

# Perform a One-Way ANOVA in SAS

Authored by  
**Mohammed looti**

November 1, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Perform a One-Way ANOVA in SAS*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=7661>

The [one-way ANOVA](#) (Analysis of Variance) is a fundamental statistical technique used extensively across various scientific disciplines, including psychology, biology, and engineering. Its primary function is to determine whether there is a [statistically significant difference](#) between the means of three or more independent, unrelated groups. This method is crucial when researchers need to compare outcomes across different treatment levels or categories, ensuring that any observed differences are likely due to the intervention rather than random chance. Understanding the underlying assumptions, such as the normality of residuals and the homogeneity of variances, is paramount for the valid application and interpretation of this test.

This comprehensive guide provides an expert, step-by-step tutorial demonstrating how to execute a [one-way ANOVA](#) using the powerful statistical software environment, [SAS](#) (Statistical Analysis System). We will walk through the entire process, starting from the meticulous creation of the sample dataset, proceeding through the specific SAS procedures required for the analysis, and concluding with a rigorous interpretation of the resulting output tables and figures. This structured approach ensures clarity and reproducibility, making complex statistical analysis accessible to analysts and researchers alike.

The scenario we will analyze involves a study investigating the impact of different studying methods on student performance. By using [SAS](#), we can efficiently test the central hypothesis: that the average exam scores differ significantly depending on which of the three assigned studying methods the students utilized. Mastering the workflow presented here is essential for anyone requiring reliable comparison of means across multiple [independent groups](#) in an academic or professional setting.

## Step 1: Defining the Research Problem and Creating the Data

To illustrate the application of the [one-way ANOVA](#), consider a practical research design where an investigator aims to evaluate the effectiveness of three distinct studying methods (labeled A, B, and C) on final exam performance. The study involves recruiting 30 participants, who are subsequently allocated to one of these three methods. This experimental design ensures that the groups are [independent groups](#), a core requirement for the ANOVA test. The resulting exam score data, representing the quantitative outcome variable, must be structured appropriately within [SAS](#) for successful analysis.

The preliminary step in any [SAS](#) analysis involves creating a data structure that clearly defines the experimental conditions and the observed measurements. In our case, this structure requires two variables: the categorical independent variable, **Method** (A, B, or C), and the continuous dependent variable, **Score** (the exam result). The raw data collected from the 30 students, which we will transcribe into SAS code, reflects the performance outcomes associated with each assigned studying technique.

The specific exam results for each participant, categorized by their assigned studying method, are summarized visually below. Careful inspection of this raw data suggests potential differences in performance, particularly noticing the higher scores achieved in Method C, which motivates the need for a formal statistical test like ANOVA to confirm these initial observations.

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

We translate this tabular data into a functional [SAS](#) dataset using the following **DATA step**. This code initializes the dataset named **my\_data**, specifies the input variables using the **INPUT** statement, and then utilizes the **DATALINES** statement to input the raw observations directly into the program. This process is foundational for ensuring the data is correctly formatted for subsequent statistical procedures.

```
/*create dataset: Defining the structure for the ANOVA analysis*/
```

```
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;
```

## Step 2: Executing the One-Way ANOVA Procedure in SAS

Once the data is successfully created and stored in the **my\_data** dataset, the next critical step is invoking the appropriate statistical procedure to conduct the analysis. In [SAS](#), the Analysis of Variance is typically performed using the [PROC ANOVA](#) statement. This procedure is specifically designed for analyzing experimental data where the independent variable is categorical and the dependent variable is continuous, perfectly matching the requirements of our study comparing three studying methods.

The core syntax within the [PROC ANOVA](#) block requires several key statements. First, the **CLASS** statement is used to identify the categorical variable--in this case, **Method**--which defines the groups we are comparing. Second, the **MODEL** statement formally specifies the structure of the statistical relationship, defining the dependent variable (**Score**) as a function of the independent variable (**Method**). These two statements are essential for instructing [SAS](#) on how to partition the total variance.

Furthermore, we incorporate the **MEANS** statement, which is vital for performing follow-up analyses if the initial ANOVA test yields a [statistically significant difference](#). We utilize the **TUKEY** option within the **MEANS** statement to request [Tukey's Honestly Significant Difference \(HSD\) Test](#). This **post-hoc** procedure controls the family-wise error rate and is used to pinpoint exactly which pairs of group means differ significantly. The **CLDIFF** option is also included to output the confidence intervals for the mean differences, enhancing the depth of our reporting.

```
/*perform one-way ANOVA: Applying the PROC ANOVA procedure*/  
proc ANOVA data=my_data;
```

```

class Method;
model Score = Method;
means Method / tukey cldiff;
run;

```

It is important to emphasize the utility of the **MEANS** statement in conditional testing. By including **TUKEY**, we ensure that if the global test indicates that the group means are not all equal, **SAS** automatically proceeds to conduct the pairwise comparison tests. This streamlined approach prevents unnecessary comparisons if the overall test fails to find a significant effect, maintaining statistical rigor and efficiency in the analysis process.

### Step 3: Interpreting the ANOVA and Post-Hoc Results

The initial output generated by **PROC ANOVA** focuses on the ANOVA summary table, which is the cornerstone for determining whether the overall effect of the independent variable is significant. This table summarizes how the total variance in the scores is partitioned between the differences among the group means (the Model or Treatment effect) and the variability within the groups (the Error or Residual effect). The key metrics to extract from this table are the F statistic and its corresponding **p-value**.

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

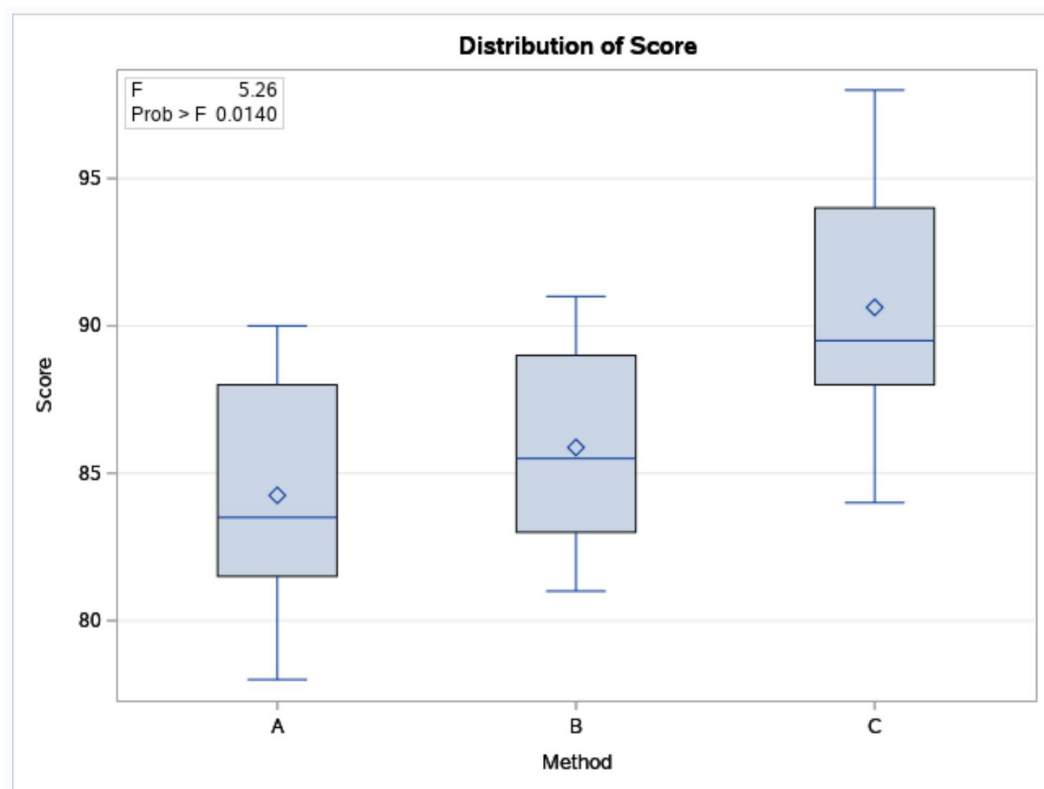
The interpretation hinges on comparing the calculated **p-value** against a predetermined significance level, typically  $\alpha = 0.05$ . The statistical hypotheses under consideration are:

**H0:** The **Null Hypothesis** states that all population group means are equal ( $\mu_A = \mu_B = \mu_C$ ).

**HA:** The Alternative Hypothesis states that at least one group mean is different from the others ( $\mu_i \neq \mu_j$  for some  $i, j$ ).

Upon reviewing the ANOVA table output, we observe the following critical values: The overall **F Value** is 5.26, and the corresponding **p-value** is 0.0140. Since 0.0140 is less than the standard significance threshold of  $\alpha = 0.05$ , we confidently reject the **null hypothesis**. This decision leads to the conclusion that there is a **statistically significant difference** in the mean exam scores across the three studying methods, necessitating further investigation using post-hoc tests to localize this difference.

Beyond the numerical tables, **SAS** often provides graphical output, such as **boxplots**, which offer an intuitive visualization of the data distribution for each group. These visual aids are crucial for confirming the pattern suggested by the ANOVA results.



The boxplots confirm that the distribution of exam scores for Method C appears visually shifted higher compared to Method A and Method B. While Method B also seems slightly higher than Method A, the most pronounced difference appears between A and C. This visual evidence supports the statistical finding that the studying methods do not lead to equivalent mean outcomes.

## Step 4: Analyzing Tukey's HSD Post-Hoc Comparisons

Because the overall [one-way ANOVA](#) indicated a significant difference, we must now turn to the results of [Tukey's HSD Test](#), which was requested using the **TUKEY** option in the **MEANS** statement. This **post-hoc** test is specifically designed for all pairwise comparisons, rigorously controlling the risk of Type I errors that arise when conducting multiple comparisons simultaneously. The output provides a detailed matrix showing the difference between every possible pair of group means.

### 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	

To identify the pairs that exhibit a [statistically significant difference](#), we examine the output for indicators of significance, typically marked by asterisks (\*\*\*) or by checking if the [p-value](#) for that specific comparison is below 0.05. The table clearly highlights the comparison between Method A and Method C. This finding confirms that the observed effectiveness of the studying methods lies in the significant superiority of Method C over Method A in terms of generating higher exam scores.

For the statistically significant pair (A vs. C), the output also includes the [95% Confidence Interval](#) for the difference in means, which provides a range of plausible values for the true population

difference. The observed 95% [Confidence Interval](#) for the difference in mean exam scores between Group A and Group C is reported as . Since this interval does not contain zero, it reinforces the conclusion that the mean scores of these two groups are reliably different in the population. The positive interval suggests that Method C yields scores that are, on average, between 1.228 and 11.522 points higher than Method A.

Conversely, the comparisons between Method A and Method B, and between Method B and Method C, do not show significance (i.e., they lack asterisks or have p-values greater than 0.05). This implies that while Method C is significantly better than A, the differences in performance between A and B, or B and C, are not strong enough to be deemed [statistically significant](#) given the sample size and variability.

## Step 5: Formal Reporting of ANOVA Results

The final stage of the analysis involves synthesizing the findings into a clear, concise, and formally accepted statistical report. When reporting the results of a [one-way ANOVA](#), it is standard practice to include the overall F statistic, the degrees of freedom (df), and the [p-value](#). This is followed by the results of the **post-hoc** tests, detailing which specific pairs were found to be different, often including the associated [confidence interval](#) for the mean difference.

The report should systematically address the primary research question regarding the effect of the studying methods on exam performance. We begin by stating the purpose of the test and the overall outcome based on the ANOVA table:

A [one-way ANOVA](#) was performed to compare the effect of three different studying methods (A, B, and C) on student exam scores.

The analysis revealed that there was a [statistically significant difference](#) in mean exam score between at least two of the groups,  $F(2, 21) = 5.26$ ,  $p = 0.014$ .

Following the global test results, we must detail the findings from the [Tukey's HSD Test](#), which provided the critical pairwise comparisons. This ensures that the reader understands the precise location of the significant effect identified by the ANOVA.

[Tukey's HSD Test](#) for multiple comparisons confirmed that the mean exam score achieved using method C was significantly higher than the mean score achieved using method A (95% [C.I.](#) = ).

Crucially, there was no [statistically significant difference](#) observed in mean exam scores between method A and method B ( $p > 0.05$ ), nor was there a significant difference between method B and method C ( $p > 0.05$ ).

This structured reporting concludes the analysis, providing robust evidence that the choice of

studying method significantly impacts performance, specifically highlighting Method C's advantage over Method A. This level of detail ensures that the findings are transparent, reproducible, and actionable for the researcher.

## Additional Resources for Statistical Analysis in SAS

For readers seeking to deepen their understanding of variance analysis and the capabilities of [SAS](#), exploring related statistical procedures is highly recommended. While [PROC ANOVA](#) is ideal for balanced designs, researchers working with unbalanced data or requiring more complex model specifications should familiarize themselves with **PROC GLM** (General Linear Models), which offers greater flexibility for various factorial and regression designs.

Understanding the assumptions underlying the [one-way ANOVA](#), such as the assumption of homogeneity of variances (tested via Levene's test or Brown-Forsythe test) and the normality of residuals, is vital for ensuring the validity of the results. If these assumptions are severely violated, alternative non-parametric tests, such as the Kruskal-Wallis H test, should be considered as robust substitutes.

The following resources provide additional information and tutorials that complement this guide on performing and interpreting [one-way ANOVA](#) analyses: