

# Perform a Two-Way ANOVA in SAS

Authored by  
**Mohammed looti**

November 1, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Perform a Two-Way ANOVA in SAS*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=7656>

## Understanding the Two-Way ANOVA Framework

The [two-way ANOVA](#) (Analysis of Variance) represents a cornerstone of statistical methodology, particularly within experimental research. This powerful technique is employed when researchers aim to simultaneously evaluate the influence of two distinct independent categorical variables, often referred to as factors, on a single continuous dependent variable. Unlike the simpler one-way ANOVA, this model provides the critical capability to assess three specific effects: the two individual main effects of each factor, and the crucial **interaction effect** between them.

The core objective of performing this analysis is to rigorously determine whether there is a [statistically significant difference](#) among the group means resulting from the crossing of these two factors. By modeling both factors concurrently, the two-way ANOVA provides an efficient and comprehensive approach to test multiple hypotheses within a single, integrated framework. This tutorial offers a meticulous, step-by-step guide designed to help you successfully execute and interpret a two-way ANOVA using the industry-standard [SAS statistical software suite](#).

It is paramount to remember that the reliability and validity of ANOVA conclusions depend heavily on meeting several underlying statistical assumptions. These key requirements include the normal distribution of residuals, the **homogeneity of variances** across all defined groups (homoscedasticity), and the fundamental independence of observations. While ANOVA is often noted for its robustness against minor violations, careful consideration and diagnostic checks of these assumptions are essential for ensuring the integrity of the results drawn from the model.

### Step 1: Defining the Experiment and Creating the SAS Dataset

To provide a clear, practical illustration of the procedure, we utilize a classic experimental design rooted in botany. Imagine a botanist investigating how **plant growth** (measured by the final height in inches, serving as the dependent variable) is influenced by two distinct categorical factors: the frequency of watering and the intensity of sunlight exposure. These two manipulated factors define the fixed effects in our statistical model.

The experiment involved planting 30 identical seeds and allowing them to grow for a fixed period of one month under strictly controlled conditions. The experimental design involved crossing the two factor levels: Watering Frequency (Daily or Weekly) and Sunlight Exposure (Low, Medium, or High). After the growth period concluded, the final height of each individual plant was meticulously recorded. This structured organization of data is the prerequisite for accurate statistical analysis.

The collected raw data, detailing the plant height under all combinations of watering and sunlight conditions, is visually represented in the table below. This structure must be carefully mapped into the SAS input format, which requires three distinct variables: the two categorical factors (`water` and `sunlight`) and the continuous quantitative response measure (`height`).

	Sunlight Exposure		
Watering Frequency	Low	Medium	High
Daily	6	5	6
	6	5	6
	6	6	7
	5	4	8
	6	5	7
Weekly	3	4	5
	4	4	6
	4	4	6
	4	4	7
	5	4	8

The following **SAS code block** is used to define and populate the dataset, which we name `my_data`. We employ the `DATALINES` statement to input the raw observations directly into the program. It is crucial to note the inclusion of the dollar sign (\$) following the factor names in the `INPUT` statement, which formally designates these variables (`water` and `sunlight`) as character variables, identifying them as categorical factors.

```
/*create dataset*/
data my_data;
input water $ sunlight $ height;
datalines;
daily low 6
daily low 6
daily low 6
daily low 5
daily low 6
daily med 5
daily med 5
daily med 6
daily med 4
daily med 5
daily high 6
daily high 6
daily high 7
daily high 8
daily high 7
```

```
weekly low 3
weekly low 4
weekly low 4
weekly low 4
weekly low 5
weekly med 4
weekly med 4
weekly med 4
weekly med 4
weekly med 4
weekly high 5
weekly high 6
weekly high 6
weekly high 7
weekly high 8
;
run;
```

## Step 2: Executing the Two-Way ANOVA Using PROC ANOVA

To execute the two-way ANOVA in the SAS environment, we utilize the specialized [PROC ANOVA](#) procedure. This procedure is optimally designed for analyzing experimental data, particularly when the design is **balanced**--meaning there are equal sample sizes in all treatment cells defined by the crossing of the factors. For experimental designs that are unbalanced or involve more complex structures, `PROC GLM` (the [General Linear Model](#) procedure) would be the recommended, more flexible alternative.

The fundamental components of the SAS syntax are the `CLASS` and `MODEL` statements. The `CLASS` statement is mandatory; it explicitly identifies the categorical variables--`water` and `sunlight`--that define the distinct groups or treatments. The `MODEL` statement then specifies the precise statistical relationship being tested, listing the dependent variable (`height`) followed by the independent terms: the two main effects and the absolutely crucial interaction term (`water*sunlight`).

Furthermore, we include the `MEANS` statement to request essential follow-up tests, which become necessary when significant main effects are detected. For these post-hoc comparisons, we specify the [Tukey HSD adjustment](#). The Tukey method is highly favored in statistical practice because it rigorously controls the family-wise error rate, thereby significantly reducing the risk of finding spurious significant results when numerous pairwise comparisons are performed. The optional `clidiff` parameter is added to request confidence intervals for the mean differences, which greatly

assists in providing a meaningful, substantive interpretation of the findings.

```
/*perform two-way ANOVA*/
proc ANOVA data=my_data;
class water sunlight;
model height = water sunlight water*sunlight;
means water sunlight / tukey cldiff;
run;
```

### Step 3: Interpreting Main Effects and Interaction Effects

The initial and most critical step in interpreting the SAS output involves a careful examination of the ANOVA Source Table. This table concisely summarizes how the total observed variance in the dependent variable (plant height) is partitioned among the two factors, their interaction, and the residual error term. It provides the necessary F-statistics and associated probabilities required to test our defined null hypotheses regarding the main and interaction effects.

Statistically, interpretation must always commence with the [interaction effect](#). If the interaction term proves to be statistically significant, the main effects cannot be interpreted independently; instead, the researcher must proceed to a simple main effects analysis to understand the influence of one factor at specific levels of the other. The provided visual output below displays the core results.

The ANOVA Procedure					
Dependent Variable: height					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	35.86666667	7.173333333	13.45	<.0001
Error	24	12.80000000	0.533333333		
Corrected Total	29	48.66666667			

  

R-Square	Coeff Var	Root MSE	height Mean
0.736986	13.69306	0.730297	5.333333

  

Source	DF	Anova SS	Mean Square	F Value	Pr > F
water	1	8.533333333	8.533333333	16.00	0.0005
sunlight	2	24.86666667	12.433333333	23.31	<.0001
water*sunlight	2	2.466666667	1.233333333	2.31	0.1207

Focusing on the [p-value](#) (labeled Pr > F) for each source of variation in the table, we derive the following conclusions:

The p-value for the `water` main effect is **.0005**. Since this value is considerably lower than the conventional  $\alpha = 0.05$  significance threshold, we confidently reject the null hypothesis, concluding that the frequency of watering significantly influences the ultimate plant height.

The p-value for the `sunlight` main effect is **<.0001**. This finding indicates a highly significant effect, unequivocally confirming that sunlight exposure is a potent and significant predictor of plant height.

The p-value for the crucial interaction term (`water*sunlight`) is **.1207**. As this value exceeds 0.05, we fail to reject the null hypothesis for the interaction.

The non-significant interaction is a highly valuable piece of information: it robustly confirms that the specific effect of watering (whether daily or weekly) remains statistically consistent across all three levels of sunlight exposure (Low, Medium, High), and conversely. This demonstrated lack of interdependence between the factors simplifies the subsequent interpretation, allowing us to proceed directly to evaluating the overall differences observed in the significant main effects.

#### **Step 4: Detailed Post-Hoc Analysis using Tukey's HSD**

Once significant main effects have been identified, post-hoc tests are required to precisely locate the source of the differences among the means of the factor levels. We specifically requested the [Tukey-Kramer Honestly Significant Difference \(HSD\) test](#), which is ideally suited for controlling error rates during these multiple pairwise comparisons.

We begin by examining the output generated for the `water` factor. Since this factor has only two levels (Daily vs. Weekly), the Tukey test effectively serves to confirm the precise nature of the significant difference previously identified in the main effect test.

### The ANOVA Procedure

#### Tukey's Studentized Range (HSD) Test for height

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

Alpha	0.05
Error Degrees of Freedom	24
Error Mean Square	0.533333
Critical Value of Studentized Range	2.91879
Minimum Significant Difference	0.5504

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

water Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
daily - weekly	1.0667	0.5163	1.6170	***
weekly - daily	-1.0667	-1.6170	-0.5163	***

The SAS output for watering frequency yields the following critical statistical information:

The calculated mean difference in height between plants watered daily and those watered weekly is **1.0667 inches**.

The 95% confidence interval (C.I.) for this true difference in mean height is . Because this interval does not encompass zero, the difference is considered statistically robust, definitively confirming that daily watering leads to significantly taller plants compared to weekly watering.

Next, we analyze the results for the `sunlight` factor, which comprises three levels (Low, Medium, High). The Tukey procedure systematically performs all three unique pairwise comparisons (High vs. Low, High vs. Medium, and Medium vs. Low) to pinpoint the source of the overall significant effect detected in Step 3.

## The ANOVA Procedure

## Tukey's Studentized Range (HSD) Test for height

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

Alpha	0.05
Error Degrees of Freedom	24
Error Mean Square	0.533333
Critical Value of Studentized Range	3.53170
Minimum Significant Difference	0.8156

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

sunlight Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
high - low	1.7000	0.8844	2.5156	***
high - med	2.1000	1.2844	2.9156	***
low - high	-1.7000	-2.5156	-0.8844	***
low - med	0.4000	-0.4156	1.2156	
med - high	-2.1000	-2.9156	-1.2844	***
med - low	-0.4000	-1.2156	0.4156	

To determine statistical significance within this output, we look for the presence of asterisks (\*\*\*), which conventionally denote a p-value below 0.05. The results clearly indicate that two of the three possible pairwise comparisons are statistically significant:

**High sunlight vs. Low sunlight:** This comparison shows a significant difference. The 95% C.I. for the difference ranges from **0.8844 to 2.5156**.

**High sunlight vs. Medium sunlight:** This comparison is also strongly significant. The 95% C.I. ranges from **1.2844 to 2.9156**.

In contrast, the comparison between Medium sunlight and Low sunlight exposure did not achieve statistical significance. We therefore conclude that while high sunlight exposure demonstrably promotes greater growth when compared to both low and medium levels, there is no statistically distinguishable difference in mean plant height between the medium and low exposure groups.

## Step 5: Formal Reporting of Statistical Findings

The final stage of the analysis is the formal, clear presentation of the results. This summary synthesizes the key findings derived from the ANOVA Source Table and the detailed post-hoc tests, providing a narrative description of the effects discovered. Formal reporting should strictly follow standardized statistical guidelines, including the F-statistic, the degrees of freedom (df), and the precise p-value for each test conducted.

The report below adheres to standard reporting guidelines, ensuring clarity and completeness. It strategically addresses the interaction effect first, before detailing the significant main effects and their corresponding [Tukey HSD](#) post-hoc results:

A two-way ANOVA was performed to analyze the effect of watering frequency (Daily vs. Weekly) and sunlight exposure (Low, Medium, High) on the dependent variable of plant growth (height).

The analysis revealed that there was **not a statistically significant interaction** between the two factors,  $F(2, 24) = 2.29$ ,  $p = .1207$ . This crucial finding indicates that the effect of watering frequency on height remained consistent regardless of the specific level of sunlight exposure.

The main effect of watering frequency was **statistically significant**,  $F(1, 24) = 14.50$ ,  $p = .0005$ . Post-hoc analysis (Tukey HSD) showed that plants watered daily were significantly taller than plants watered weekly ( $M_{\text{diff}} = 1.07$  inches, 95% CI: ).

The main effect of sunlight exposure was also **highly statistically significant**,  $F(2, 24) = 19.98$ ,  $p < .0001$ . [Tukey HSD](#) comparisons indicated that High sunlight exposure resulted in significantly greater plant height compared to both Low and Medium exposure groups. Crucially, no statistically significant difference was found between the Medium and Low sunlight groups.

## Advanced Techniques and Further Resources

For users seeking to expand their knowledge beyond the basic fixed-factor two-way ANOVA, SAS provides sophisticated procedures capable of handling substantially more complex experimental designs. These advanced analyses include designs involving **repeated measures**, intricate nested factors, and the inclusion of continuous covariates (ANCOVA). A strong understanding of these advanced analytical techniques is essential for applying the full power of ANOVA to real-world research problems, particularly those involving longitudinal data collection or the necessity of controlling for nuisance variables.

The following resources offer additional information regarding two-way ANOVAs and related statistical methods, enabling deeper exploration of these topics: