

4.9: Nested Anova

Learning Objectives

- Use nested anova when you have one measurement variable and more than one nominal variable, and the nominal variables are nested (form subgroups within groups). It tests whether there is significant variation in means among groups, among subgroups within groups, etc.

When to use it

Use a nested anova (also known as a hierarchical anova) when you have one measurement variable and two or more nominal variables. The nominal variables are nested, meaning that each value of one nominal variable (the subgroups) is found in combination with only one value of the higher-level nominal variable (the groups). All of the lower level subgroupings must be random effects (model II) variables, meaning they are random samples of a larger set of possible subgroups.

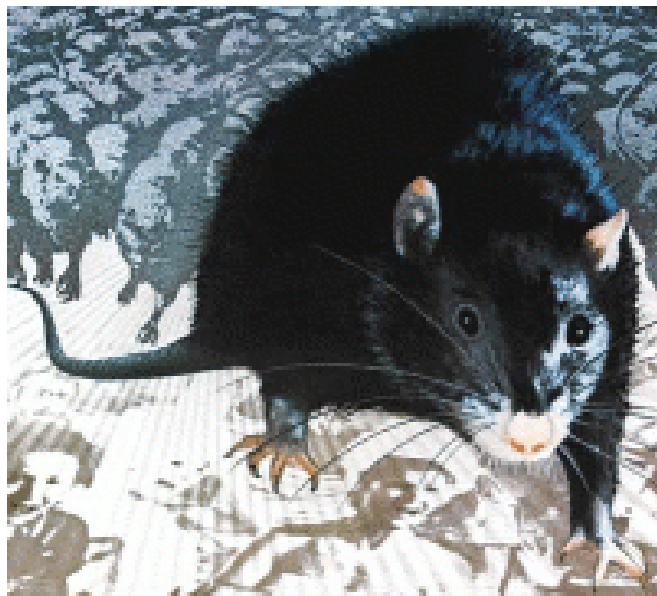


Fig. 4.9.1 Ben.

Nested analysis of variance is an extension of one-way anova in which each group is divided into subgroups. In theory, you choose these subgroups randomly from a larger set of possible subgroups. For example, a friend of mine was studying uptake of fluorescently labeled protein in rat kidneys. He wanted to know whether his two technicians, who I'll call Brad and Janet, were performing the procedure consistently. So Brad randomly chose three rats, and Janet randomly chose three rats of her own, and each technician measured protein uptake in each rat.

If Brad and Janet had measured protein uptake only once on each rat, you would have one measurement variable (protein uptake) and one nominal variable (technician) and you would analyze it with one-way anova. However, rats are expensive and measurements are cheap, so Brad and Janet measured protein uptake at several random locations in the kidney of each rat:

| Technician: | | Brad | | | Janet | |
|-------------|--------|--------|---------|--------|--------|--------|
| Rat: | Arnold | Ben | Charlie | Dave | Eddy | Frank |
| | 1.119 | 1.045 | 0.9873 | 1.3883 | 1.3952 | 1.2574 |
| | 1.2996 | 1.1418 | 0.9873 | 1.104 | 0.9714 | 1.0295 |
| | 1.5407 | 1.2569 | 0.8714 | 1.1581 | 1.3972 | 1.1941 |
| | 1.5084 | 0.6191 | 0.9452 | 1.319 | 1.5369 | 1.0759 |
| | 1.6181 | 1.4823 | 1.1186 | 1.1803 | 1.3727 | 1.3249 |

| Technician: | | Brad | | | Janet | |
|-------------|--------|--------|--------|--------|--------|--------|
| | 1.5962 | 0.8991 | 1.2909 | 0.8738 | 1.2909 | 0.9494 |
| | 1.2617 | 0.8365 | 1.1502 | 1.387 | 1.1874 | 1.1041 |
| | 1.2288 | 1.2898 | 1.1635 | 1.301 | 1.1374 | 1.1575 |
| | 1.3471 | 1.1821 | 1.151 | 1.3925 | 1.0647 | 1.294 |
| | 1.0206 | 0.9177 | 0.9367 | 1.0832 | 0.9486 | 1.4543 |

Because there are several observations per rat, the identity of each rat is now a nominal variable. The values of this variable (the identities of the rats) are nested under the technicians; **rat A** is only found with Brad, and **rat D** is only found with Janet. You would analyze these data with a nested anova. In this case, it's a two-level nested anova; the technicians are groups, and the rats are subgroups within the groups. If the technicians had looked at several random locations in each kidney and measured protein uptake several times at each location, you'd have a three-level nested anova, with kidney location as subsubgroups within the rats. You can have more than three levels of nesting, and it doesn't really make the analysis that much more complicated.

Note that if the subgroups, subsubgroups, etc. are distinctions with some interest (fixed effects, or model I, variables), rather than random, you should not use a nested anova. For example, Brad and Janet could have looked at protein uptake in two male rats and two female rats apiece. In this case you would use a two-way anova to analyze the data, rather than a nested anova.

When you do a nested anova, you are often only interested in testing the null hypothesis about the group means; you may not care whether the subgroups are significantly different. For this reason, you may be tempted to ignore the subgrouping and just use all of the observations in a one-way anova, ignoring the subgrouping. This would be a mistake. For the rats, this would be treating the 30 observations for each technician (10 observations from each of three rats) as if they were 30 independent observations. By using all of the observations in a one-way anova, you compare the difference in group means to the amount of variation within each group, pretending that you have 30 independent measurements of protein uptake. This large number of measurements would make it seem like you had a very accurate estimate of mean protein uptake for each technician, so the difference between Brad and Janet wouldn't have to be very big to seem "significant." You would have violated the assumption of independence that one-way anova makes, and instead you have what's known as pseudoreplication.

What you could do with a nested design, if you're only interested in the difference among group means, is take the *average* for each subgroup and analyze them using a one-way anova. For the example data, you would take the average protein uptake for each of the three rats that Brad used, and each of the three rats that Janet used, and you would analyze these six values using one-way anova. If you have a balanced design (equal sample sizes in each subgroup), comparing group means with a one-way anova of subgroup means is mathematically identical to comparing group means using a nested anova (and this is true for a nested anova with more levels, such as subsubgroups). If you don't have a balanced design, the results won't be identical, but they'll be pretty similar unless your design is very unbalanced. The advantage of using one-way anova is that it will be more familiar to more people than nested anova; the disadvantage is that you won't be able to compare the variation among subgroups to the variation within subgroups. Testing the variation among subgroups often isn't biologically interesting, but it can be useful in the optimal allocation of resources, deciding whether future experiments should use more rats with fewer observations per rat.

Null hypotheses

A nested anova has one null hypothesis for each level. In a two-level nested anova, one null hypothesis is that the groups have the same mean. For our rats, this null would be that Brad's rats had the same mean protein uptake as the Janet's rats. The second null hypothesis is that the subgroups within each group have the same means. For the example, this null would be that all of Brad's rats have the same mean, and all of Janet's rats have the same mean (which could be different from the mean for Brad's rats). A three-level nested anova would have a third null hypothesis, that all of the locations within each kidney have the same mean (which could be a different mean for each kidney), and so on.

How the test works

Remember that in a one-way anova, the test statistic, F_s , is the ratio of two mean squares: the mean square among groups divided by the mean square within groups. If the variation among groups (the group mean square) is high relative to the variation within groups, the test statistic is large and therefore unlikely to occur by chance. In a two-level nested anova, there are two F statistics,

one for subgroups (F_{subgroup}) and one for groups (F_{group}). You find the subgroup F -statistic by dividing the among-subgroup mean square, MS_{subgroup} (the average variance of subgroup means within each group) by the within-subgroup mean square, MS_{within} (the average variation among individual measurements within each subgroup). You find the group F -statistic by dividing the among-group mean square, MS_{group} (the variation among group means) by MS_{subgroup} . You then calculate the P value for the F -statistic at each level.

For the rat example, the within-subgroup mean square is 0.0360 and the subgroup mean square is 0.1435, making the $F_{\text{subgroup}} = 0.1435/0.0360 = 3.9818$. There are 4 degrees of freedom in the numerator (the total number of subgroups minus the number of groups) and 54 degrees of freedom in the denominator (the number of observations minus the number of subgroups), so the P value is 0.0067. This means that there is significant variation in protein uptake among rats within each technician. The F_{group} is the mean square for groups, 0.0384, divided by the mean square for subgroups, 0.1435, which equals 0.2677. There is one degree of freedom in the numerator (the number of groups minus 1) and 4 degrees of freedom in the denominator (the total number of subgroups minus the number of groups), yielding a P value of 0.632. So there is no significant difference in protein abundance between the rats Brad measured and the rats Janet measured.

For a nested anova with three or more levels, you calculate the F -statistic at each level by dividing the MS at that level by the MS at the level immediately below it.

If the subgroup F -statistic is not significant, it is possible to calculate the group F -statistic by dividing MS_{group} by MS_{pooled} , a combination of MS_{subgroup} and MS_{within} . The conditions under which this is acceptable are complicated, and some statisticians think you should never do it; for simplicity, I suggest always using $MS_{\text{group}}/MS_{\text{subgroup}}$ to calculate F_{group} .

Partitioning variance and optimal allocation of resources

In addition to testing the equality of the means at each level, a nested anova also partitions the variance into different levels. This can be a great help in designing future experiments. For our rat example, if most of the variation is among rats, with relatively little variation among measurements within each rat, you would want to do fewer measurements per rat and use a lot more rats in your next experiment. This would give you greater statistical power than taking repeated measurements on a smaller number of rats. But if the nested anova tells you there is a lot of variation among measurements but relatively little variation among rats, you would either want to use more observations per rat or try to control whatever variable is causing the measurements to differ so much.

If you have an estimate of the relative cost of different parts of the experiment (in time or money), you can use this formula to estimate the best number of observations per subgroup, a process known as optimal allocation of resources:

$$n = \sqrt{\frac{(C_{\text{subgroup}} \times V_{\text{within}})}{(C_{\text{within}} \times V_{\text{subgroup}})}} \quad (4.9.1)$$

where n is the number of observations per subgroup, C_{within} is the cost per observation, C_{subgroup} is the cost per subgroup (not including the cost of the individual observations), V_{subgroup} is the percentage of the variation partitioned to the subgroup, and V_{within} is the percentage of the variation partitioned to within groups. For the rat example, V_{subgroup} is 23.0% and V_{within} is 77% (there's usually some variation partitioned to the groups, but for these data, groups had 0% of the variation). If we estimate that each rat costs \$200 to raise, and each measurement of protein uptake costs \$10, then the optimal number of observations per rat is

$\sqrt{\frac{(200 \times 77)}{(10 \times 23)}}$, which equals 8 rats per subgroup. The total cost per subgroup will then be \$200 to raise the rat and $8 \times \$10 = \80 for the observations, for a total of \$280; based on your total budget for your next experiment, you can use this to decide how many rats to use for each group.

For a three-level nested anova, you would use the same equation to allocate resources; for example, if you had multiple rats, with multiple tissue samples per rat kidney, and multiple protein uptake measurements per tissue sample. You would start by determining the number of observations per subsubgroup; once you knew that, you could calculate the total cost per subsubgroup (the cost of taking the tissue sample plus the cost of making the optimal number of observations). You would then use the same equation, with the variance partitions for subgroups and subsubgroups, and the cost for subgroups and the total cost for subsubgroups, and determine the optimal number of subsubgroups to use for each subgroup. You could use the same procedure for as higher levels of nested anova.

It's possible for a variance component to be zero; the groups (Brad vs. Janet) in our rat example had 0% of the variance, for example. This just means that the variation among group means is smaller than you would expect, based on the amount of variation among subgroups. Because there's variation among rats in mean protein uptake, you would expect that two random samples of

three rats each would have different means, and you could predict the average size of that difference. As it happens, the means of the three rats Brad studied and the three rats Janet studied happened to be closer than expected by chance, so they contribute 0% to the overall variance. Using zero, or a very small number, in the equation for allocation of resources may give you ridiculous numbers. If that happens, just use your common sense. So if V_{subgroup} in our rat example (the variation among rats within technicians) had turned out to be close to 0%, the equation would have told you that you would need hundreds or thousands of observations per rat; in that case, you would design your experiment to include one rat per group, and as many measurements per rat as you could afford.

Often, the reason you use a nested anova is because the higher level groups are expensive and lower levels are cheaper. Raising a rat is expensive, but looking at a tissue sample with a microscope is relatively cheap, so you want to reach an optimal balance of expensive rats and cheap observations. If the higher level groups are very inexpensive relative to the lower levels, you don't need a nested design; the most powerful design will be to take just one observation per higher level group. For example, let's say you're studying protein uptake in fruit flies (*Drosophila melanogaster*). You could take multiple tissue samples per fly and make multiple observations per tissue sample, but because raising 100 flies doesn't cost any more than raising 10 flies, it will be better to take one tissue sample per fly and one observation per tissue sample, and use as many flies as you can afford; you'll then be able to analyze the data with one-way anova. The variation among flies in this design will include the variation among tissue samples and among observations, so this will be the most statistically powerful design. The only reason for doing a nested anova in this case would be to see whether you're getting a lot of variation among tissue samples or among observations within tissue samples, which could tell you that you need to make your laboratory technique more consistent.

Unequal sample sizes

When the sample sizes in a nested anova are unequal, the P values corresponding to the F -statistics may not be very good estimates of the actual probability. For this reason, you should try to design your experiments with a "balanced" design, meaning equal sample sizes in each subgroup. (This just means equal numbers at each level; the rat example, with three subgroups per group and 10 observations per subgroup, is balanced). Often this is impractical; if you do have unequal sample sizes, you may be able to get a better estimate of the correct P value by using modified mean squares at each level, found using a correction formula called the Satterthwaite approximation. Under some situations, however, the Satterthwaite approximation will make the P values less accurate. If you cannot use the Satterthwaite approximation, the P values will be conservative (less likely to be significant than they ought to be), so if you never use the Satterthwaite approximation, you're not fooling yourself with too many false positives. Note that the Satterthwaite approximation results in fractional degrees of freedom, such as 2.87; don't be alarmed by that (and be prepared to explain it to people if you use it). If you do a nested anova with an unbalanced design, be sure to specify whether you use the Satterthwaite approximation when you report your results.

Assumptions

Nested anova, like all anovas, assumes that the observations within each subgroup are normally distributed and have equal standard deviations.

Example

Keon and Muir (2002) wanted to know whether habitat type affected the growth rate of the lichen *Usnea longissima*. They weighed and transplanted 30 individuals into each of 12 sites in Oregon. The 12 sites were grouped into 4 habitat types, with 3 sites in each habitat. One year later, they collected the lichens, weighed them again, and calculated the change in weight. There are two nominal variables (site and habitat type), with sites nested within habitat type. You could analyze the data using two measurement variables, beginning weight and ending weight, but because the lichen individuals were chosen to have similar beginning weights, it makes more sense to use the change in weight as a single measurement variable. The results of a nested anova are that there is significant variation among sites within habitats ($F_{8, 200} = 8.11$, $P = 1.8 \times 10^{-9}$) and significant variation among habitats ($F_{3, 8} = 8.29$, $P = 0.008$). When the Satterthwaite approximation is used, the test of the effect of habitat is only slightly different ($F_{3, 8.13} = 8.76$, $P = 0.006$).



Fig. 4.9.2 Old man's beard lichen, *Usnea longissima*.

Graphing the results

The way you graph the results of a nested anova depends on the outcome and your biological question. If the variation among subgroups is not significant and the variation among groups is significant—you're really just interested in the groups, and you used a nested anova to see if it was okay to combine subgroups—you might just plot the group means on a bar graph, as shown for one-way anova. If the variation among subgroups is interesting, you can plot the means for each subgroup, with different patterns or colors indicating the different groups.

Similar tests

Both nested anova and two-way anova (and higher level anovas) have one measurement variable and more than one nominal variable. The difference is that in a two-way anova, the values of each nominal variable are found in all combinations with the other nominal variable; in a nested anova, each value of one nominal variable (the subgroups) is found in combination with only one value of the other nominal variable (the groups).

If you have a balanced design (equal number of subgroups in each group, equal number of observations in each subgroup), you can perform a one-way anova on the subgroup means. For the rat example, you would take the average protein uptake for each rat. The result is mathematically identical to the test of variation among groups in a nested anova. It may be easier to explain a one-way anova to people, but you'll lose the information about how variation among subgroups compares to variation among individual observations.

How to do the test

Spreadsheet

I have made spreadsheets to do two-level nested anova nested2.xls, with equal or unequal sample sizes, on up to 50 subgroups with up to 1000 observations per subgroup. It does significance tests and partitions the variance. The spreadsheet tells you whether the Satterthwaite approximation is appropriate, using the rules on p. 298 of Sokal and Rohlf (1983), and gives you the option to use it. F_{group} is calculated as $MS_{group} / MS_{subgroup}$. The spreadsheet gives the variance components as percentages of the total. If the estimate of the group component would be negative (which can happen), it is set to zero.

I have also written spreadsheets to do three-level nested anova nested3.xls and four-level nested anova nested4.xls.

Web page

I don't know of a web page that will let you do nested anova.

R

Salvatore Mangiafico's *R Companion* has a sample [R program for nested anova](#).

SAS

You can do a nested anova with either PROC GLM or PROC NESTED. PROC GLM will handle both balanced and unbalanced designs, but does not partition the variance; PROC NESTED partitions the variance but does not calculate P values if you have an unbalanced design, so you may need to use both procedures.

You may need to sort your dataset with PROC SORT, and it doesn't hurt to include it.

In PROC GLM, list all the nominal variables in the CLASS statement. In the MODEL statement, give the name of the measurement variable, then after the equals sign give the name of the group variable, then the name of the subgroup variable followed by the group variable in parentheses. SS1 (with the numeral one, not the letter el) tells it to use type I sums of squares. The TEST statement tells it to calculate the F -statistic for groups by dividing the group mean square by the subgroup mean square, instead of the within-group mean square (H stands for "hypothesis" and E stands for "error"). "HTYPE=1 ETYPE=1" also tells SAS to use "type I sums of squares"; I couldn't tell you the difference between them and types II, III and IV, but I'm pretty sure that type I is appropriate for a nested anova.

Here is an example of a two-level nested anova using the rat data.

```
DATA bradvsjanet;
INPUT tech $ rat $ protein @@;
DATALINES;

Janet 1 1.119 Janet 1 1.2996 Janet 1 1.5407 Janet 1 1.5084 Janet 1 1.6181
Janet 1 1.5962 Janet 1 1.2617 Janet 1 1.2288 Janet 1 1.3471 Janet 1 1.0206
Janet 2 1.045 Janet 2 1.1418 Janet 2 1.2569 Janet 2 0.6191 Janet 2 1.4823
Janet 2 0.8991 Janet 2 0.8365 Janet 2 1.2898 Janet 2 1.1821 Janet 2 0.9177
Janet 3 0.9873 Janet 3 0.9873 Janet 3 0.8714 Janet 3 0.9452 Janet 3 1.1186
Janet 3 1.2909 Janet 3 1.1502 Janet 3 1.1635 Janet 3 1.151 Janet 3 0.9367
Brad 5 1.3883 Brad 5 1.104 Brad 5 1.1581 Brad 5 1.319 Brad 5 1.1803
Brad 5 0.8738 Brad 5 1.387 Brad 5 1.301 Brad 5 1.3925 Brad 5 1.0832
Brad 6 1.3952 Brad 6 0.9714 Brad 6 1.3972 Brad 6 1.5369 Brad 6 1.3727
Brad 6 1.2909 Brad 6 1.1874 Brad 6 1.1374 Brad 6 1.0647 Brad 6 0.9486
Brad 7 1.2574 Brad 7 1.0295 Brad 7 1.1941 Brad 7 1.0759 Brad 7 1.3249
Brad 7 0.9494 Brad 7 1.1041 Brad 7 1.1575 Brad 7 1.294 Brad 7 1.4543
;
PROC SORT DATA=bradvsjanet;
BY tech rat;
PROC GLM DATA=bradvsjanet;
CLASS tech rat;
MODEL protein=tech rat(tech) / SS1;
TEST H=tech E=rat(tech) / HTYPE=1 ETYPE=1;
RUN;
```

The output includes F_{group} calculated two ways, as MS_{group}/MS_{within} and as $MS_{group}/MS_{subgroup}$.

Source DF Type I SS Mean Sq. F Value Pr > F

tech 1 0.03841046 0.03841046 1.07 0.3065 **MS_{group}/MS_{within}; don't use this**
 rat(tech) 4 0.57397543 0.14349386 3.98 0.0067 **use this for testing subgroups**

Tests of Hypotheses Using the Type I MS for rat(tech) as an Error Term

Source DF Type I SS Mean Sq. F Value Pr > F

tech 1 0.03841046 0.03841046 0.27 0.6322 **MS_{group}/MS_{subgroup}; use this for testing groups**

You can do the Tukey-Kramer test to compare pairs of group means, if you have more than two groups. You do this with a MEANS statement. This shows how (even though you wouldn't do Tukey-Kramer with just two groups):

```
PROC GLM DATA=bradvsjanet;
CLASS tech rat;
MODEL protein=tech rat(tech) / SS1;
TEST H=tech E=rat(tech) / HTYPE=1 ETYPE=1;
MEANS tech /LINES TUKEY;
RUN;
```

PROC GLM does not partition the variance. PROC NESTED will partition the variance, but it only does the hypothesis testing for a balanced nested anova, so if you have an unbalanced design you'll want to run both PROC GLM and PROC NESTED. In PROC NESTED, the group is given first in the CLASS statement, then the subgroup.

```
PROC SORT DATA=bradvsjanet;
BY tech rat;
PROC NESTED DATA=bradvsjanet;
CLASS tech rat;
VAR protein;
RUN;
```

Here's the output; if the data set was unbalanced, the "F Value" and "Pr>F" columns would be blank.

| | Variance | Sum of F | Error Mean | Variance Percent |
|--------|----------|----------|------------|--------------------|
| Source | DF | Squares | Value | Pr>F |
| | | Term | Square | Component of Total |

| | | | | | |
|----------|----------|----------|----------|----------------|------------------|
| Total | 59 | 2.558414 | 0.043363 | 0.046783 | 100.0000 |
| tech 1 | 0.038410 | 0.27 | 0.6322 | rat 0.038410 | -0.003503 0.0000 |
| rat 4 | 0.573975 | 3.98 | 0.0067 | Error 0.143494 | 0.010746 22.9690 |
| Error 54 | 1.946028 | 0.036038 | 0.036038 | 77.0310 | |

You set up a nested anova with three or more levels the same way, except the MODEL statement has more terms, and you specify a TEST statement for each level. Here's how you would set it up if there were multiple rats per technician, with multiple tissue samples per rat, and multiple protein measurements per sample:

```
PROC GLM DATA=bradvsjanet;
CLASS tech rat sample;
MODEL protein=tech rat(tech) sample(rat tech)/ SS1;
TEST H=tech E=rat(tech) / HTYPE=1 ETYPE=1;
TEST H=rat E=sample(rat tech) / HTYPE=1 ETYPE=1;
RUN;
PROC NESTED DATA=bradvsjanet;
CLASS sample tech rat;
VAR protein;
RUN;
```


References

Picture of a rat from [Posterwire](#).

Picture of lichen from Lichenological Society of Japan Lichen Photo Gallery

Keon, D.B., and P.S. Muir. 2002. Growth of *Usnea longissima* across a variety of habitats in the Oregon coast range. Bryologist 105: 233-242.

This page titled [4.9: Nested Anova](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [John H. McDonald](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.