

Learn How to Perform a Kruskal-Wallis Test in SAS for Non-Parametric Data Analysis

Authored by
Mohammed looti

October 30, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Perform a Kruskal-Wallis Test in SAS for Non-Parametric Data Analysis*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=6250>

When statistical analysis requires comparing the medians of three or more [independent groups](#), the preferred methodology often shifts away from traditional parametric tests. Researchers frequently utilize the [Kruskal-Wallis Test](#) (KWT), a powerful [non-parametric](#) statistical procedure. This test is designed to determine whether there is a statistically significant difference in the distribution of scores across these groups. It serves as a robust and reliable alternative to the [one-way ANOVA](#), particularly when the underlying assumptions required for parametric testing, such as the normality of data distribution, cannot be satisfied.

Mastering the application and interpretation of the Kruskal-Wallis Test is essential for drawing accurate conclusions from non-normally distributed data. This comprehensive tutorial is meticulously structured to guide you through every necessary phase of conducting the KWT using [SAS](#) statistical software. We will cover everything from structuring your raw data to executing the analysis and interpreting the final output, ensuring you gain clarity and precision in your advanced statistical endeavors.

Understanding the Theoretical Foundation of the Kruskal-Wallis Test

The Kruskal-Wallis H-test, commonly known as the KWT, is fundamentally a non-parametric technique used to assess whether multiple independent samples originate from the same population or from populations with identical distributions. Its utility shines when analyzing data involving one independent variable that defines three or more categorical groups, and a dependent variable that is either ordinal or continuous but significantly skewed or non-normally distributed. Unlike the ANOVA, which focuses on comparing group means, the KWT leverages the ranks of the data points, enabling it to compare the [medians](#) of the groups. This reliance on ranks makes the test inherently less susceptible to the distorting influence of outliers and extreme skewness within the data distributions.

The core methodology of the test involves pooling all observations from all groups and ranking them from smallest to largest, irrespective of which group they belong to. Once ranked, the sum of ranks is calculated separately for each group, and these sums are then compared to assess variability. The statistical inference is based on the magnitude of the differences between these summed ranks. The test posits a strict [null hypothesis](#) (H_0), which states that the medians of all groups under investigation are equal, meaning the samples are drawn from populations sharing the same location parameter. Consequently, the alternative hypothesis (H_1) asserts that at least one group median is statistically different from the others.

Due to its robust non-parametric nature, the Kruskal-Wallis Test is an invaluable analytical instrument across diverse research disciplines, including biomedical sciences, psychological research, and ecology. Whenever collected data fails to meet the strict prerequisites of parametric methods--specifically the assumption of homogeneity of variances or normality--the KWT provides

a statistically sound pathway to reliable conclusions, broadening the scope of applicability for inferential testing.

Step 1: Preparing and Importing Data into SAS

To provide a clear demonstration of the Kruskal-Wallis Test application, let us outline a common research scenario. Imagine a team of agricultural scientists aiming to evaluate the comparative effectiveness of three distinct fertilizer types on maximizing plant growth. Their primary goal is to determine if applying these fertilizers results in significantly varied growth rates. For this study, 30 seedlings are randomly selected and divided into three equal groups of 10. Each group is systematically treated with one specific fertilizer (Fertilizer 1, Fertilizer 2, or Fertilizer 3) over a standard experimental period of one month.

At the conclusion of the experiment, the researchers precisely measure the total height increase (recorded in inches) for every single plant. This collected raw data must be structured meticulously for subsequent analysis within [SAS](#). Our resultant dataset will comprise two essential variables: `fertilizer`, which functions as the categorical independent variable defining the treatment groups, and `growth`, which is the continuous dependent variable representing the measured plant growth in inches.

The following SAS code block demonstrates the foundational step of data preparation--creating a dataset named `fertilizer_data` and entering the observed plant growth measurements corresponding to each fertilizer group. This foundational step is absolutely critical, as it ensures the data is correctly structured and available for the robust non-parametric analysis that follows.

```
/*create dataset*/  
data fertilizer_data;  
input fertilizer $ growth;  
datalines;  
fert1 7  
fert1 14  
fert1 14  
fert1 13  
fert1 12  
fert1 9  
fert1 6  
fert1 14  
fert1 12  
fert1 8  
fert2 15
```

```
fert2 17  
fert2 13  
fert2 15  
fert2 15  
fert2 13  
fert2 9  
fert2 12  
fert2 10  
fert2 8  
fert3 6  
fert3 8  
fert3 8  
fert3 9  
fert3 5  
fert3 14  
fert3 13  
fert3 8  
fert3 10  
fert3 9  
;  
run;
```

Step 2: Executing the Kruskal-Wallis Test using PROC NPAR1WAY

Once the data is successfully loaded and structured in SAS, the subsequent and most critical step is the execution of the Kruskal-Wallis Test itself. In SAS, this test is executed with efficiency and accuracy using the powerful [PROC NPAR1WAY](#) procedure. This specialized procedure is designed specifically for performing non-parametric one-way analysis, making it the ideal tool for comparing the distributions across three or more independent groups.

Within the `PROC NPAR1WAY` statement, several essential options must be clearly specified to ensure the correct analysis is performed. The option `data=fertilizer_data` directs the procedure to the dataset prepared in Step 1. The key option `WILCOXON` is included specifically to request the computation of the Kruskal-Wallis test statistic, which is derived from the generalized Wilcoxon rank-sum statistic. Furthermore, the inclusion of the `DSCF` option is highly recommended; this crucial option instructs SAS to perform [pairwise comparisons](#) between all group combinations, which are necessary if the overall Kruskal-Wallis test reveals a statistically significant result.

The subsequent statements define the variables for the analysis: `class fertilizer;` explicitly

identifies `fertilizer` as the classification variable, which partitions the data into the three independent groups. The statement `var growth;` designates `growth` as the dependent variable whose median values are being compared across the treatment groups. Finally, the `run;` statement executes the procedure, prompting SAS to generate the detailed output required for thorough interpretation.

```
/*perform Kruskal-Wallis test*/
proc npar1way data=fertilizer_data wilcoxon dscf;
class fertilizer;
var growth;
run;
```

Step 3: Analyzing the Overall Test Results (Chi-Squared Statistic)

Upon the successful execution of the SAS code, the output generated by `PROC NPAR1WAY` will present several tables containing the critical information needed to interpret the results of the Kruskal-Wallis Test. The initial and most important section of the output focuses on the overall test statistic, which is typically reported as the [Chi-Squared test statistic](#), along with its associated degrees of freedom and, most importantly, the corresponding [p-value](#).

The NPAR1WAY Procedure					
Wilcoxon Scores (Rank Sums) for Variable growth Classified by Variable fertilizer					
fertilizer	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
fert1	10	153.0	155.0	22.578089	15.30
fert2	10	205.0	155.0	22.578089	20.50
fert3	10	107.0	155.0	22.578089	10.70
Average scores were used for ties.					

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
6.2878	2	0.0431

In the results displayed above, the p-value for the overall Kruskal-Wallis test is calculated as **0.0431**. To determine the outcome of the study, we must compare this value against our predetermined level of significance (alpha, α), which is conventionally set at 0.05. Since the

calculated p-value of 0.0431 is less than 0.05, we must consequently reject the [null hypothesis](#). This decisive rejection provides strong statistical evidence to support the conclusion that the median plant growth is not identical across all three fertilizer types. Stated differently, the specific type of fertilizer utilized yields [statistically significant differences](#) in observed plant growth.

Step 4: Conducting Post-Hoc Pairwise Comparisons

While the overall Kruskal-Wallis Test successfully confirms that differences exist somewhere among the groups, it does not specify which particular pairs of groups are significantly different from one another. This necessitates a follow-up analysis known as post-hoc testing, which is precisely why we included the `DSCF` option in the `PROC NPAR1WAY` code. The subsequent output table, generated by the DSCF (Dwass, Steel, Critchlow-Fligner) option, presents the adjusted p-values for all possible [pairwise comparisons](#) between the three fertilizer groups, controlling for the increased risk of Type I error inherent in multiple testing.

The NPAR1WAY Procedure			
Pairwise Two-Sided Multiple Comparison Analysis			
Dwass, Steel, Critchlow-Fligner Method			
Variable: growth			
fertilizer	Wilcoxon Z	DSCF Value	Pr > DSCF
fert1 vs. fert2	-1.3700	1.9374	0.3567
fert1 vs. fert3	1.2220	1.7281	0.4401
fert2 vs. fert3	2.4402	3.4509	0.0390

A careful examination of this table reveals that only one pairwise comparison yields a p-value below our established 0.05 significance threshold: the comparison between Fertilizer 2 and Fertilizer 3, which resulted in a p-value of **0.0390**. This specific finding demonstrates a statistically significant difference in median plant growth between the plants treated with Fertilizer 2 and those treated with Fertilizer 3. Conversely, the comparisons involving Fertilizer 1 (Fertilizer 1 vs. Fertilizer 2, and Fertilizer 1 vs. Fertilizer 3) both show p-values greater than 0.05, indicating no statistically significant difference in median growth between these pairs. Therefore, the key finding is that Fertilizer 2 and Fertilizer 3 have a differential and significant effect on plant growth, whereas Fertilizer 1's impact does not statistically differ from either of the other two median growth rates.

Conclusion

This tutorial has provided a comprehensive, step-by-step guide on how to successfully execute

and interpret the Kruskal-Wallis Test within the SAS environment. By applying this rigorous non-parametric approach to our hypothetical agricultural study, we were able to confidently establish that the type of fertilizer significantly influences plant growth, with the specific difference localized to the comparison between Fertilizer 2 and Fertilizer 3. The KWT remains an indispensable analytical procedure for comparing multiple [independent groups](#), especially when the crucial assumptions of parametric tests, such as normality, cannot be adequately met. By following these outlined steps, researchers can confidently apply the Kruskal-Wallis Test to their own data, ensuring statistically sound and meaningful conclusions are drawn.

Additional Resources

To further advance your proficiency in statistical analysis using SAS, we encourage you to explore the following related tutorials, which detail the application of other common statistical tests: