

# How to Perform a Nested ANOVA in R: A Step-by-Step Guide

Authored by  
**stats writer**

December 6, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Perform a Nested ANOVA in R: A Step-by-Step Guide*.  
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106265>

The Nested ANOVA, sometimes referred to as a hierarchical ANOVA, is a specialized statistical methodology used when the levels of one categorical factor (the nested factor) are unique to the levels of a higher-order factor (the nesting factor). This design contrasts sharply with a traditional factorial ANOVA, where all factors cross, meaning every level of one factor occurs with every level of the other factor. The core application of the ANOVA technique remains the same: comparing the means of two or more groups of observations, but it is tailored specifically for data exhibiting this hierarchical structure.

Implementing a Nested ANOVA in R requires precise specification of the model formula to correctly account for the dependencies between factors. We primarily utilize the built-in `aov()` function to define the nested structure. Subsequently, functions like `summary()` or `Anova()` (often from specialized packages) are employed to extract the resulting F-statistics, degrees of freedom, and p-values necessary for drawing conclusions about the factors' effects on the response variable.

A key advantage of using R for this analysis is the flexibility it offers in defining complex experimental designs. While standard ANOVA models assess main effects and interactions, the nested structure is specified using a specialized notation (FactorA / FactorB), ensuring that the variability attributed to the nested factor is correctly partitioned only within the levels of the nesting factor. This robust approach ensures accurate hypothesis testing, particularly when dealing with potentially confounding sources of variation.

Understanding the concept of nesting is crucial before proceeding with the statistical implementation. A Nested ANOVA design features a factor where the levels of a secondary factor are unique within the levels of a primary factor. In essence, the secondary factor does not cross all levels of the primary factor; it only exists within its specific parent level.

Consider a practical research example: a study aiming to determine if three distinct types of fertilizer (A, B, and C) yield different levels of plant growth. To ensure efficiency and account for potential technician-to-technician variability, the experiment is structured hierarchically.

The experimental setup involves nine different technicians. Each fertilizer type is applied by three unique technicians, who are responsible for four plants each. Specifically:

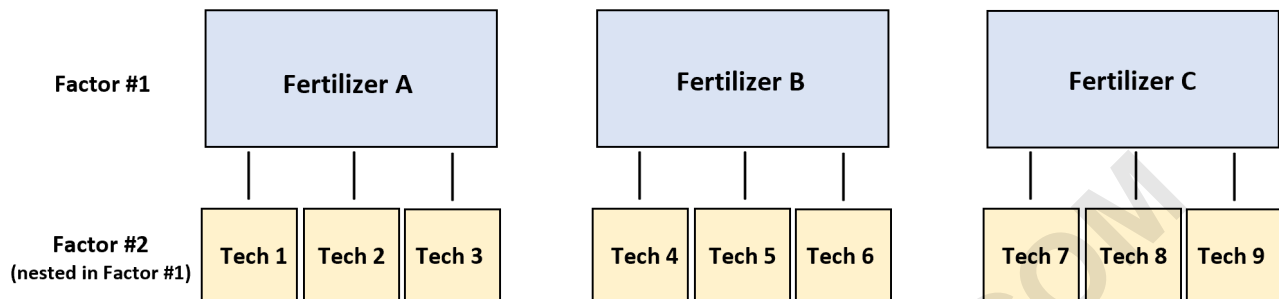
Fertilizer A is applied by Technicians 1, 2, and 3.

Fertilizer B is applied by Technicians 4, 5, and 6.

Fertilizer C is applied by Technicians 7, 8, and 9.

In this carefully constructed scenario, the measured outcome, or response variable, is **plant growth** and the two categorical factors under investigation are **fertilizer type** and **technician identity**. Crucially, the technician factor is **nested** within the fertilizer factor. Technician 1 only

handles Fertilizer A and never Fertilizer B or C; thus, the variability introduced by Technician 1 is entirely contained within the Fertilizer A group. This hierarchical relationship is fundamental to the nested design structure.



The following comprehensive, step-by-step tutorial demonstrates how to execute and interpret this specific Nested ANOVA model using the powerful statistical capabilities of R.

## Preparing the Data Structure in R

Before fitting any statistical model, it is essential to structure the data appropriately in a data frame within R. For a Nested ANOVA, we require columns for the response variable (growth) and for both the nesting factor (fertilizer) and the nested factor (technician). The total number of observations in this example is 36 (3 fertilizers \* 3 technicians per fertilizer \* 4 plants per technician).

We will create a data frame named `df`, ensuring that the levels of the technician variable are correctly assigned to their respective fertilizer groups, matching the experimental design described previously. Pay close attention to the way the categorical variables are constructed using the `rep()` function to ensure the correct hierarchical relationships are established.

The initial step involves using the `data.frame()` function to consolidate the raw measurements and group identifiers. This setup is crucial for the subsequent modeling step, as the `aov()` function relies on this organized structure.

**# Create the data frame containing plant growth measurements, fertilizer type, and technician ID.**

```
df <- data.frame(growth=c(13, 16, 16, 12, 15, 16, 19, 16, 15, 15, 12, 15,
19, 19, 20, 22, 23, 18, 16, 18, 19, 20, 21, 21,
21, 23, 24, 22, 25, 20, 20, 22, 24, 22, 25, 26),
fertilizer=c(rep(c('A', 'B', 'C'), each=12)),
tech=c(rep(1:9, each=4)))
```

# Verify the structure by viewing the first six rows of the data frame.

```
head(df)
```

```
growth fertilizer tech
```

```
1 13 A 1
```

```
2 16 A 1
```

```
3 16 A 1
```

```
4 12 A 1
```

```
5 15 A 2
```

```
6 16 A 2
```

## Specifying and Fitting the Nested ANOVA Model

The next critical step involves using the `aov()` function to fit the model. Unlike standard two-way ANOVA where interaction is denoted by a multiplication sign (\*) or a colon (:), a nested design uses the forward slash (/) notation to explicitly define the hierarchy.

The general syntax for fitting a Nested ANOVA in R is structured as follows, where the response variable is modeled as a function of Factor A, and Factor B is modeled only within Factor A:

```
aov(response ~ FactorA / FactorB)
```

Detailed explanation of the formula components:

**response:** Represents the continuous dependent variable being measured (e.g., plant growth).

**FactorA:** Is the primary, higher-level factor (the nesting factor, e.g., fertilizer type).

**FactorB:** Is the secondary factor whose levels are unique within each level of Factor A (the nested factor, e.g., technician identity).

For our specific dataset, we apply this syntax. Note that we wrap the nested factor, `df$tech`, within the `factor()` function to ensure R treats the technician IDs as distinct categorical levels rather than continuous numerical data. This is crucial for correct variance partitioning in ANOVA.

**# Fit the nested ANOVA model. The '/' indicates that the variation due to technician (tech) is nested within fertilizer.**

```
nest <- aov(df$growth ~ df$fertilizer / factor(df$tech))
```

```
# Display the results summary of the fitted nested ANOVA.
```

```
summary(nest)
```

```
Df Sum Sq Mean Sq F value Pr(>F)
```

```
df$fertilizer 2 372.7 186.33 53.238 4.27e-10 ***
```

```
df$fertilizer:factor(df$tech) 6 31.8 5.31 1.516 0.211
```

Residuals 27 94.5 3.50

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Interpreting the Analysis of Variance Output

The output generated by the `summary()` function provides the core results of the Nested ANOVA in the standard ANOVA table format. We must interpret the results based on the F-values and corresponding p-values ( $\Pr(>F)$ ). The output is partitioned into three key rows: the main nesting factor (`df$fertilizer`), the nested factor (`df$fertilizer:factor(df$tech)`), and the unexplained variation (`Residuals`).

Let's analyze the effects:

**Effect of Fertilizer (Nesting Factor):** The row labeled `df$fertilizer` shows a high F-value (53.238) and an extremely small p-value (4.27e-10, represented by '\*\*\*' significance code). Since this p-value is far below the conventional significance level of  $\alpha = 0.05$ , we conclude that the type of fertilizer used has a highly statistically significant effect on plant growth. This suggests that the mean growth across the three fertilizer types is not equal.

**Effect of Technician (Nested Factor):** The row labeled `df$fertilizer:factor(df$tech)` represents the variation due to technicians **within** each fertilizer group. The F-value here is 1.516, resulting in a p-value of 0.211. As 0.211 is greater than 0.05, we fail to reject the null hypothesis. This indicates that the differences in growth observed among the technicians applying the same fertilizer type are not statistically significant.

The practical implication of these findings is profound. If the goal is to maximize plant growth, the research strongly suggests that efforts should be concentrated on selecting the most effective fertilizer type. Conversely, there is no evidence to suggest that the identity or specific technique of the individual technician applying the fertilizer introduces significant unwanted variation into the results. This model successfully isolates the sources of variability, leading to clear, actionable conclusions.

## Post-Hoc Analysis Consideration

Although not strictly required by the non-significant technician effect, when a main factor, such as fertilizer, is found to be significant, researchers typically perform post-hoc tests to determine exactly which pairs of means differ from one another. Since the technician factor was not significant, we do not need to investigate differences between technicians.

For the fertilizer factor, had we required a comparison between A, B, and C, we would typically

employ Tukey's Honestly Significant Difference (HSD) test or similar methods. In R, this is often done using the `TukeyHSD()` function applied directly to the `aoV` object, providing adjusted p-values for all pairwise comparisons among the significant factor's levels. This crucial step provides granularity beyond the overall F-test result.

## Visualizing the Hierarchical Results with Boxplots

Statistical output, while precise, benefits greatly from visual confirmation. To further solidify our interpretation, we can use graphical tools like boxplots to visualize the distribution of plant growth, grouped first by fertilizer and then subdivided by technician within those groups. We rely on the sophisticated data visualization package, `ggplot2`, for this task.

The visualization strategy involves mapping the fertilizer type to the fill aesthetic, allowing us to see the primary differences, while mapping the individual technician (as a factor) to the X-axis. This arrangement clearly highlights the nesting: technicians are organized sequentially on the X-axis, but the coloring clearly delineates which technicians belong to which fertilizer group.

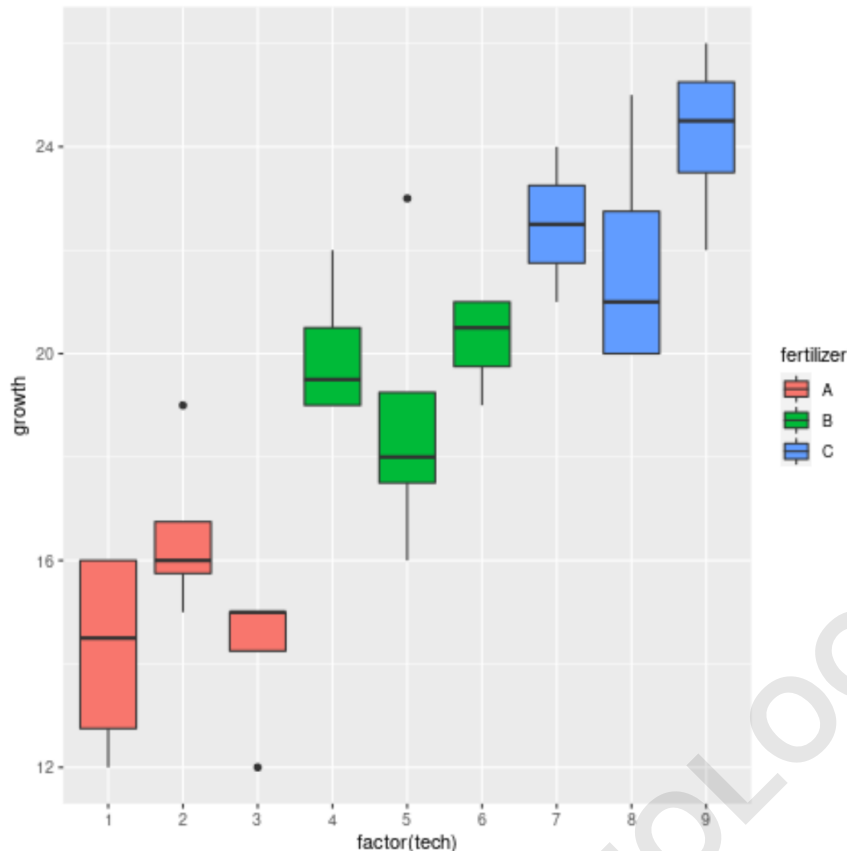
First, we must load the necessary package. Then, we construct the plot using `ggplot()`, specifying the aesthetics (`aes`) for X and Y axes and the fill color, followed by the `geom_boxplot()` layer to generate the final chart.

### # Load the essential ggplot2 data visualization package

```
library(ggplot2)
```

```
# Create boxplots to visualize plant growth across technicians, nested within fertilizer types.
```

```
ggplot(df, aes(x=factor(tech), y=growth, fill=fertilizer)) +  
geom_boxplot()
```



Examination of the resulting boxplot provides a clear qualitative confirmation of the ANOVA results. We observe substantial vertical separation and distinct differences in the median growth levels (the central lines of the boxplots) between the three major fertilizer groups (A, B, and C), indicated by the different colors. Fertilizer C clearly leads to higher overall growth compared to A and B.

Conversely, within each fertilizer group (e.g., technicians 1, 2, and 3 under Fertilizer A), the boxplots are generally positioned at similar vertical levels, indicating minimal variation among those technicians. This visual pattern aligns perfectly with the statistical finding that while fertilizer type significantly impacts plant growth, the variance attributable to the specific technician within those groups is not statistically significant.

## Summary of Nested ANOVA Methodology

The implementation of Nested ANOVA in R is a powerful tool for analyzing complex hierarchical data structures. By correctly using the aov() function with the appropriate nested syntax (FactorA / FactorB), researchers can accurately partition the total variance. This precision is vital for experiments where sampling units are grouped within larger treatment categories, such as patients within clinics, students within classrooms, or, as demonstrated here, technicians within fertilizer treatments.

This procedure ensures that the effects are tested against the appropriate error term. For instance, the main effect of fertilizer is tested against the variability among technicians within fertilizers, not against the residual error (which represents plant-to-plant variation). This adherence to the correct statistical testing protocol prevents Type I errors and ensures the reliability of the conclusions drawn from the study. Mastering this technique allows for advanced statistical inference tailored to hierarchical experimental designs.

## Concluding Remarks

The Nested ANOVA provides a robust framework for assessing treatment effects while controlling for variation introduced by nested factors. Our analysis confirmed that fertilizer type is the dominant factor influencing plant growth, while technician variability is negligible in this experimental design. This outcome guides researchers toward focusing resources on optimizing the primary treatment factor, demonstrating the practical value of correctly specifying and executing hierarchical statistical models in R.