

Learn How to Conduct Tukey's HSD Test in SAS: A Step-by-Step Guide

Authored by
Mohammed looti

November 14, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Conduct Tukey's HSD Test in SAS: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=1556>

Introduction: The Necessity of Post Hoc Analysis Following ANOVA

The [one-way ANOVA](#) (Analysis of Variance) is a foundational statistical tool used extensively across research disciplines. Its primary function is to determine whether significant differences exist among the means of three or more independent groups. Researchers rely on ANOVA as an essential screening procedure when comparing multiple treatment arms or distinct population segments to establish if the independent variable exerts any measurable effect on the continuous dependent variable.

Following the execution of an ANOVA, the immediate focus shifts to the overall [p-value](#) derived from the output table. If this probability value falls below the predefined [significance level](#) (alpha, conventionally set at 0.05), researchers possess sufficient evidence to reject the null hypothesis. This rejection confirms the important conclusion that at least one group mean is statistically different from the others within the population.

However, a significant overall ANOVA result is inherently non-specific. While it affirms that the group means are not all equal, it critically fails to identify which specific pairs of groups are responsible for driving this observed statistical difference. To precisely isolate and confirm the exact groups contributing to the variance, a subsequent, more granular level of analysis focusing on specific comparisons is mandatory.

To overcome this limitation and pinpoint the precise location of the differences, it is crucial to conduct a [post hoc test](#). These specialized tests are rigorously designed to perform systematic multiple comparisons between all possible group means while concurrently imposing strict control over the inflated risk of committing a **Type I error**--the false rejection of a true null hypothesis. This risk naturally increases exponentially as the number of simultaneous comparisons grows.

Tukey's Honestly Significant Difference (HSD) Test: A Controlled Approach

Among the repertoire of available statistical procedures for post hoc analysis, **Tukey's Honestly Significant Difference (HSD) Test** stands out as one of the most robust, reliable, and widely employed methods. It is the preferred technique for executing all possible [pairwise comparisons](#) between group means, provided the initial ANOVA has already established a statistically significant result. Tukey's HSD ensures that these multiple comparisons are made fairly and consistently under a controlled error environment.

The primary statistical strength of [Tukey's Test](#) lies in its powerful ability to regulate the [family-wise error rate](#) (FWER). This regulation is achieved by carefully adjusting the critical value required for significance, thereby mitigating the heightened probability of incorrectly identifying a difference as significant when numerous hypothesis tests are performed simultaneously across the family of comparisons. By ensuring a uniform and controlled error rate across the entire set of comparisons,

Tukey's HSD provides more trustworthy and definitive conclusions regarding specific group differentiations.

While this test is optimally suited for experimental designs where the sample sizes are equal across all groups (known as a **balanced design**), its methodology remains sufficiently flexible to accommodate unbalanced designs, albeit sometimes with a slight reduction in statistical power. Tukey's long-standing reputation for stringent error control and straightforward interpretability makes it indispensable for researchers who require clear, conservative, and statistically sound comparisons following the rejection of the overall ANOVA null hypothesis.

Preparing and Structuring Data for Analysis in SAS

To demonstrate the practical application of [Tukey's Test](#), we will utilize the powerful statistical software, [SAS](#). Consider a hypothetical study involving 24 students, where each participant is randomly assigned to one of three distinct studying methodologies: Method A, Method B, or Method C.

The core objective of this study is to definitively ascertain whether a measurable and significant difference exists in the final exam scores achieved by students based on the methodology they employed. The resulting exam scores are systematically collected and prepared in a two-column format: the studying method serves as the categorical independent variable, and the numerical score serves as the continuous response variable, making the data ready for immediate statistical processing.

Method A	Method B	Method C
78	81	84
81	83	88
82	83	88
82	85	89
85	86	90
88	88	93
88	90	95
90	91	98

To commence the analysis within the [SAS](#) environment, the essential first step is the creation and population of a data structure that faithfully represents the study's observations. The following SAS data step code snippet illustrates the precise definition and input process, explicitly specifying 'Method' as a character variable and 'Score' as the numerical measure.

```
/*create dataset*/  
data my_data;  
input Method $ Score;  
datalines;  
A 78  
A 81  
A 82  
A 82  
A 85  
A 88  
A 88  
A 90  
B 81  
B 83  
B 83  
B 85  
B 86  
B 88  
B 90  
B 91  
C 84  
C 88  
C 88  
C 89  
C 90  
C 93  
C 95  
C 98  
;  
run;
```

In this foundational data preparation block, each command serves a specific analytical purpose:

`data my_data;` initiates the procedure to construct a new SAS dataset, which we have named `my_data`.

`input Method $ Score;` defines the variable structure. The dollar sign (\$) following `Method` designates it explicitly as a **character variable** (the categorical factor). `Score` is defined as the **numeric variable** (the continuous outcome).

`datalines;` is a procedural instruction informing SAS that the raw data values immediately follow

within the program script.

The subsequent lines contain the actual subject-level data, linking each student's studying method to their respective exam score.

`run;` is the final execution command, compiling and saving the structured `my_data` dataset for all subsequent statistical procedures.

Executing One-Way ANOVA and Tukey's Test in SAS

With the dataset successfully defined and populated, the next essential step involves executing the [one-way ANOVA](#). Within the [SAS](#) environment, this analysis is most efficiently managed using the `PROC ANOVA` procedure. This specialized procedure is adept at handling the general linear model specification and, crucially for our purpose, allows for the seamless integration of post hoc tests, such as Tukey's HSD, directly within the primary analysis command structure.

The following SAS code block provides the precise syntax required to execute the primary one-way ANOVA and simultaneously trigger the request for [Tukey's Test](#). This integrated approach is highly recommended as it streamlines the analytical workflow and ensures that the post hoc comparison is executed using the correct error term derived from the initial ANOVA F-test.

```
/*perform one-way ANOVA*/  
proc ANOVA data=my_data;  
class Method;  
model Score = Method;  
means Method / tukey cldiff;  
run;
```

A detailed understanding of the parameters within the `PROC ANOVA` command is vital for grasping the model specification:

`proc ANOVA data=my_data;` invokes the ANOVA statistical procedure, explicitly specifying the input data source as the `my_data` dataset.

`class Method;` declares 'Method' as a **classification variable** (or factor). This tells SAS that 'Method' consists of discrete categories (A, B, C) whose group means are to be compared.

`model Score = Method;` specifies the linear model structure. 'Score' is designated as the outcome or **dependent variable**, modeled by the independent categorical variable, 'Method'.

`means Method / tukey cldiff;` is the command requesting detailed mean comparisons. The `tukey` option executes [Tukey's HSD post hoc test](#). The `cldiff` option further requests the calculation and display of [confidence intervals](#) for the differences between every pair of group means.

For maintaining statistical rigor, it is important to note that SAS's `PROC ANOVA` typically executes the post hoc test conditionally. The detailed results for [Tukey's Test](#) will only be generated and displayed in the output if the overall [p-value](#) from the primary ANOVA indicates [statistical significance](#), thereby justifying the subsequent pairwise investigation.

Interpreting the Initial ANOVA Results

Upon the successful execution of the `PROC ANOVA` code, the initial analytical focus must immediately center on the ANOVA Summary Table, which forms the foundation of the SAS output. This table provides the critical statistical evidence necessary to determine whether the variation observed between the group means is a genuine effect of the studying methods or merely attributable to random chance fluctuations.

The ANOVA Procedure
Dependent Variable: Score

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	175.5833333	87.7916667	5.26	0.0140
Error	21	350.2500000	16.6785714		
Corrected Total	23	525.8333333			

R-Square	Coeff Var	Root MSE	Score Mean
0.333914	4.698685	4.083941	86.91667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Method	2	175.5833333	87.7916667	5.26	0.0140

From the summarized ANOVA table displayed above, two specific metrics are extracted for formal analysis:

The calculated [F-value](#) is reported as **5.26**. This F-statistic quantifies the ratio of the variance explained by the differences between the studying methods (Method) to the unexplained, random variation within those methods (Error).

The corresponding [p-value](#) for this F-value is **0.0140**. This p-value represents the probability of observing our data (or data more extreme) if, hypothetically, there were absolutely no true differences between the population means.

To properly contextualize these findings, we review the framework of [hypothesis testing](#) governing the one-way ANOVA:

H0 (Null Hypothesis): All population group means are mathematically equivalent. (e.g., Mean score for Method A = Mean score for Method B = Mean score for Method C).

HA (Alternative Hypothesis): At least one population group mean is statistically different from the others. (e.g., Not all mean scores for the studying methods are equal).

In our scenario, the calculated [p-value](#) of 0.0140 must be compared against the standard pre-selected [significance level](#), typically set at $\alpha = 0.05$. Since 0.0140 is unequivocally smaller than 0.05, we possess sufficient statistical evidence to confidently reject the [null hypothesis](#). This decisive rejection confirms a finding of overall [statistical significance](#), validating the need to proceed to the next analytical stage: interpreting the post hoc test results to precisely localize which method pairs differ.

Decoding the Pairwise Differences via Tukey's HSD Results

With the overall significance firmly established by the ANOVA F-test, our investigation now transitions entirely to the detailed output generated by the Tukey's HSD procedure. This segment of the output provides the core evidence necessary for drawing specific conclusions, presenting a comprehensive list of all possible [pairwise comparisons](#) between the group means. This allows us to accurately identify which studying methods resulted in statistically distinct average exam scores.

To systematically isolate these differences, we must consult the final table produced by SAS, typically titled 'Tukey's Studentized Range (HSD) Test'. This structured table organizes the comparisons and provides the adjusted p-values that inherently account for the multiple testing environment, ensuring that the FWER is maintained at the acceptable alpha level.

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for Score

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	21
Error Mean Square	16.67857
Critical Value of Studentized Range	3.56462
Minimum Significant Difference	5.1469

Comparisons significant at the 0.05 level are indicated by ***.

Method Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
C - B	4.750	-0.397	9.897	
C - A	6.375	1.228	11.522	***
B - C	-4.750	-9.897	0.397	
B - A	1.625	-3.522	6.772	
A - C	-6.375	-11.522	-1.228	***
A - B	-1.625	-6.772	3.522	

The most straightforward method for quickly discerning statistically meaningful differences is by examining the indicators of [statistical significance](#), which in SAS output are often represented by numerical groupings or asterisks (***) adjacent to the p-value or the difference measure. These markers highlight the comparisons where the observed difference between the means is large enough to be considered significant at the chosen alpha level of 0.05.

Analysis of the provided Tukey HSD output reveals a clear and definitive pattern: a statistically significant difference in mean exam scores exists exclusively between **Group A** and **Group C**. This robust finding suggests that students utilizing studying Method C achieved significantly higher or lower scores compared to those who used Method A.

Conversely, the table confirms that all other possible pairs--specifically the comparison between Group A and Group B, and the comparison between Group B and Group C--exhibit no statistically significant differences. This implies that while the overall methods differ significantly, the primary driving force behind that overall variance is the performance gap between Method A and Method C. This granular, controlled insight, which is unattainable solely from the initial ANOVA, represents the central and most valuable outcome of properly employing Tukey's HSD Test.

Conclusion: Mastering Controlled Multiple Comparisons

This guide has provided a comprehensive, step-by-step walkthrough of implementing and interpreting [Tukey's Test](#) within the [SAS](#) statistical software environment. Our journey spanned from the essential initial definition of the dataset to the final, precise interpretation of the pairwise differences. We began by establishing the conceptual necessity of performing [post hoc tests](#) subsequent to a significant [one-way ANOVA](#), emphasizing that ANOVA only signals a general difference, not the specific sources of variation.

We meticulously detailed the required data preparation steps in SAS, followed by the specific execution of the `PROC ANOVA` procedure, highlighting the critical inclusion of the `tukey` and `clidiff` options. The subsequent analysis of the ANOVA table confirmed the overall effect, thereby authorizing the detailed examination of Tukey's output. Our final and most crucial step involved careful scrutiny of the [pairwise comparisons](#), which successfully isolated the significant performance difference exclusively between studying Method A and Method C.

Adopting this systematic, two-stage analytical strategy guarantees that conclusions drawn from multi-group comparisons are statistically robust, valid, and effectively control the critical family-wise error rate. Proficiently applying and interpreting [Tukey's Test](#) remains an indispensable, advanced skill for any professional researcher engaged in rigorous experimental data analysis.

For those interested in delving deeper into ANOVA models, post hoc procedures, and related statistical concepts, the following resources offer additional valuable information and tutorials:

[A Guide to Using Post Hoc Tests with ANOVA](#)