

# Scheffe's Test in R

Authored by  
**stats writer**

December 17, 2025

## RECOMMENDED CITATION

stats writer (2025). *Scheffe's Test in R*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=107712>

The Scheffe's Test is a powerful and robust post-hoc test essential for modern statistical analysis, particularly when dealing with multiple group comparisons following a significant result from an Analysis of Variance (ANOVA). It is specifically designed for conducting statistical inference regarding the differences among population means, helping researchers pinpoint exactly which pairs or combinations of groups differ significantly from one another.

Unlike some other post-hoc procedures, Scheffe's method is considered highly conservative, meaning it is excellent at controlling the family-wise error rate (FWER) across all possible comparisons, not just pairwise ones. The methodology is mathematically rigorous, basing its critical value on the F-distribution and incorporating linear combinations of sample means. This tutorial provides a comprehensive guide on implementing and interpreting the Scheffe's Test within the statistical environment of R, utilizing established packages for clean, reproducible results.

## The Context: Why Post-Hoc Tests Are Necessary

Before diving into the specifics of Scheffe's procedure, it is vital to understand the initial statistical hurdle: the one-way ANOVA. This foundational test determines if there is a statistically significant difference among the means of three or more independent groups. The primary null hypothesis tested by ANOVA is that all group population means are equal. If the ANOVA returns a statistically significant result--indicated by a small p-value, typically below the alpha level of 0.05--we reject the null hypothesis.

However, the limitation of the ANOVA is that it is an omnibus test. While it tells us that a difference exists somewhere among the groups, it provides no information on the exact location of that difference. We know that at least one group mean is distinct from the others, but we don't know whether Group A differs from Group B, Group A differs from Group C, or if more complex relationships exist. This ambiguity necessitates a secondary, targeted analysis.

If we were to simply run multiple independent T-tests comparing every pair of groups following a significant ANOVA result, we would inflate our overall Type I error rate. Each T-test carries an individual risk of falsely rejecting the null hypothesis (alpha), and performing many tests multiplies this risk, leading to an unacceptable family-wise error rate. Therefore, to maintain statistical integrity and confidence in our findings, we must employ a specialized post-hoc test that correctly adjusts for multiple comparisons.

## Defining Scheffe's Test and Its Key Advantages

The Scheffe's Test, developed by Henry Scheffé, is renowned for being one of the most flexible and conservative post-hoc methods available. It is particularly advantageous because it can be used to test not only all pairwise comparisons (e.g., A vs. B, B vs. C) but also any and all possible linear combinations or "contrasts" of the group means (e.g.,  $(A+B)/2$  vs. C). This flexibility makes it

ideal for exploratory research where complex hypotheses about group means might emerge after the data has been collected.

The test's high degree of conservatism stems from its stringent control of the FWER. By using the F-distribution critical value derived from the initial ANOVA, the Scheffe procedure ensures that the probability of making one or more Type I errors across the entire set of comparisons remains equal to or less than the chosen alpha level. Consequently, while it is robust and safe, it often possesses less statistical power for simple pairwise comparisons compared to slightly less conservative tests like Tukey's Honestly Significant Difference (HSD) test. Researchers typically select Scheffe's test when complex comparisons or a rigorous safeguard against error are paramount.

## Example Scenario: Evaluating Studying Techniques

To illustrate the application of Scheffe's Test in R, consider a practical scenario in educational research. Suppose a teacher is interested in determining whether three distinct studying techniques--Tech 1, Tech 2, and Tech 3--result in different mean exam scores among students. The teacher sets up an experiment, randomly assigning 10 students to use each technique, resulting in 30 total observations, and records their final exam scores. The goal is first to establish if any difference exists using a one-way ANOVA, and then, if the ANOVA is significant, to use Scheffe's Test to pinpoint the exact difference(s) between the techniques.

We will follow a structured four-step process within the R environment: dataset creation, data visualization, ANOVA execution, and finally, the application of Scheffe's Test. This example demonstrates how to transition seamlessly from a global test of difference to specific, protected post-hoc comparisons, ensuring the integrity of the statistical conclusions drawn from the study.

### Step 1: Creating the Dataset in R

The first crucial step in any analysis is setting up the data correctly. We need to structure the exam scores along with their corresponding studying technique groups. In R, this is best accomplished by creating a data frame where one column identifies the group (the independent variable, technique) and the other holds the continuous outcome measure (the dependent variable, score). The following code block demonstrates how to generate the sample dataset containing 30 observations, 10 for each technique, mimicking the results of the teacher's experiment.

The generated data frame is essential for the subsequent ANOVA and Scheffe's Test, as R functions rely on this tidy structure to correctly partition the variance and calculate the necessary statistics. Reviewing the initial rows helps confirm the data structure and variable types before proceeding.

**#create data frame**

```
data <- data.frame(technique = rep(c("tech1", "tech2", "tech3"), each = 10),
score = c(76, 77, 77, 81, 82, 82, 83, 84, 85, 89,
81, 82, 83, 83, 83, 84, 87, 90, 92, 93,
77, 78, 79, 88, 89, 90, 91, 95, 95, 98))
```

```
#view first six rows of data frame
head(data)
```

```
technique score
```

```
1 tech1 76
```

```
2 tech1 77
```

```
3 tech1 77
```

```
4 tech1 81
```

```
5 tech1 82
```

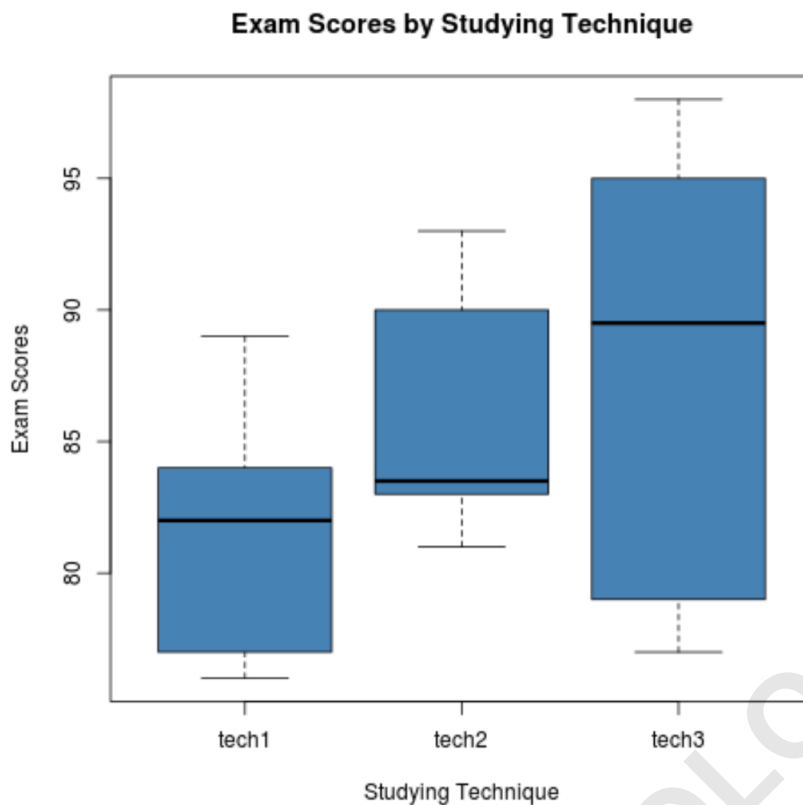
```
6 tech1 82
```

## Step 2: Visualizing Data Distributions with Boxplots

A crucial preliminary step in statistical analysis, often preceding the formal ANOVA calculation, is data visualization. Visualizing the data provides immediate insights into the distribution of scores within each group, potential differences in means, and the degree of variability. For comparing distributions across categorical groups, a boxplot is an ideal tool, clearly displaying the median, quartiles, and any outliers for each studying technique.

The following R code uses the base plotting system to generate boxplots comparing the `score` across the three levels of `technique`. This visualization can help us anticipate the results of the ANOVA; if the boxes (representing the interquartile range) show minimal overlap and distinct medians, it suggests that statistical differences are likely to be found. Conversely, substantial overlap may indicate minimal or no significant difference between the means.

```
boxplot(score ~ technique,
data = data,
main = "Exam Scores by Studying Technique",
xlab = "Studying Technique",
ylab = "Exam Scores",
col = "steelblue",
border = "black")
```



### Step 3: Performing the One-Way ANOVA

With the data prepared and visualized, we proceed to the formal test of overall difference: the one-way ANOVA. In R, this is conventionally achieved using the `aov()` function, which fits the ANOVA model to the data, followed by the `summary()` function to generate the ANOVA table. This table summarizes the partitioning of variance into components explained by the group differences (`technique`) and the unexplained residual variance.

The primary result of interest in the output is the F value and its corresponding p-value. If the p-value falls below our predetermined significance threshold (commonly  $\alpha = 0.05$ ), we conclude that the studying techniques, as a whole, have a statistically significant effect on exam scores. This significant finding is the necessary precursor to justify moving on to the post-hoc analysis using Scheffe's Test.

#### #fit the one-way ANOVA model

```
model <- aov(score ~ technique, data = data)
```

```
#view model output
```

```
summary(model)
```

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

```

technique 2 211.5 105.73 3.415 0.0476 *
Residuals 27 836.0 30.96
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The analysis shows an overall p-value of **0.0476**. Since this value is less than the typical significance level of 0.05, we reject the null hypothesis and confirm that there is a statistically significant difference in the average exam scores among the three studying techniques. The crucial task now is to identify precisely where these differences lie, which mandates the use of Scheffe's Test.

#### Step 4: Executing Scheffe's Test in R using DescTools

To execute Scheffe's Test, we need an appropriate R package capable of performing multiple comparisons while implementing the Scheffe adjustment. While several packages exist, we will leverage the DescTools package, which provides the highly convenient ScheffeTest() function. This function takes the ANOVA model object (created in Step 3) as its input and automatically calculates the adjusted p-values for all pairwise comparisons.

It is important to ensure the package is installed and loaded into the R session before attempting to call the function. The ScheffeTest() function calculates the mean difference (diff), the lower and upper bounds of the confidence interval (lwr.ci and upr.ci), and the adjusted p-value (pval) for every pairwise contrast. Crucially, the confidence level is set for the 95% family-wise protection, guaranteeing that the overall probability of a Type I error remains controlled.

##### #load DescTools package

```
library(DescTools)
```

```
#perform Scheffe's test
```

```
ScheffeTest(model)
```

```
Posthoc multiple comparisons of means : Scheffe Test
```

```
95% family-wise confidence level
```

```
$technique
```

```
diff lwr.ci upr.ci pval
```

```
tech2-tech1 4.2 -2.24527202 10.645272 0.2582
```

```
tech3-tech1 6.4 -0.04527202 12.845272 0.0519 .
```

```
tech3-tech2 2.2 -4.24527202 8.645272 0.6803
```

```
---
```

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

## Interpreting the Results of Scheffe's Test

The output provides a clear, tabulated summary of all possible pairwise comparisons. Interpreting these results requires careful attention to the `diff` column, which shows the mean difference between the groups, and the `pval` column, which contains the Scheffe-adjusted p-value for that specific comparison. A comparison is considered statistically significant if its adjusted p-value is less than the predetermined alpha level (e.g., 0.05).

We analyze the output line by line:

The mean difference in exam scores between technique 2 and technique 1 is **4.2** (Tech 2 scored 4.2 points higher than Tech 1). The corresponding adjusted p-value is **0.2582**. Since 0.2582 is much greater than 0.05, we conclude there is no statistically significant difference between Studying Technique 2 and Technique 1.

The mean difference in exam scores between technique 3 and technique 1 is **6.4** (Tech 3 scored 6.4 points higher than Tech 1). The adjusted p-value is **0.0519**. This value is just marginally above the conventional 0.05 cutoff, indicated by the dot (.) in the significance codes. If a researcher were to use a slightly more relaxed alpha (e.g., 0.10) or consider this result highly suggestive, they might interpret this as an approaching significance. Based strictly on alpha = 0.05, this difference is not significant.

The mean difference in exam scores between technique 3 and technique 2 is **2.2** (Tech 3 scored 2.2 points higher than Tech 2). The adjusted p-value is **0.6803**. This is clearly not statistically significant, indicating no reliable difference between Technique 3 and Technique 2.

Overall, based on a standard significance level of alpha = 0.05, the Scheffe's Test indicates that while the overall ANOVA was significant, none of the pairwise differences met the strict threshold for statistical significance after the conservative Scheffe adjustment. The only comparison approaching significance was Technique 3 versus Technique 1. This scenario highlights the conservative nature of Scheffe's test; it may fail to detect differences detected by less conservative procedures (like Tukey's HSD) when sample sizes are moderate or when the overall F-ratio is close to the critical value, reinforcing the protection against inflating the family-wise error rate.

## Conclusion and Further Exploration

The Scheffe's Test is an indispensable tool in the statistical toolkit when conducting post-hoc analysis, especially when complex contrasts or rigorous error control are required following a significant ANOVA result. Its implementation in R via packages like `DescTools` is straightforward,

providing comprehensive output for mean comparisons and adjusted significance levels. Understanding the context of FWER control and the distinction between omnibus and specific post-hoc tests is essential for drawing accurate conclusions from experimental data.

For researchers needing alternatives or further statistical exploration in R, the following resources detail related procedures:

### **Related Statistical Resources in R**

[How to Conduct a One-Way ANOVA in R](#)

[How to Perform Tukey's Test in R](#)

[How to Perform a Bonferroni Correction in R](#)

ARABPSYCHOLOGY.COM