

# Dunn's Test in Python

Authored by  
**stats writer**

December 21, 2025

## RECOMMENDED CITATION

stats writer (2025). *Dunn's Test in Python*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=108250>

## Understanding Dunn's Test: A Non-Parametric Post Hoc Analysis

Dunn's Test is a critically important statistical procedure utilized for conducting multiple comparisons among two or more independent samples. It is a robust non-parametric method, meaning it does not necessitate the restrictive assumptions of normality or homogeneity of variances often required by parametric alternatives. This makes it exceptionally valuable when analyzing data that are ordinal, non-normally distributed, or derived from groups with unequal variances. The fundamental purpose of Dunn's Test is to identify precisely which pairs of groups exhibit statistically significant differences, following an initial omnibus test that indicated an overall difference exists.

The methodology relies on ranking the combined data set and then calculating a test statistic based on the differences in the sums of ranks between the groups being compared. Unlike simpler pairwise comparisons, Dunn's Test incorporates a crucial mechanism: the adjustment of p-values. This adjustment is essential for controlling the inflated risk of committing a Type I error (false positive) that inherently arises when performing numerous simultaneous comparisons across multiple groups in a single experiment.

Implementing Dunn's Test effectively in a programming environment such as Python provides researchers with efficiency and precision. By using dedicated statistical libraries, complex rank-based calculations and subsequent p-value corrections, such as the widely used Bonferroni correction, are executed seamlessly, allowing for reliable interpretation of results pertaining to group differences.

### The Role of Kruskal-Wallis H Test

Dunn's Test is categorized as a post hoc test, meaning it must only be applied subsequent to the successful rejection of the null hypothesis by a prerequisite omnibus test. For non-parametric data comparing three or more independent groups, this prerequisite test is the Kruskal-Wallis test. The Kruskal-Wallis H test is considered the non-parametric analog to the traditional One-Way ANOVA.

The Kruskal-Wallis test determines if there is a statistically significant difference between the medians of the groups under comparison. If the resulting p-value from the Kruskal-Wallis test is greater than the chosen significance level (e.g.,  $\alpha = 0.05$ ), we conclude that the data do not provide sufficient evidence to suggest any difference among the groups, and the analysis halts. However, if the Kruskal-Wallis test yields a statistically significant result, it only confirms that the medians are not all equal, but fails to identify which specific pairs of groups are dissimilar.

Therefore, a significant Kruskal-Wallis result necessitates a follow-up test. Directly proceeding to multiple unadjusted pairwise comparisons (like repeated Mann-Whitney U tests) would dramatically increase the likelihood of spurious findings. This is precisely where Dunn's Test becomes

indispensable, providing the controlled environment required to validate specific group differences.

## Why Dunn's Test is Necessary for Controlled Comparisons

The necessity of using a controlled post hoc procedure like Dunn's Test stems directly from the problem of multiple comparisons. When researchers compare every possible pair among  $k$  groups, they perform  $k(k-1)/2$  tests. Each individual test has a chance of yielding a false positive (Type I error). As the number of comparisons increases, the overall probability of making at least one Type I error across the entire set of comparisons, known as the family-wise error rate (FWER), escalates rapidly.

Dunn's Test is specifically tailored to address this inflation by integrating a specialized test statistic derived from the rank sums and applying corrections to the resulting p-values. This rigorous approach ensures that the overall FWER is maintained at or below the predefined significance level, typically  $\alpha = 0.05$ . By controlling the FWER, Dunn's Test grants higher confidence in the specific pairwise conclusions drawn from the data.

In contrast to other popular post hoc tests used after Kruskal-Wallis, Dunn's test is often preferred for its clear methodology and the availability of various correction methods, providing flexibility while maintaining statistical integrity. It ensures that researchers only report differences that are robust after accounting for the inherent risks associated with repeated hypothesis testing.

## Implementing Dunn's Test in Python

To perform Dunn's Test efficiently within the Python ecosystem, researchers rely on specialized external libraries. The most authoritative and practical library for non-parametric post hoc analysis is `scikit-posthocs`. This library provides advanced statistical functions that extend the capabilities of foundational libraries like NumPy and SciPy, making complex statistical procedures accessible and reproducible.

The key function within this library for our purpose is `posthoc_dunn()`. This function accepts the data organized by groups and performs the ranking, calculates the pairwise test statistics, and applies the chosen p-value adjustment method. Utilizing this function simplifies the otherwise manually intensive process of conducting rank-sum calculations and subsequent error correction.

Effective implementation requires understanding how to structure the input data--typically a list containing sub-lists, where each sub-list represents an independent sample group--and correctly specifying parameters like the `p_adjust` method. The precision offered by scikit-posthocs allows for rapid analysis consistent with stringent statistical standards.

## Prerequisites: Setting up the Python Environment

Before commencing the statistical analysis, the necessary package must be installed. The `scikit-posthocs` library is not typically included in standard Python distributions or virtual environments and must be explicitly added using the package manager `pip`. This installation ensures that all required statistical functionalities are available for use in the subsequent script.

The installation process is simple and performed through the terminal or command prompt associated with your Python environment. Once installed, the library can be imported into any Python script or Jupyter Notebook for immediate use.

### Step 1: Install scikit-posthocs.

To install the necessary library, execute the following command:

```
pip install scikit-posthocs
```

Successful execution confirms that the environment is prepared to handle the complex non-parametric computations required by Dunn's Test. This foundational step is crucial for maintaining a clean and functional statistical workflow.

## Case Study: Fertilizer Impact on Plant Growth

To illustrate the application of Dunn's Test, consider a research question focusing on the efficacy of three distinct fertilizer treatments on plant height. Thirty plants are randomly allocated into three groups (Group 1, Group 2, Group 3), each receiving a different fertilizer formulation. After a fixed period, the height of each plant is measured. The researchers hypothesize that the median growth across these groups is not equal.

Upon gathering the data, the researchers first apply the Kruskal-Wallis test. This initial test confirms a statistically significant result ( $p < 0.05$ ), indicating a global difference in plant height exists among the three fertilizers. However, this result does not tell the full story; it merely confirms that a difference is present somewhere within the group comparisons.

To isolate the specific differences--for instance, to confirm if Fertilizer B is statistically better than Fertilizer C--the researchers must employ a post hoc procedure. This necessity leads directly to the application of Dunn's Test to perform controlled, pairwise comparisons. This procedure utilizes the raw data gathered in the experiment:

Group 1 Data:

Group 2 Data:

Group 3 Data:

## Executing the Dunn's Test in Python

With the data collected and the prerequisite Kruskal-Wallis test completed, the next step is to execute the pairwise comparisons using the `posthoc_dunn()` function from the `scikit-posthocs` library in `Python`. A critical decision during this step is the selection of the p-value adjustment method. Here, we select the `'bonferroni'` correction due to its ability to strictly control the family-wise error rate.

The code below structures the group data and performs the necessary rank calculations and adjustments:

### Step 2: Perform Dunn's test.

**#specify the growth of the 10 plants in each group**

**group1 =**

**group2 =**

**group3 =**

**data =**

**#perform Dunn's test using a Bonferonni correction for the p-values**

**import scikit\_posthocs as sp**

**sp.posthoc\_dunn(data, p\_adjust = 'bonferroni')**

1 2 3

1 1.000000 0.550846 0.718451

2 0.550846 1.000000 0.036633

3 0.718451 0.036633 1.000000

The resulting matrix provides the adjusted p-values for the three possible pairwise comparisons (1 vs 2, 1 vs 3, and 2 vs 3).

## Interpreting the Results Matrix

Interpretation of the output matrix is straightforward. Each cell represents the adjusted p-value for the comparison between the row group and the column group. We compare these adjusted p-values against our significance threshold,  $\alpha = 0.05$ . If the adjusted p-value is less than 0.05, the difference is considered statistically significant.

Based on the results generated in the previous step, using the Bonferroni correction:

The adjusted p-value for the comparison between Group 1 and Group 2 is **0.550846**. Since this value is significantly larger than 0.05, we conclude there is no statistical difference in plant height between Fertilizer 1 and Fertilizer 2.

The adjusted p-value for the comparison between Group 1 and Group 3 is **0.718451**. This value is also much greater than 0.05, indicating no significant difference between Fertilizer 1 and Fertilizer 3.

The adjusted p-value for the comparison between Group 2 and Group 3 is **0.036633**. Since  $0.036633 < 0.05$ , we conclude that the difference between Fertilizer 2 and Fertilizer 3 is statistically significant.

Thus, the post hoc analysis using Dunn's Test reveals that the overall significant result from the Kruskal-Wallis test is attributable solely to the difference between Group 2 and Group 3.

## Controlling Errors: P-Value Adjustment Methods

The robustness of Dunn's Test hinges on the selection of an appropriate p-value adjustment method, which mitigates the risk of inflating the Type I error rate. While the Bonferroni correction is often the default choice due to its strong control over the FWER, it is also highly conservative. Researchers may choose less conservative methods if they require greater statistical power, accepting a slightly higher risk of Type I errors in exchange for detecting more subtle effects.

The `posthoc_dunn()` function in `scikit-posthocs` accommodates a wide array of alternative adjustment methods. These methods vary in their approach to balancing Type I and Type II error rates, some focusing strictly on FWER control (like Bonferroni), while others prioritize control of the False Discovery Rate (FDR).

Other potential choices for the `p_adjust` argument include:

sidak  
holm-sidak  
simes-hochberg  
hommel  
fdr\_bh (Benjamini-Hochberg False Discovery Rate)  
fdr\_by (Benjamini-Yekutieli False Discovery Rate)  
fdr\_tsbh

Researchers should consult the specific guidelines for their field of study and the official documentation for comprehensive details on how each of these adjustments operates and which is most appropriate for their specific experimental design.

Further resources on non-parametric statistical methods and post hoc testing:

---

An Introduction to Dunn's Test for Multiple Comparisons

How to Perform Dunn's Test in R

ARABPSYCHOLOGY.COM