

Learning Tukey's Honest Significant Difference (HSD) Test for ANOVA in R

Authored by
Mohammed looti

November 7, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learning Tukey's Honest Significant Difference (HSD) Test for ANOVA in R*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=12282>

The [Analysis of Variance \(ANOVA\)](#), particularly the one-way design, stands as a fundamental statistical procedure in quantitative research. Its primary purpose is to ascertain whether statistically significant differences exist among the mean values of three or more independent groups. Conceptually, the ANOVA serves as an omnibus test, providing a critical initial assessment of group heterogeneity.

When researchers execute an ANOVA, the core output is the overall F-statistic and its corresponding [p-value](#). If this p-value falls below the predetermined level of significance (typically $\alpha = 0.05$), the researcher is justified in rejecting the null hypothesis of equal means. This rejection confirms a crucial finding: there is evidence that at least one group mean differs significantly from the others. However, the ANOVA result is inherently limited; it signals the presence of a difference but fails to identify precisely **where** that difference lies among the specific pairs of groups.

To move beyond the general confirmation of variance and pinpoint the exact group comparisons that are statistically distinct, a subsequent procedure known as a [post hoc test](#) is mandatory. Without this step, any conclusions about specific group performance would be purely speculative.

Tukey's HSD Test: Controlling the Family-Wise Error Rate

Among the available post hoc methods, [Tukey's Honest Significant Difference \(HSD\) Test](#) is one of the most widely respected and statistically robust choices. Its primary strength lies in its design for performing **all possible pairwise comparisons** between group means while maintaining rigorous control over the critical statistical challenge known as the [family-wise error rate \(FWER\)](#).

Controlling the FWER is essential because performing numerous individual t-tests dramatically inflates the probability of committing a Type I error (falsely rejecting a true null hypothesis). Tukey's HSD mitigates this risk by basing its calculations on the [studentized range distribution](#), which adjusts the critical value necessary to declare a difference significant across the entire set of comparisons.

The goal of this comprehensive tutorial is to provide a step-by-step guide to executing [Tukey's Test](#) within the powerful [R statistical environment](#). We will cover proper model fitting, function execution, and the correct interpretation of the resulting statistical output.

Important Note: The choice of post hoc test should align with your experimental design. If your study includes a designated baseline or **control group**, and you only wish to compare each treatment group against that specific control, Dunnett's Test is statistically more appropriate than Tukey's HSD, as it is optimized for that specific comparison structure.

Prerequisite: Confirming Group Heterogeneity

Tukey's HSD test is fundamentally dependent on the outcome of the initial ANOVA. It is only justifiable to proceed with Tukey's HSD after a significant result has been obtained from the ANOVA F-test, thereby confirming that heterogeneity exists among the underlying group means.

The term "Honest Significant Difference" originates from the test's methodology: it determines the minimum difference between group means that must be observed to declare a pair statistically distinct while maintaining the overall desired alpha level. This rigorous threshold prevents spurious findings that often plague uncorrected multiple comparison procedures.

For the purpose of demonstration, we will simulate a common research scenario involving three distinct experimental conditions or groups (labeled A, B, and C). We will generate data where we intentionally introduce differences in the underlying population means, a setup typical in fields ranging from biology to experimental psychology.

Step 1: Fitting the One-Way ANOVA Model in R

The essential first step in R is fitting the [ANOVA](#) model using the appropriate functions. This process defines the relationship between the continuous response variable and the categorical grouping factor, enabling R to test the overall null hypothesis of mean equality.

The following R code block demonstrates how to create a reproducible sample dataset of 90 observations, equally distributed across the three hypothetical groups (A, B, and C). Following data creation, we utilize R's base `aov()` function to fit the standard one-way [ANOVA](#) model, testing whether the mean **values** are equivalent across the **group** factor levels.

```
#make this example reproducible  
set.seed(0)
```

```

#create data
data <- data.frame(group = rep(c("A", "B", "C"), each = 30),
  values = c(runif(30, 0, 3),
  runif(30, 0, 5),
  runif(30, 1, 7)))

#view first six rows of data
head(data)

group values
1 A 2.6900916
2 A 0.7965260
3 A 1.1163717
4 A 1.7185601
5 A 2.7246234
6 A 0.6050458

#fit one-way ANOVA model
model <- aov(values~group, data=data)

#view the model output
summary(model)

Df Sum Sq Mean Sq F value Pr(>F)
group 2 98.93 49.46 30.83 7.55e-11 ***
Residuals 87 139.57 1.60
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Reviewing the output from the **summary(model)** command, our attention is drawn to the [p-value](#) associated with the **group** factor, which is reported as **7.55e-11**. Since this value is orders of magnitude smaller than the standard alpha level of 0.05, we have robust statistical evidence to reject the null hypothesis. This confirms that significant differences are indeed present among the group means, thereby validating the necessity of proceeding to the [post hoc test](#).

Step 2: Executing Tukey's HSD Test in R

Once the significant ANOVA result is established, the transition to the pairwise comparisons is straightforward using the R statistical package. The built-in **TukeyHSD()** function is specifically

designed to accept the output object from an **aov()** model. This function automatically performs all necessary calculations, including applying the specific adjustment required to control the [family-wise error rate](#) across all comparisons simultaneously.

We apply the **TukeyHSD()** function to our fitted **model** object, explicitly setting the **conf.level** argument to 0.95. This confidence level dictates the range of the confidence intervals displayed in the output, which correspond directly to an alpha level (α) of 0.05 for hypothesis testing.

#perform Tukey's Test

```
TukeyHSD(model, conf.level=.95)
```

```
Tukey multiple comparisons of means
95% family-wise confidence level
```

```
Fit: aov(formula = values ~ group, data = data)
```

```
$group
diff lwr upr p adj
B-A 0.9777414 0.1979466 1.757536 0.0100545
C-A 2.5454024 1.7656076 3.325197 0.0000000
C-B 1.5676610 0.7878662 2.347456 0.0000199
```

The resulting output matrix generated by **TukeyHSD()** comprehensively lists the results for all pairwise comparisons: B vs. A, C vs. A, and C vs. B. Correctly interpreting the columns of this table is paramount to drawing statistically sound conclusions.

Interpreting the Tukey's HSD Numerical Results

The results table provides four key columns that must be analyzed to determine which specific groups are significantly different from one another:

diff: This column presents the raw mean difference between the two groups being compared (e.g., the mean of group B minus the mean of group A).

lwr and **upr:** These define the lower and upper bounds of the specified 95% confidence interval for the true mean difference. The most crucial interpretation rule here is: if the interval brackets zero (i.e., one bound is negative and one is positive), we fail to reject the null hypothesis of equal means for that pair.

p adj: This is the adjusted [p-value](#). Because Tukey's method uses the studentized range statistic to account for the multiplicity of comparisons, this value is adjusted to ensure that the overall probability of making a Type I error (the FWER) does not exceed 0.05.

A statistically significant difference between any pair of means is established only if the **p adj** value is strictly less than the alpha level of 0.05. Conversely, if the p-adjusted value is 0.05 or greater, we must conclude that there is insufficient evidence to claim a difference between those two specific groups.

By analyzing the p-adjusted values in our example, we observe highly significant differences across all three pairs at the 0.05 level:

Comparison B-A: **p adj = 0.0100545**. Since $0.010 < 0.05$, we conclude that Group B is significantly different from Group A.

Comparison C-A: **p adj = 0.0000000**. This indicates an extremely highly significant difference.

Comparison C-B: **p adj = 0.0000199**. This also confirms a highly significant difference between these two groups.

Visualizing Pairwise Differences

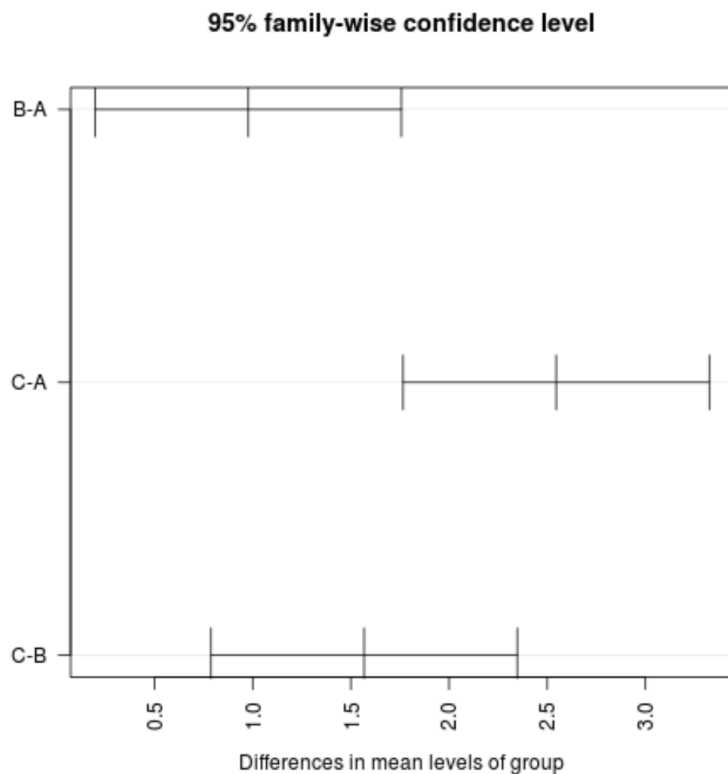
While the numerical table provides the necessary precision, visualizing the confidence intervals offers an intuitive and powerful way to present and confirm the findings. The **plot()** function, when applied directly to the output object of **TukeyHSD()**, generates a graphical representation of these pairwise confidence intervals.

The key feature of this plot is the central vertical line, which represents a mean difference of zero. If a confidence interval for a specific comparison spans across or includes this zero line, it graphically demonstrates that the null hypothesis (that the true difference is zero) cannot be rejected. Conversely, if the entire confidence interval lies entirely on one side of the zero line (either all positive or all negative), the difference between those two means is statistically significant.

#plot confidence intervals

plot(TukeyHSD(model, conf.level=.95), las = 2)

For enhanced readability, especially when comparison labels are lengthy, the optional argument **las = 2** is utilized. This ensures that the tick mark labels representing the group comparisons on the y-axis are displayed perpendicularly.



The visual output perfectly corroborates our numerical findings. We observe that none of the plotted 95% confidence intervals for the mean differences overlap with the vertical line at zero. This strong graphical evidence confirms that every pairwise comparison (A vs B, A vs C, and B vs C) resulted in a statistically significant difference in mean values.

Drawing Definitive Conclusions

Based on the combined evidence from the highly significant [ANOVA](#) F-test and the precise pairwise analysis performed by [Tukey's HSD test](#), we can move from the general finding of variance to specific, actionable conclusions regarding the performance or characteristics of the groups:

Group C exhibited mean values that were significantly superior (or different) when compared to the mean values of both Group A and Group B.

Group B's mean values were also found to be significantly higher (or different) when compared specifically to the mean values of Group A.

This detailed [post hoc analysis](#) provides researchers with the necessary statistical foundation not only to confirm that differences exist but also to confidently rank and describe the statistical separation among all groups studied.

Additional Resources for ANOVA and Post Hoc Analysis

[A Guide to Using Post Hoc Tests with ANOVA](#)

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

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