

Learn How to Perform Scheffe's Post-Hoc Test in R: A Step-by-Step Guide

Authored by
Mohammed looti

November 6, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Perform Scheffe's Post-Hoc Test in R: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11644>

The Foundation: Understanding ANOVA and Post-Hoc Testing

The [one-way ANOVA](#) (Analysis of Variance) represents a fundamental procedure in statistical inference, meticulously designed to determine if statistically significant differences exist among the mean values of three or more independent groups. This test serves as the crucial initial gateway, efficiently assessing all population means simultaneously within a single inferential framework.

When the overall [p-value](#) derived from the ANOVA summary falls below the predetermined significance threshold (commonly established at 0.05), we possess sufficient statistical evidence to confidently reject the null hypothesis. This rejection signifies that at least one group mean is distinct from the others. However, the ANOVA procedure is inherently limited; it only indicates the presence of a difference but fails to specify precisely **which** specific pairs of groups are the source of this detected variation.

To accurately localize these differences between specific combinations of group pairs, researchers must subsequently employ a specialized procedure known as a [post-hoc test](#). A critical requirement for any reliable multiple comparison procedure is the robust capability to control the overall cumulative probability of making a Type I error across all comparisons, formally referred to as the [family-wise error rate](#) (FWER).

Among the comprehensive suite of available methods for multiple comparison correction, [Scheffe's test](#) is highly regarded. It is recognized as one of the most powerful and notably conservative approaches. This detailed tutorial provides a thorough, step-by-step methodology on how to correctly implement, execute, and interpret **Scheffe's test** within the versatile statistical programming environment of [R](#).

Case Study: Investigating Study Technique Efficacy

To illustrate the practical application of Scheffe's method, consider a focused case study involving an educator. This teacher aims to rigorously investigate the differential efficacy of three distinct studying techniques on the final exam performance of students. For experimental control, the teacher randomly allocates 10 students to utilize each technique over a specified period, meticulously recording their final exam scores. This design results in three clearly defined, independent groups available for subsequent statistical comparison.

Our analytical objective is structured in two sequential phases: Initially, we must fit a one-way ANOVA model to establish whether any overall significant difference exists between the three techniques. If this initial test yields a significant result, the second phase involves applying **Scheffe's test**. This post-hoc analysis will systematically identify the specific pairs of techniques that produced statistically different mean exam scores, all while maintaining rigorous control over the risk of inflated Type I errors due to multiple comparisons.

Step 1: Preparing and Structuring the Sample Data in R

The commencement of any statistical analysis in R necessitates the appropriate structuring of the raw data. For this case study, we will construct a data frame containing 30 total observations. This structure maps each student's numerical exam `score` to their assigned categorical variable, `technique`. The assignment is balanced, with 10 observations per technique group.

The following R commands are used to efficiently establish the required data frame. Subsequent viewing of the initial rows confirms the successful data input and verifies the foundational structure necessary for the modeling process:

```
#create data frame for 3 techniques, 10 students each
```

```
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 to verify input
```

```
head(data)
```

```
technique score
```

```
1 tech1 76
```

```
2 tech1 77
```

```
3 tech1 77
```

```
4 tech1 81
```

```
5 tech1 82
```

```
6 tech1 82
```

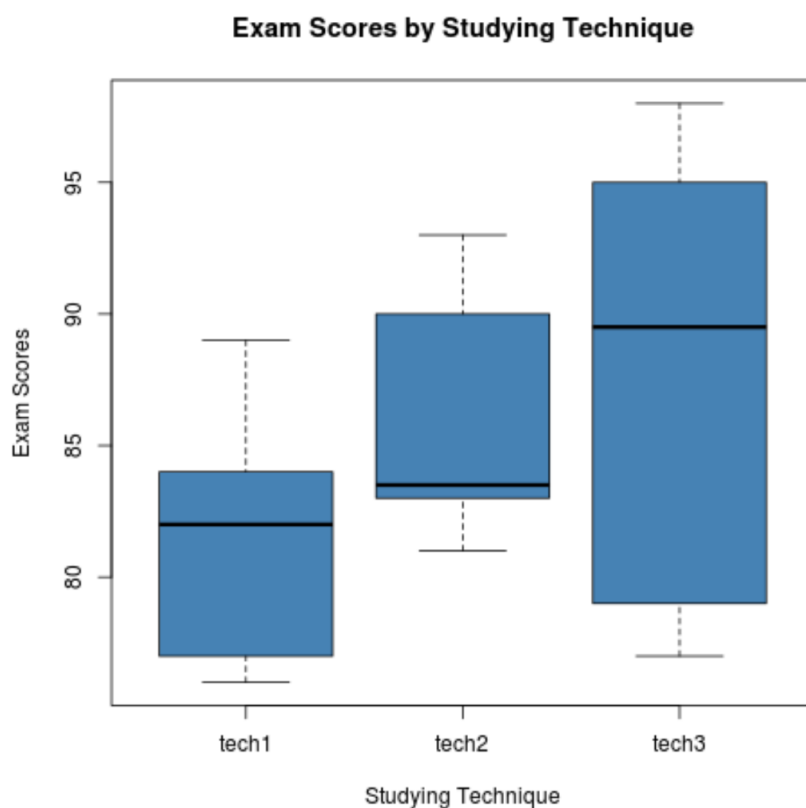
Step 2: Visual Confirmation of Group Distributions

Before proceeding with formal hypothesis testing, it is considered best statistical practice to conduct a thorough visual inspection of the data. We employ boxplots, which provide an immediate and effective means to compare the distribution characteristics, spread, and central tendency of the exam scores across the three distinct studying techniques.

This critical visualization step offers an intuitive understanding of the data's inherent structure. Furthermore, examining the boxplots helps analysts check for potential violations of core ANOVA assumptions, such as severe lack of symmetry (skewness) or the presence of influential outliers that could distort the model results. The R code below generates the comparative boxplot:

```
boxplot(score ~ technique,
```

```
data = data,  
main = "Exam Scores by Studying Technique",  
xlab = "Studying Technique",  
ylab = "Exam Scores",  
col = "steelblue",  
border = "black")
```



Step 3: Executing the Preliminary One-Way ANOVA

The next analytical phase involves fitting the one-way ANOVA statistical model using R's native `aov()` function. This test is specifically designed to assess the null hypothesis, which posits that the mean exam scores for all three techniques are statistically equivalent. A significant outcome from this initial test is absolutely essential, as it validates the necessity and utility of proceeding with a more granular post-hoc analysis.

The resulting summary output of the **Analysis of Variance** provides crucial metrics, including the degrees of freedom (Df), the sums of squares (Sum Sq), the mean squares (Mean Sq), the calculated F-statistic, and most importantly, the corresponding overall [p-value](#). This comprehensive summary allows for a direct assessment of the model's fit:

```
#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
```

Upon reviewing the output, we observe that the overall [p-value](#) quantifying the effect of the `technique` variable is calculated as **0.0476**. Given that this value is only marginally below the established conventional alpha level of 0.05, we conclude that a statistically significant difference in mean exam scores exists somewhere among the three studying techniques. This marginal but significant finding provides the necessary justification to proceed directly to a specific post-hoc procedure to precisely identify the locus of this overall effect.

Step 4: Implementing the Conservative Scheffe's Test

Because the preliminary ANOVA indicated a significant overall effect, we now move to implement [Scheffe's test](#). This test is particularly recommended in situations involving planned or unplanned complex comparisons, or when the assumption of equal sample sizes might be violated, due to its exceptional strength in controlling the [family-wise error rate](#) (FWER).

To successfully execute Scheffe's test within R, it is necessary to utilize the powerful `ScheffeTest()` function, which is conveniently packaged within the widely used [DescTools](#) library. The following sequence of R commands demonstrates how to load the required library and subsequently apply the test directly to the previously fitted ANOVA model:

```
#load DescTools package
```

```
library(DescTools)
```

```
#perform Scheffe's test using the previously fitted ANOVA model
```

```
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
```

Detailed Interpretation of Pairwise Comparisons

The generated output provides a clear, structured summary of all possible pairwise comparisons between the three techniques. For each pair, the output displays the calculated mean difference (`diff`), the 95% family-wise confidence interval, and the crucial adjusted [p-value](#). Since Scheffe's procedure is inherently conservative, these p-values have been strictly adjusted to ensure the overall FWER remains controlled at the specified 5% level.

The comparison between **Technique 2 and Technique 1** shows a mean score difference of 4.2 points. The corresponding adjusted p-value is calculated as **0.2582**. Since this value significantly exceeds the 0.05 criterion, we must conclude that there is no statistically significant difference between the average performance achieved using these two techniques.

The comparison between **Technique 3 and Technique 1** reveals the largest numerical mean difference, recorded at 6.4 points. The associated adjusted p-value is **0.0519**. This result is notable as it falls extremely close to, but critically fails to cross, the stringent 0.05 significance threshold. This near-significance is visually indicated by the marginal dot (.) in the significance codes column.

Finally, the comparison between **Technique 3 and Technique 2** exhibits the smallest mean difference of 2.2 points, accompanied by a high adjusted p-value of **0.6803**. This outcome decisively indicates that there is no statistical difference between the mean scores of these two specific groups.

In conclusion, when applying the strict $\alpha = 0.05$ cutoff, [Scheffe's test](#), due to its conservative nature, ultimately fails to identify any single pair of groups as being statistically significantly different. This result is frequently observed when the overall ANOVA result is marginal ($p=0.0476$), as Scheffe's method requires a substantially larger effect or difference to surpass its heightened significance threshold compared to alternative, less conservative procedures like Tukey's Honestly Significant Difference (HSD) test.

Further Statistical Resources

To further enhance your mastery of statistical comparison methods and their robust

implementation within the R environment, we recommend exploring the following related tutorials covering similar post-hoc and ANOVA procedures:

[Detailed Guide on Conducting a One-Way ANOVA in R](#)

[Implementation of Tukey's HSD Test in R](#)

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