

Perform Dunn's Test in R

Authored by
Mohammed looti

November 7, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Perform Dunn's Test in R*. PSYCHOLOGICAL STATISTICS.
Retrieved from <https://statistics.arabpsychology.com/?p=12175>

Understanding Non-Parametric Post-Hoc Analysis

When researchers need to compare the central tendencies of three or more independent groups, the standard approach is often the [One-Way ANOVA](#). However, this parametric test relies on strict assumptions, notably that the data within each group are normally distributed and that the variances are homogeneous. When these assumptions are violated, particularly when dealing with small sample sizes, ordinal data, or heavily skewed distributions, a robust alternative is required.

The [Kruskal-Wallis test](#) serves as the non-parametric equivalent of the One-Way ANOVA. It determines whether a statistically significant difference exists between the medians of three or more independent groups. By working with ranks rather than the raw data values, the Kruskal-Wallis test bypasses the restrictive assumptions of normality and equal variance, making it an indispensable tool in certain statistical contexts.

A significant result from a Kruskal-Wallis test confirms only that at least one group median is different from the others. It does not identify which specific pairs of groups are causing this overall significant difference. To isolate these specific group differences, a subsequent post-hoc analysis is mandatory. Failing to conduct this step would leave the researcher with an incomplete and non-actionable conclusion.

Why and When to Use Dunn's Test

If the [Kruskal-Wallis test](#) yields a statistically significant p-value, indicating that differences exist, the next logical step is to perform a post-hoc test. While one might be tempted to use simple pairwise [non-parametric](#) tests (like the Wilcoxon rank-sum test) for all combinations, this approach is problematic because it increases the overall family-wise error rate (the probability of making at least one Type I error).

The most statistically appropriate procedure for post-hoc analysis following a Kruskal-Wallis test is [Dunn's Test](#). Developed by Olive Jean Dunn, this test is specifically designed to perform multiple comparisons based on the rank sums derived from the preceding Kruskal-Wallis procedure. Unlike running multiple independent Wilcoxon tests, Dunn's Test incorporates the pooled variance structure inherent in the original Kruskal-Wallis rank calculation, providing a more reliable comparison.

Dunn's Test is crucial because it inherently includes a mechanism to adjust the p-values for multiple comparisons, thus controlling the inflation of the Type I error rate. Without this adjustment, the likelihood of finding a spurious significant difference--a false positive--increases dramatically with the number of groups compared. This tutorial provides a definitive guide on executing and interpreting Dunn's Test efficiently within the statistical programming environment, **R**.

Designing the Study: A Research Example

Consider a practical scenario: A medical researcher aims to determine whether three distinct pharmacological agents (Drug A, Drug B, and Drug C) have varying effects on chronic back pain levels. The researcher recruits 30 individuals who report similar levels of persistent back pain and randomly assigns them into three groups, with 10 participants receiving one of the three drugs.

After a one-month treatment period, each participant is asked to rate their current back pain severity on a continuous scale ranging from 1 to 100, where 100 signifies the maximum possible pain. Given the nature of subjective pain ratings, which often do not follow a normal distribution, the researcher decides to proceed with the robust [Kruskal-Wallis test](#), setting the significance threshold (α) at **0.05**. The objective is to first establish if any difference exists across the group medians.

The following **R** code demonstrates the creation of the synthetic dataset for this experiment and the execution of the initial Kruskal-Wallis omnibus test.

Setting Up the Data and Initial Kruskal-Wallis Test in R

To ensure reproducibility of the simulated data, the `set.seed()` function is used. We then construct a data frame containing two variables: the factor variable `drug` (A, B, or C) and the response variable `pain` (the self-reported pain score).

Ensure the example is reproducible across sessions

```
set.seed(0)
```

```
# Create the primary data frame
```

```
# Assign 10 observations to each of the three drug groups
```

```
data <- data.frame(drug = rep(c("A", "B", "C"), each = 10),
```

```
pain = c(runif(10, 40, 60), # Drug A scores
```

```
runif(10, 45, 65), # Drug B scores
```

```
runif(10, 55, 70))) # Drug C scores
```

```
# Display the first few rows to verify structure
```

```
head(data)
```

```
# drug pain
```

```
#1 A 57.93394
```

```
#2 A 45.31017
```

```
#3 A 47.44248
```

```
#4 A 51.45707
```

```
#5 A 58.16416
```

```
#6 A 44.03364
```

```
# Perform the Kruskal-Wallis Rank Sum Test
```

```
kruskal.test(pain ~ drug, data = data)
```

```
Kruskal-Wallis rank sum test
```

```
data: pain by drug
```

```
Kruskal-Wallis chi-squared = 11.105, df = 2, p-value = 0.003879
```

Upon reviewing the output of the Kruskal-Wallis test, the overall p-value is **0.003879**. Since this value is considerably less than the predetermined significance level of $\alpha = 0.05$, we reject the null hypothesis. This finding strongly suggests that there is a statistically significant difference in the reported median pain levels across the three drug treatment groups. Because the omnibus test is significant, we are justified--and required--to proceed with the post-hoc multiple comparison procedure, which is Dunn's Test.

Executing Dunn's Test for Multiple Comparisons

To execute [Dunn's Test](#) in R, we must utilize the `dunnTest()` function, which is available within the **FSA** package. If this library is not already installed, it must be loaded using the `library()` command before proceeding.

A critical component of the Dunn's Test procedure is the selection of a p-value adjustment method. As discussed, running multiple tests (comparing A vs B, A vs C, and B vs C) without adjustment drastically inflates the family-wise error rate. For this example, we will initially employ the stringent [Bonferroni correction](#), which is known for its strong control over Type I error, though sometimes at the cost of statistical power.

```
# Load the necessary library for Dunn's Test
```

```
library(FSA)
```

```
# Perform Dunn's Test
```

```
# Specify the formula (pain by drug), the data source, and the adjustment method
```

```
dunnTest(pain ~ drug,
```

```
data=data,
```

```
method="bonferroni")
```

```
Dunn (1964) Kruskal-Wallis multiple comparison
```

```
p-values adjusted with the Bonferroni method.
```

```

Comparison Z P.unadj P.adj
1 A - B -0.8890009 0.374002602 1.000000000
2 A - C -3.2258032 0.001256197 0.003768591
3 B - C -2.3368023 0.019449464 0.058348393

```

The output table provides the results for all possible pairwise comparisons (A-B, A-C, B-C). The key columns for interpretation are Z (the test statistic), P . unadj (the raw p-value before correction), and P . adj (the p-value after applying the specified correction method, in this case, Bonferroni).

Interpretation of Results and P-Value Adjustments

The final column, P . adj, dictates the conclusion. We compare these adjusted p-values against our chosen significance level, $\alpha = 0.05$.

Comparison A - B: The adjusted p-value is 1.000. Since $1.000 > 0.05$, there is no statistically significant difference in median pain levels between Drug A and Drug B.

Comparison A - C: The adjusted p-value is **0.003768591**. Since 0.003768 is substantially less than 0.05, we conclude that there is a highly statistically significant difference between Drug A and Drug C.

Comparison B - C: The adjusted p-value is 0.058348393. This value is slightly greater than 0.05, meaning we fail to reject the null hypothesis for this pair. Under the strict Bonferroni criterion, the difference between Drug B and Drug C is not considered statistically significant.

Based on the analysis, only drugs A and C are definitively statistically different from one another regarding their effect on back pain, assuming an alpha level of 0.05.

Alternative P-Value Correction Methods

While the [Bonferroni correction](#) is straightforward and reliable, it is highly conservative, meaning it can sometimes fail to detect true effects (i.e., increase the risk of Type II errors). The `dunnTest()` function supports several alternative methods for controlling the family-wise error rate or the false discovery rate (FDR). Researchers should select the method that best balances the risk of Type I and Type II errors for their specific field of study.

Other widely accepted p-value adjustment methods available in the `dunnTest()` function include:

"sidak" (Sidak adjustment): Less conservative than Bonferroni, often used when tests are independent.

"holm" (Holm Adjustment): Generally considered a step-down procedure that is uniformly more powerful than Bonferroni, making it a highly recommended default.

"hs" (Holm-Sidak Adjustment): A variation combining aspects of both Holm and Sidak procedures.

"bs" (Bonferroni-Sidak Adjustment): Another hybrid method.

"by" (Benjamini-Yekutieli Adjustment): Controls the False Discovery Rate (FDR) under dependency assumptions.

"bh" ([Benjamini-Hochberg procedure](#)): A popular method for controlling the False Discovery Rate (FDR), often preferred in exploratory studies where many hypotheses are tested simultaneously.

For most comparative research following a Kruskal-Wallis test, the Holm method provides an excellent balance of rigor and power, making it a strong alternative to the default Bonferroni setting used in the example above. Choosing the correct adjustment technique is a critical step in ensuring the validity of the final statistical conclusion.