Learn How to Perform a Kruskal-Wallis Test in Python

Authored by

Mohammed looti

November 8, 2025

The Kruskal-Wallis Test, frequently termed the Kruskal-Wallis H Test, is a cornerstone procedure within non-parametric statistics. Data analysts and researchers rely on this robust test to systematically determine if statistically significant differences exist among the medians of three or more independent population groups. This analytical approach proves indispensable when datasets fail to satisfy the demanding assumptions required by traditional parametric models, particularly the requirements of population normality and homogeneity of variances. Mastering the application of this test is vital for producing reliable insights from complex, real-world data.

In contrast to its parametric counterpart, the One-way ANOVA, the Kruskal-Wallis Test does not analyze the raw data values directly. Instead, it operates on the ranks assigned to the observations. This critical transformation process makes the test inherently resistant to the influence of extreme outliers and highly appropriate for analyzing data measured on an **ordinal scale** or continuous data that exhibits severe skewness. A fundamental skill in modern data science involves understanding how to correctly implement and interpret this powerful statistical procedure using sophisticated software environments, such as Python's specialized SciPy library.

This comprehensive tutorial is designed to provide both the theoretical foundation and the practical implementation necessary for applying the Kruskal-Wallis Test. We will guide you through the core principles, introduce a detailed, step-by-step example executed in Python, and ensure clarity in data preparation, execution of the statistical function, and the meaningful interpretation of the resultant statistics. By the end of this guide, you will be equipped to confidently apply this methodology to your own non-normally distributed datasets.

Introduction to the Kruskal-Wallis H Test

The Kruskal-Wallis Test functions as the primary non-parametric alternative when researchers need to compare three or more groups, serving the same comparative purpose as the One-way ANOVA. Its central objective is to rigorously test the **Null hypothesis** (H_0), which posits that all populations from which the samples were drawn possess identical median values. Should the test yield a significant result and lead to the rejection of this null hypothesis, it statistically confirms that at least one group median is significantly distinct from the others. This non-parametric approach holds considerable value across diverse disciplines, including medicine, social sciences, and engineering, where collected data frequently deviates significantly from the idealized normal distribution.

The procedural methodology begins by pooling all observational data points from every group into a single, unified set. These observations are then assigned numerical ranks, starting with rank 1 for the smallest value and progressing sequentially to the largest observation. The test statistic, denoted as H, is subsequently calculated based on the summation of these assigned ranks within each individual group. If the true median values across all groups are indeed similar, the

average rank sums for each respective group should be approximately equal, resulting in a low calculated H statistic. Conversely, substantial differences in the rank sums across groups indicate a profound disparity in medians, thereby generating a large H statistic and, critically, a small p-value.

It is essential to understand that while the Kruskal-Wallis Test signals the existence of a difference, it is fundamentally an **omnibus test**. When the H_0 is successfully rejected, the result does not specify which particular pairs of groups are statistically different. Consequently, subsequent post-hoc analyses are mandatory to pinpoint the exact location of the significant disparities. The test's powerful ability to resist violations of parametric assumptions solidifies its position as an indispensable analytical tool when handling complex datasets characterized by non-ideal structures and high variability.

When to Use the Kruskal-Wallis Test

The decision to employ a specific statistical test rests entirely on the inherent characteristics of the data and the validity of the assumptions that can be met. The Kruskal-Wallis Test is the statistically sound selection in specific scenarios, primarily those requiring the comparison of three or more independent groups where the dependent variable is measured at an ordinal or continuous level, but where the crucial parametric assumptions have been violated.

There are three primary circumstances that compel researchers to choose this non-parametric method over the One-way ANOVA. Firstly, if the data distributions within the groups are observably not **normally distributed**, and the sample sizes are insufficient to rely on the robust principles of the Central Limit Theorem. Secondly, the presence of significant **heteroscedasticity** (unequal variances across the different groups) undermines the fundamental reliability of ANOVA, making a rank-based test necessary. Thirdly, if the dependent variable itself is measured on a strict ordinal scale, calculating a true mean is statistically misleading, making the median the most appropriate and representative measure of central tendency.

Although the Kruskal-Wallis Test demands fewer assumptions than ANOVA, the requirement for **independence of groups** remains paramount. This ensures that the observations collected within one fertilizer group, for instance, have absolutely no influence on the observations collected in any other group. Furthermore, while the test ideally assumes that the distributions across groups share a similar shape (even if non-normal), modern statistical interpretations acknowledge that if distribution shapes vary significantly, the test still compares central tendencies. However, in such cases, results must be interpreted cautiously, as the detected differences might be attributable to variations in distributional shape rather than strictly median disparities.

Setting Up the Experiment and Data in Python

To demonstrate the practical utility of the Kruskal-Wallis Test, we will utilize a classic experimental design scenario. Imagine agricultural scientists are investigating the relative effectiveness of three distinct fertilizer formulations (labeled A, B, and C) on overall plant growth. Their research hypothesis suggests that these fertilizer types will produce statistically different median plant heights after a standardized measurement period. To conduct this experiment, 30 plants are randomly selected and divided into three independent groups of 10 plants each. Each group is then exclusively treated with one fertilizer type. Following a one-month growth period, the height of every single plant is meticulously recorded in centimeters.

Our objective is to apply the Kruskal-Wallis Test to determine if the median plant growth measurements are statistically identical across the three treatment groups. We will perform this analysis using the Python programming language, specifically relying on the specialized statistical functions provided by the comprehensive SciPy library, which includes the necessary `kruskal` function. The initial and most critical step is the accurate representation of the collected data within Python, typically structured as separate arrays or lists corresponding to each independent group.

The following data represents the measured plant heights for the three fertilizer groups. We define these measurements as distinct variables--`group1`, `group2`, and `group3`--which will serve as the required input arguments for the statistical function call. This precise data entry step is non-negotiable for ensuring the data structure conforms to the requirements of the SciPy function.

Step 1: Enter the data.

We will create three arrays to hold our plant measurements for each of the three fertilizer groups:

group1 =
group2 =
group3 =

Step-by-Step Implementation using SciPy

Once the experimental data has been successfully loaded and structured into appropriate Python variables, the subsequent stage involves the actual execution of the statistical test. The `scipy.stats` module is recognized as the industry standard for conducting statistical analyses in Python, owing to its accurate and efficient implementation of numerous functions, including `kruskal`. This function is designed to accept the data arrays of the independent groups as arguments and, in return, provides the calculated test statistic (H) and the corresponding p-value.

A primary benefit of utilizing the SciPy library is the immediate simplification of complex calculations. By merely invoking `stats.kruskal(group1, group2, group3)`, the function autonomously manages the entire necessary sequence of steps: pooling the observations, assigning ranks, calculating the sum of ranks for each group, and ultimately computing the H statistic. Furthermore, it determines the probability of observing a difference of this magnitude (or one more extreme) under the strict assumption that the Null hypothesis is true.

The numerical output generated by the function is essential for hypothesis testing and decision-making. The test statistic quantifies the degree to which the observed data deviates from the null hypothesis expectation, while the p-value furnishes the critical probability metric required for formal statistical inference. The process requires importing the specific module and then executing the function using our previously defined data variables, as demonstrated clearly in the following code block.

Step 2: Perform the Kruskal-Wallis Test.

Next, we will perform a Kruskal-Wallis Test using the `kruskal` function from the `scipy.stats` library:

from scipy import stats

#perform Kruskal-Wallis Test
stats.kruskal(group1, group2, group3)

(statistic=6.2878, pvalue=0.0431)

Interpreting the Statistical Results

Accurately interpreting the output from the Kruskal-Wallis Test necessitates a precise grasp of the formal hypotheses under examination and the meaning conveyed by the derived statistics. The test is structured around the comparison of two competing statements concerning the population medians:

The Kruskal-Wallis Test employs the following definitions for its null and alternative hypotheses:

The Null Hypothesis (H0): The population median is equal across all groups (MedianGroup1 = MedianGroup2 = MedianGroup3).

The Alternative Hypothesis (Ha): The population median is *not* equal across all groups (at least one group median differs significantly).

In the context of our fertilizer experiment, the calculated test statistic, H, is **6.2878**, and the

corresponding p-value is **0.0431**. The p-value fundamentally represents the probability of obtaining our observed data, or data showing a more extreme difference, assuming that the null hypothesis were entirely true. To reach a conclusion, this p-value is compared against a predefined significance level ($alpha$), conventionally set at 0.05. The statistical decision rule is straightforward: if the p-value is less than or equal to $alpha$, we must reject H_0.

Since the calculated p-value of 0.0431 is less than our threshold of 0.05, we formally reject the null hypothesis. This outcome constitutes a statistically significant finding. Based on the collected plant height data, we have sufficient empirical evidence to conclude that there are statistically significant differences in the median plant growth among the three distinct fertilizer groups. Practically speaking, the choice of fertilizer formulation does have a measurable impact on the median growth outcome, though the Kruskal-Wallis Test itself does not reveal which specific fertilizer is superior or inferior.

Post-Hoc Analysis and Pairwise Comparisons

As previously established, the Kruskal-Wallis Test is classified as an omnibus test; while rejecting the Null hypothesis confirms that differences exist somewhere among the groups, it fails to localize those differences. To precisely identify which specific pairs of fertilizer groups (e.g., A vs. B, A vs. C, or B vs. C) exhibit statistically significant disparities, a robust post-hoc analysis is absolutely required.

A major concern when performing multiple pairwise comparisons without appropriate adjustment is the inflation of the **family-wise error rate**--the overall probability of making at least one Type I error across the entire set of comparisons. Therefore, specialized post-hoc tests specifically designed for non-parametric data must be utilized, incorporating methods to rigorously control this error rate. The most widely recommended post-hoc procedure following a significant Kruskal-Wallis result is the **Dunn's Test**, typically paired with a correction method such as the Bonferroni adjustment, or the more statistically powerful Holm or Benjamini-Hochberg procedures.

Although alternatives like Conover's Test or the Nemenyi Test are viable, Dunn's Test is generally favored due to its computational simplicity and broad applicability. For data analysts working in Python, while `scipy.stats` handles the initial Kruskal-Wallis test efficiently, implementing the necessary non-parametric post-hoc comparisons often requires specialized extension libraries or dedicated statistical packages (sometimes relying on R integrations) to correctly apply the required error rate corrections, thereby ensuring the confidence in the specific group differences identified.

Critical Assumptions and Inherent Limitations

Despite the Kruskal-Wallis Test being a foundational element of non-parametric statistics, it is not entirely assumption-free. A clear understanding of its prerequisites is essential for guaranteeing the

validity of the results obtained in Python. Firstly, the test strictly demands the **independence of observations**. This means that the height measurement taken from any single plant must not be influenced by the measurement taken from any other plant in the experiment. This assumption is typically secured through rigorous experimental design, including random assignment of subjects to groups.

Secondly, the dependent variable must be measurable at an **ordinal or continuous level**. Because the Kruskal-Wallis Test relies solely on the ranks of the data, it is perfectly capable of handling both highly skewed continuous data and pure ordinal data (such as responses collected on a Likert scale). Thirdly, the test relies on the assumption that the samples are drawn from populations that possess the **same shape of distribution**. If the shapes of the distributions differ significantly across the groups, the Kruskal-Wallis Test may still indicate a significant difference; however, it is technically testing for differences in stochastic dominance rather than purely median differences. If the shapes are known to be similar, the comparison of population medians remains valid.

A key operational limitation, as previously detailed, is its inherent nature as an omnibus test: it merely signals the presence of a difference without specifying its location. Moreover, like all rank-based procedures, the Kruskal-Wallis Test inevitably sacrifices a degree of **statistical power** when compared to its parametric alternative, the One-way ANOVA, provided that the assumptions for ANOVA are perfectly satisfied. Consequently, researchers should primarily elect to use the Kruskal-Wallis Test only when parametric assumptions (normality and homogeneity of variance) are demonstrably violated, or when the data is intrinsically ordinal, ensuring that the statistical choice aligns optimally with the data's true underlying properties.