed) analysis.

Data analysis in R

See [script3.html](#) for notes and R code for analyzing nested models.

See [script3.html](#) for notes and R code for analyzing nested models.

Example

Example. Class learning scores (data set class). In R (script3):

```
> tmp <- lm( score ~ school.f + instruct.f %in% school.f )
> anova(tmp)
```

The F -statistic and P -value for instructor within school are correct. If school and instructor are both random effects, the F -statistic for school should be recalculated as: $MS(\text{school}) / MS(\text{instructor within school})$.

Variance components (class example):

Factor	Variance component	Percent of total
school	$\hat{\sigma}_a^2 = \frac{78.25 - 189.2}{2(2)} < 0 !$	0
instructor (school)	$\hat{\sigma}_{b(a)}^2 = \frac{189.17 - 7.00}{2} = 91.1$	93
class (instructor)	$\hat{\sigma}_{c(b)}^2 = 7.00$	7

So, instructor is by far the most important source of variation in the class scores. School contributes no additional variability.

Recall the example where the factor “instructor” is nested in the factor school.

When fitting the model, note the special operator “%in%”, this is telling R that the factor instructor (‘instruct.f’) is nested in the factor school (‘school.f’).

In this example, we estimate the variance components using the output from the ANOVA table, we will find out that the variance component corresponding to school is actually less than 0.

In such a case, one common practice is to let the estimated value be 0 for that variance component. It indicates that the between-school variation is overwhelmed by other sources of variation: in this example, between-instructor variation and between-class variation.

In this example, we see that the factor “instructor” is the most important source of variation in the class scores.

In practice, when you see a negative estimated variance component, you may want to talk to the person who conducted the experiment and consult with a statistician to discuss possible reasons, and the best way to handle the situation. In some cases, a negative estimated variance component may be the result of negative correlation among observations.

DESIGN WITH NESTED AND CROSSED FACTORS

Example

Example. Suppose that two teaching methods are compared in the preceding example. Each × in the table represents a class score.

	School					
	1		2		3	
	Instructor					
	1	2	3	4	5	6
Method 1	×	×				
Method 2	×	×				

Note:

- method is crossed with instructor
- method is crossed with school
- instructor is nested within school

In more complex designs, it is possible to have both crossed factors and nested factors.

Suppose that two teaching methods are compared in the preceding example, and each instructor will try both methods. The treatment structure is summarized in this table.

Under this arrangement, we see that the factor instructor is still nested in the factor school: the instructors from different schools are different.

But each instructor will try the same two methods, so the factor method is crossed with instructor and with school.

Note that the table only summarizes the treatment structure, it does not indicator how the treatments will be randomized. In our textbook, it sometimes use the term “treatment design” to describe the structure of the treatments.

Model

The model (using the above example):

$$y_{ijkl} = \mu + \alpha_i + b_j + c_{k(j)} + (ab)_{ij} + (ac)_{ik(j)} + e_{l(ijk)},$$

$(i = 1, 2; j = 1, 2, 3; k = 1, 2; l = 1, \dots, n)$

where

- α_i is the fixed effect of the i th method ($i = 1, 2$)
- b_j is the random effect of school ($j = 1, 2, 3$)
- $c_{k(j)}$ is the random effect of instructor within school ($k = 1, 2$)
- $(ab)_{ij}$ is the method \times school interaction (random)
- $(ac)_{ik(j)}$ is the method \times instructor interaction, nested within school
- $e_{l(ijk)}$ is the random error associated with class ($l = 1, 2$)

Notice that there can not be a school \times instructor interaction, nor a method \times school \times instructor interaction, because instructor is nested within school.

Here we give the statistical model corresponding to the example.

Note that there can not be a school by instructor interaction, nor a method by school by instructor interaction, because instructor is nested within school.

In practice, when fitting models in R with nested factors, it is often possible to simply re-code the instructors to avoid confusion, but it can be unwieldy when a factor is nested under another factor that has a large number factor levels. In those situations, it is better to use the special operator “%in%” to inform R about the nesting structure.

Analysis

See Kuehl, p. 252-254. You're not responsible for details of the analysis, but you should be able to recognize the design.

In this class, we will not talk about analysis of complex designs that have multiple levels of nesting and crossing. For those experiments, it is only asked that you are able to correctly recognize the design structures.

UNBALANCED DATA

Notes

If sample sizes are unequal, the notation and computations become much more complicated, and there are no exact significance tests for the upper $k - 1$ levels in a k -level nested ANOVA. Sokal and Rohlf (Biometry, 4th ed., p. 301) tackle an example illustrating the difficulties.

Unlike the one-factor experiment, with multiple levels, when the sample sizes are not balanced, the analysis becomes more complex. We will cover the analysis in such situations. If you see such examples in practice, consider talking to an experienced statistician.

Summary

- In a crossed design, every level of factor A appears with each level of factor B.
- In a nested design, each level of factor A contains different levels of factor B.
- Study the R code and notes on how to analyze a nested design.
- Study the ANOVA table to see how the sum of squares is partitioned in a nested design.

In a crossed design, every level of factor A appears with each level of factor B.

In a nested design, each level of factor A contains different levels of factor B.

Please study the R code and notes on how to analyze a nested design.

Study the ANOVA table to see how the sum of squares is partitioned in a nested design.