

How to Easily Perform Post-Hoc Pairwise Comparisons in R

Authored by
stats writer

December 2, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform Post-Hoc Pairwise Comparisons in R*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103562>

Performing post-hoc pairwise comparisons in the R statistical environment is a critical step following a significant omnibus test, such as an Analysis of Variance (ANOVA). While R offers the foundational built-in function, `pairwise.t.test()`, for these comparisons, researchers often utilize specialized functions or packages to apply specific correction methods like Tukey, Scheffe, Bonferroni, or Holm.

The `pairwise.t.test()` function is versatile; it accepts the dataset as input and systematically compares every pair of levels within a specified categorical variable. For each comparison, it returns a corrected p-value, which is essential for determining the statistical significance of the difference observed between those specific pairs. Key arguments often tailored in this function include the choice of the alternative hypothesis, the required confidence level, and, most importantly, the method used for p-value adjustment, which controls the family-wise error rate.

Understanding the Need for Post-Hoc Analysis

An ANOVA is fundamentally designed to assess whether or not there is a statistically significant difference occurring anywhere among the means of three or more independent groups. It acts as an initial screen, informing us if the group means are heterogeneous. However, it does not specify which particular pairs of groups differ from one another; it only confirms that the overall model is significant.

For instance, if an ANOVA yields a significant result, rejecting the null hypothesis, we know that at least one group mean is different from the others. To precisely locate where these differences lie--that is, whether Group A differs from Group B, Group B from Group C, and so on--we must employ post-hoc pairwise comparisons. These subsequent tests are designed to mitigate the increased risk of Type I errors (false positives) that arises from conducting multiple comparisons on the same dataset.

The choice of a specific post-hoc test, whether it is Tukey, Scheffe, or a Bonferroni-type correction, depends heavily on the research design, specifically whether the comparisons were planned before the data collection (a priori) or decided upon only after seeing the data (post-hoc), and whether the group sample sizes are equal.

The Role of One-Way ANOVA

A one-way ANOVA is the standard methodology when investigating the effect of a single categorical independent variable (with multiple levels) on a continuous dependent variable. The procedure partitions the total variance observed in the data into components attributable to differences between the groups and components attributable to error (within the groups). The core hypotheses tested by this method are clearly defined:

H0: All group means are equal ($\mu_1 = \mu_2 = \mu_3 = \dots$).

HA: Not all group means are equal (at least one mean differs).

If the overall p-value derived from the F-statistic of the ANOVA model is less than the predetermined significance level (commonly $\alpha = .05$), we reject the null hypothesis. This rejection signals that the independent variable significantly influences the dependent variable. However, rejecting the null hypothesis is only the first step; we must then perform post-hoc pairwise comparisons to determine which specific techniques or treatments caused the significant difference.

R Setup and Initial ANOVA Execution Example

Consider a practical scenario where a teacher is interested in evaluating the effectiveness of three distinct studying techniques on student exam scores. The goal is to determine if the mean scores differ significantly across these techniques. Thirty students are randomly assigned, 10 to each technique, and their subsequent exam scores are recorded. This design perfectly aligns with a one-way ANOVA test.

To analyze this data in R, we first structure the data into a data frame and then apply the `aov()` function (Analysis of Variance). This standard implementation allows us to quickly assess the overall effect of the independent variable (technique) on the dependent variable (score). The following code block demonstrates the setup and the initial ANOVA calculation:

```
#create data frame
df <- 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))

#perform one-way ANOVA
model <- aov(score ~ technique, data = df)

#view output of ANOVA
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
```

Interpreting the ANOVA Results

Upon reviewing the ANOVA output above, we focus on the p-value associated with the 'technique' factor, listed under `Pr(>F)`. In this example, the p-value is 0.0476. Since 0.0476 is less than the conventional significance threshold of $\alpha = .05$, we are compelled to reject the null hypothesis. This critical finding confirms that there is a statistically significant difference in the mean exam scores across the three studying techniques.

However, simply knowing that a difference exists is insufficient for drawing practical conclusions. We still need to identify which specific technique pairs (tech1 vs. tech2, tech1 vs. tech3, tech2 vs. tech3) are driving this overall significance. Therefore, having established the significance of the omnibus test, the next logical and necessary step is to apply specific post-hoc pairwise comparisons to pinpoint the source of the variance.

The Tukey Method: Comparison for Equal Samples

The Tukey Honestly Significant Difference (HSD) method is one of the most widely used post-hoc tests, particularly appropriate when two conditions are met: first, all possible pairwise comparisons are being made, and second, the sample sizes across all groups are equal (which is the case in our example, where $n=10$ for each technique). The Tukey method controls the family-wise error rate, ensuring the probability of making at least one Type I error across all comparisons remains below the specified alpha level.

In R, the Tukey HSD procedure is conveniently accessed using the built-in `TukeyHSD()` function, applied directly to the ANOVA model object. This function outputs the mean difference (`diff`), the lower and upper bounds of the confidence interval (`lwr` and `upr`), and the adjusted p-value (`p adj`) for every possible pair:

#perform the Tukey post-hoc method

TukeyHSD(model, conf.level=.95)

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = score ~ technique, data = df)

\$technique

diff lwr upr p adj

tech2-tech1 4.2 -1.9700112 10.370011 0.2281369

tech3-tech1 6.4 0.2299888 12.570011 0.0409017

tech3-tech2 2.2 -3.9700112 8.370011 0.6547756

Based on the adjusted p-values (p_{adj}), we can interpret the results. The comparison between technique 3 and technique 1 yields a p-value of 0.0409017. Since this value is less than 0.05, we conclude that there is a statistically significant difference in mean exam scores between students who used technique 1 and those who used technique 3. Both other comparisons (tech2-tech1 and tech3-tech2) fail to reach statistical significance under the Tukey correction.

The Scheffe Method: The Conservative Approach

The Scheffe method is recognized as the most conservative post-hoc pairwise comparison technique. It is particularly robust and flexible, as it can be used for any number of comparisons, including complex contrasts, not just simple pairwise ones, and it performs well even with unequal sample sizes. Because of its inherent conservatism, it typically produces the widest confidence intervals compared to other methods, making it harder to achieve statistical significance.

To implement the Scheffe method in R, we must utilize external packages, such as the `DescTools` package, which provides the necessary `ScheffeTest()` function. It is important to load the required library before execution, as shown below:

library(DescTools)

```
#perform the Scheffe post-hoc method
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
```

When analyzing the Scheffe results, we observe that the smallest p-value is 0.0519 (for tech3-tech1). Since none of the p-values in this output are strictly less than the alpha level of 0.05, the conclusion drawn using the Scheffe test is that there is no statistically significant difference in mean exam scores among any of the groups. This result highlights the conservative nature of the Scheffe method compared to the Tukey test, which previously identified one significant difference.

Planned Comparisons: Bonferroni and Holm Adjustments

Unlike the Tukey or Scheffe methods, which are designed for all possible (unplanned) comparisons, the Bonferroni correction is best suited for situations where the researcher has a specific, limited set of pairwise comparisons planned before conducting the experiment. The Bonferroni method controls the family-wise error rate by dividing the original alpha level (α) by the total number of comparisons (m). While simple to calculate, it often proves overly conservative, leading to a loss of statistical power--that is, a greater risk of committing a Type II error (false negative).

The Holm method (also known as the Holm-Bonferroni method) is an improvement upon the traditional Bonferroni approach. It is also designed for pre-planned comparisons but uses a sequentially rejective procedure. By ordering the p-values from smallest to largest and adjusting the critical alpha level step-by-step, the Holm method ensures that the family-wise error rate is still controlled while offering significantly higher power than the standard Bonferroni correction. Consequently, the Holm method is generally preferred when controlling for Type I errors in planned comparisons.

Executing the Bonferroni Correction in R

In R, we can efficiently apply the Bonferroni correction using the versatile `pairwise.t.test()` function, specifying the adjustment method via the `p.adj` argument. This function conducts standard t-tests for each pair and then modifies the resulting p-values according to the chosen technique. We input the scores, the grouping variable (technique), and set `p.adj='bonferroni'`:

```
#perform the Bonferroni post-hoc method  
pairwise.t.test(df$score, df$technique, p.adj='bonferroni')
```

Pairwise comparisons using t tests with pooled SD

data: df\$score and df\$technique

```
tech1 tech2  
tech2 0.309 -  
tech3 0.048 1.000
```

P value adjustment method: bonferroni

The resulting matrix displays the adjusted p-values for all pairwise comparisons. Analyzing this output, the only p-value that falls below the 0.05 threshold is the comparison between technique 1 and technique 3 ($p = 0.048$). Therefore, under the Bonferroni method, we reach the same

conclusion as the Tukey test: only the difference in mean exam scores between students who used technique 1 and those who used technique 3 is deemed statistically significant.

The Power of the Holm Method

The Holm method, due to its sequential nature, generally offers greater statistical power while maintaining robust control over the family-wise error rate, making it a preferred alternative to the traditional Bonferroni correction for planned comparisons. If a researcher suspects that the Bonferroni method might be too restrictive, the Holm adjustment provides a more lenient, yet statistically sound, approach to multiple testing.

Implementing the Holm correction in R is identical in syntax to the Bonferroni execution, except for changing the `p.adj` argument to `'holm'`. This tells the `pairwise.t.test()` function to apply the sequential Holm adjustment procedure:

```
#perform the Holm post-hoc method  
pairwise.t.test(df$score, df$technique, p.adj='holm')
```

Pairwise comparisons using t tests with pooled SD

data: df\$score and df\$technique

```
tech1 tech2  
tech2 0.206 -  
tech3 0.048 0.384
```

P value adjustment method: holm

Reviewing the output for the Holm method reveals that the p-values for the non-significant comparisons (tech2-tech1: 0.206; tech3-tech2: 0.384) are slightly smaller than those produced by the Bonferroni method, reflecting the Holm method's higher power. Critically, the p-value for tech3-tech1 remains 0.048, reaffirming the conclusion derived from the other less conservative tests: there is a statistically significant difference only between technique 1 and technique 3.

Summary of Post-Hoc Comparison Methods

Selecting the appropriate post-hoc test is vital for accurate interpretation of experimental results, particularly after a significant ANOVA. The choice hinges on whether all possible comparisons are of interest, whether the comparisons were planned, and the tolerance for Type I versus Type II error risk. Researchers must carefully weigh the balance between maximizing power and rigorously controlling the family-wise error rate when choosing among these established statistical

techniques.

The following list summarizes the primary considerations for choosing between the methods detailed above:

The Tukey Method: Ideal for all possible pairwise comparisons when sample sizes are equal, offering a good balance of power and error control.

The Scheffe Method: The most conservative choice; suitable for complex contrasts and when sample sizes are unequal, resulting in fewer significant findings.

The Bonferroni Method: Best for a small number of planned comparisons, though it is often overly conservative.

The Holm Method: Preferred over Bonferroni for planned comparisons, as it retains better power while still controlling the family-wise error rate effectively.

Further information regarding ANOVA procedures and the detailed mathematical workings of these post-hoc adjustments can be found in standard statistical textbooks and the official documentation for the relevant R packages.