

How to Easily Perform a Two-Way ANOVA in SAS

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform a Two-Way ANOVA in SAS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103431>

The two-way ANOVA (Analysis of Variance) is a powerful statistical procedure designed to assess how two independent categorical variables, often termed factors, simultaneously influence a single continuous dependent variable. Unlike the one-way ANOVA, this test allows researchers to evaluate not only the main effect of each factor but also their combined influence through an interaction effect. Mastering this procedure in **SAS** is essential for analysts working with experimental or observational data where multiple influences are present.

Executing a two-way ANOVA in SAS involves specific steps, including data preparation, defining the statistical model, and interpreting the detailed output, which typically includes the formal ANOVA table, least squares means, and post-hoc comparison results. This comprehensive tutorial will guide you through a practical, step-by-step example using the **proc ANOVA** procedure.

Defining the Research Scenario and Data Structure

To demonstrate the application of the two-way ANOVA, we will use a classic experimental scenario. Imagine a botanist investigating the impact of two distinct factors--sunlight exposure and watering frequency--on plant growth. The goal is to determine if either or both factors, or their combination, significantly influence the final height of the plants.

The botanist conducts an experiment using 30 seeds, allowing them to grow for one month under varying conditions defined by two factors: **Watering Frequency** (Daily vs. Weekly) and **Sunlight Exposure** (Low, Medium, High). After the experimental period, the height of each plant is measured, providing the continuous dependent variable for our analysis.

The resulting dataset is structured with three key variables: `water` (categorical factor 1), `sunlight` (categorical factor 2), and `height` (continuous dependent variable). This structure is ideal for the two-way ANOVA, which assesses the variability in height attributable to these two factors.

Preparing the Data in SAS

The initial requirement for analysis in **SAS** is the creation of a valid dataset reflecting the experimental observations. The following code uses the `DATA` step and `DATALINES` statement to input the variables `water`, `sunlight`, and `height` for all 30 observations. Note that the categorical variables (factors) are defined as character variables using the dollar sign (`$`).

The raw results collected by the botanist are summarized in the following visual representation:

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

We implement this dataset in the SAS environment using the following code block:

```
/*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;
```

Specifying the ANOVA Model using PROC ANOVA

Once the data is successfully loaded, the primary tool for executing the Analysis of Variance in SAS is the **proc ANOVA** procedure. This procedure requires precise specification of the factors and the statistical model to be tested. The `CLASS` statement identifies the categorical independent variables, and the `MODEL` statement defines the relationship between the dependent variable and the factors.

The model statement syntax `height = water sunlight water*sunlight` instructs SAS to test three distinct effects: the main effect of `water`, the main effect of `sunlight`, and the interaction effect between `water` and `sunlight`. Analyzing the interaction term (`water*sunlight`) is the unique and critical capability of the two-way ANOVA.

Furthermore, we include the `MEANS` statement to request post-hoc comparisons. By specifying `tukey cldiff`, we request the Tukey post-hoc test for both factors, which is essential for determining exactly which group means differ significantly if the overall main effect is found to be significant.

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

Analyzing the Overall ANOVA Results

Upon execution of the PROC ANOVA step, SAS generates a series of output tables. The most critical table for initial assessment is the primary Analysis of Variance table, which summarizes the sources of variation and their associated *p*-values. This table allows us to immediately assess the statistical significance of the two main effects and the interaction term.

We focus on the F-tests and associated *p*-values ($Pr > F$) to draw conclusions about the null hypotheses. The **null hypothesis** states that the means across the levels of a factor (or combination of factors) are equal. If the *p*-value is below the predetermined significance level (e.g., $\alpha = 0.05$), we reject the null hypothesis.

The resulting ANOVA table from our plant growth experiment is shown below:

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	12.43333333	23.31	<.0001
water*sunlight	2	2.46666667	1.233333333	2.31	0.1207

Analysis of this output reveals the following critical significance levels:

The *p*-value for the **Water** factor is **.0005**.

The *p*-value for the **Sunlight** factor is **<.0001**.

The *p*-value for the **Water * Sunlight** interaction is **.1207**.

Since the *p*-values for both *water* (.0005) and *sunlight* (<.0001) are less than 0.05, we conclude that both watering frequency and sunlight exposure are statistically significant predictors of plant height (main effects). Conversely, the interaction *p*-value (.1207) is greater than 0.05, indicating

that there is no statistically significant interaction effect; the effect of watering does not depend on the level of sunlight exposure, and vice versa.

Interpreting the Post-Hoc Comparisons: Factor 1 (Watering)

A statistically significant main effect only tells us that differences exist somewhere among the group means; it does not specify which pairs of groups are different. Since the `water` factor has a significant main effect, we must examine the **Tukey post-hoc test** output to perform pairwise comparisons between its levels (Daily vs. Weekly).

The Tukey procedure controls the overall Type I error rate when conducting multiple comparisons. The following table provides the results of the Tukey analysis for the watering frequency factor:

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 output highlights the comparison between plants watered Daily and those watered Weekly. The mean difference in plant height is calculated as **1.0667** inches. This confirms that, on average, plants watered daily were 1.0667 inches taller than those watered weekly.

Crucially, the 95% confidence interval for this difference is provided as . Since this interval does not include zero, we are 95% confident that the true difference in mean height between the daily and weekly watering groups lies between 0.5163 and 1.6170 inches, further solidifying the statistical difference between these two groups.

Interpreting the Post-Hoc Comparisons: Factor 2 (Sunlight)

Similarly, because the `sunlight` factor was highly significant ($p < .0001$), we proceed to review its specific pairwise comparisons using the Tukey method. The sunlight factor has three levels (Low, Medium, High), resulting in three potential pairwise comparisons (High vs. Low, High vs. Medium, and Medium vs. Low).

We look for indicators of statistical significance, often marked by asterisks (***) or based on whether the confidence interval for the mean difference excludes zero. The output for the sunlight factor is displayed below:

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	

By examining the pairwise comparisons in the table, we identify which combinations of sunlight exposure levels yield statistically significant differences in plant height:

High sunlight vs. Low sunlight: This comparison shows a significant difference, with a 95% Confidence Interval of .

High sunlight vs. Medium sunlight: This comparison is also statistically significant, with a 95% Confidence Interval of .

Medium sunlight vs. Low sunlight: Reviewing this comparison, we observe that it is not marked as statistically significant, indicating that plant height does not differ reliably between medium and low sunlight conditions.

In summary, the highest plant growth occurred under high sunlight exposure, and this level was significantly superior to both low and medium exposure conditions.

Reporting the Findings

The final step in any statistical analysis is clearly and concisely reporting the results in a format suitable for scientific documentation. The report should explicitly state the procedure used, the significant findings for the main effects, and the conclusion regarding the interaction effect.

When reporting results from a two-way ANOVA, it is standard practice to first address the interaction term, followed by the main effects, particularly if the interaction is non-significant, allowing for the interpretation of simple main effects.

A formal report of the two-way ANOVA findings should include the following points:

A two-way ANOVA was performed using SAS to analyze the effects of watering frequency and sunlight exposure on the continuous dependent variable, plant growth (height).

The analysis revealed that there was not a statistically significant interaction effect between watering frequency and sunlight exposure on plant height ($p = .1207$).

Simple main effects analysis was thus justified and showed that watering frequency had a statistically significant effect on plant growth ($p = .0005$). Post-hoc comparisons indicated that daily watering resulted in significantly taller plants than weekly watering.

Furthermore, sunlight exposure also had a statistically significant effect on plant growth ($p < .0001$). Post-hoc testing confirmed that high sunlight exposure led to significantly greater growth compared to both medium and low sunlight exposure conditions.

Further Resources

For readers seeking additional depth and context regarding the analysis of variance models and their application within the SAS environment, the following tutorials and documentation provide valuable supplementary information.

These resources cover advanced topics such as assumptions of ANOVA, handling unbalanced data, and generating detailed interaction plots using SAS graphical procedures.