

# Perform Tukey's Test in Python

Authored by  
**Mohammed loot**

November 6, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Perform Tukey's Test in Python*. PSYCHOLOGICAL STATISTICS.  
Retrieved from <https://statistics.arabpsychology.com/?p=11497>

When analyzing experimental data, researchers often need to determine if there is a statistically significant difference among the means of multiple independent groups. The [one-way ANOVA](#) (Analysis of Variance) is the primary statistical tool used for this purpose.

The ANOVA procedure tests the null hypothesis that all group means are equal. If the resulting overall [p-value](#) falls below a predefined significance level (such as 0.05), we reject the null hypothesis, concluding that at least one group mean differs from the others.

However, the ANOVA result is non-specific; it only indicates that differences exist, but not which specific pairs of groups are causing the variation. To pinpoint these exact differences, we must employ a subsequent procedure known as a [post hoc test](#).

## Understanding ANOVA and the Need for Post Hoc Analysis

ANOVA is a powerful initial step, but relying solely on multiple t-tests after a significant ANOVA result is inappropriate. Performing numerous t-tests increases the likelihood of making a Type I error--falsely rejecting the null hypothesis--because each test carries its own intrinsic risk of error.

This accumulation of error across multiple comparisons is quantified by the [family-wise error rate](#). To maintain statistical rigor, any procedure used for pairwise comparisons must effectively control this error rate, ensuring that the overall probability of making a false positive across the entire set of comparisons remains acceptable.

## Introducing Tukey's Honestly Significant Difference (HSD) Test

One of the most robust and widely used methods for controlling the family-wise error rate during post hoc analysis is [Tukey's Test](#), also known as the Tukey Honestly Significant Difference (HSD) test.

Tukey's HSD allows for simultaneous pairwise comparisons of all group means. It uses a studentized range distribution to determine the minimum significant difference required between two group means to declare them statistically distinct, all while holding the overall Type I error probability constant.

This tutorial provides a detailed, step-by-step example demonstrating how to execute and interpret **Tukey's Test** using the Python programming language, leveraging key statistical libraries.

## Step 1: Preparing the Environment and Loading Necessary Packages

Before proceeding with the analysis, we must import the necessary libraries in Python. We will rely on **Pandas** for data manipulation, **NumPy** for numerical operations, **SciPy** for the initial ANOVA calculation, and the powerful [Statsmodels](#) library, which contains the specialized function for

Tukey's HSD.

```
import pandas as pd
import numpy as np
from scipy.stats import f_oneway
from statsmodels.stats.multicomp import pairwise_tukeyhsd
```

## Step 2: Conducting the Prerequisite One-Way ANOVA

Our analysis begins by defining a sample dataset containing measurements for three independent groups: A, B, and C. We will then fit the [one-way ANOVA](#) model using the `f_oneway` function from SciPy to test the preliminary hypothesis of mean equality.

The code below sets up the synthetic data and executes the initial ANOVA test:

```
#enter data for three groups
a =
b =
c =

#perform one-way ANOVA
f_oneway(a, b, c)

F_onewayResult(statistic=5.167774552944481, pvalue=0.012582197136592609)
```

Reviewing the output, we observe that the calculated [p-value](#) is **0.01258**.

Since this value is considerably less than our standard significance level of 0.05, we confidently reject the null hypothesis. This significant result confirms that the mean scores across groups A, B, and C are not all equal, thereby justifying the necessity of conducting a [post hoc test](#) to isolate the specific differences.

## Step 3: Performing Tukey's HSD Pairwise Comparison

To execute **Tukey's Test**, the data must first be structured into a Pandas DataFrame suitable for the `pairwise_tukeyhsd()` function. This function requires the response variable (score) and the grouping variable (group) to be explicitly defined.

We use the `pairwise_tukeyhsd()` function from the **statsmodels** library, setting the endogenous variable (endog) to the scores and the grouping variable (groups) to the group labels. The alpha level is set to 0.05, representing the maximum acceptable [family-wise error rate](#).

```
#create DataFrame to hold data
df = pd.DataFrame({'score': ,
'group': np.repeat(, repeats=10)})
```

```
# perform Tukey's test
tukey = pairwise_tukeyhsd(endog=df,
groups=df,
alpha=0.05)
```

```
#display results
print(tukey)
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

```
=====
group1 group2 meandiff p-adj lower upper reject
-----
a b 8.4 0.0158 1.4272 15.3728 True
a c 1.3 0.8864 -5.6728 8.2728 False
b c -7.1 0.0453 -14.0728 -0.1272 True
-----
```

## Interpreting the Results of the Tukey HSD Test

The resulting table provides a concise summary of the pairwise comparisons. The key columns to examine are `meandiff` (the difference between the group means), `p-adj` (the adjusted [p-value](#), corrected for multiple comparisons), and `reject` (a boolean indicating whether the null hypothesis of equal means for that specific pair is rejected).

Based on the output, we can analyze the significance of each comparison against the critical alpha level (0.05):

Comparison between Group A and Group B: The adjusted p-value is **0.0158**. Since  $0.0158 < 0.05$ , the null hypothesis is rejected (True). This indicates a statistically significant difference between the means of groups A and B.

Comparison between Group A and Group C: The adjusted p-value is **0.8864**. Since 0.8864 is much greater than 0.05, we fail to reject the null hypothesis (False). There is no statistically significant difference between groups A and C.

Comparison between Group B and Group C: The adjusted p-value is **0.0453**. Since  $0.0453 < 0.05$ , the null hypothesis is rejected (True). This confirms a statistically significant difference between the

means of groups B and C.

In summary, the [Tukey's Test](#) definitively showed that Group A differs significantly from Group B, and Group B differs significantly from Group C. However, Group A and Group C do not exhibit a statistically significant difference in their mean scores.

## Conclusion and Further Resources

Mastering the application of [Tukey's Test](#) is vital when conducting multi-group comparisons following a significant [ANOVA](#) result. By utilizing Python's **statsmodels** library, researchers can efficiently control the [family-wise error rate](#) and accurately identify which specific groups contribute to the overall variance.

This methodology ensures that conclusions drawn about group mean differences are statistically sound and reliable.

For those interested in exploring related statistical concepts or alternative post hoc procedures, please consult the resources linked within this article.