

# How do you perform a Two-Way ANOVA in Python?

Authored by  
**stats writer**

April 17, 2024

## RECOMMENDED CITATION

stats writer (2024). *How do you perform a Two-Way ANOVA in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136191>

A Two-Way ANOVA (Analysis of Variance) is a statistical test used to determine the effect of two categorical variables on a continuous outcome variable. In Python, this test can be performed using the "statsmodels" library. This involves importing the necessary modules, loading the data into a dataframe, and using the "ols" function to fit a linear model. The "anova\_lm" function can then be used to calculate the ANOVA table and obtain the p-values for the variables. Additionally, the "TukeyHSD" function can be used for post-hoc analysis to determine which groups have significant differences. Overall, performing a Two-Way ANOVA in Python involves a few simple steps and can provide valuable insights into the relationship between categorical variables and a continuous outcome variable.

## Perform a Two-Way ANOVA in Python

**A is used to determine whether or not there is a statistically significant difference between the means of three or more independent groups that have been split on two factors.**

**The purpose of a two-way ANOVA is to determine how two factors impact a response variable, and to determine whether or not there is an interaction between the two factors on the response variable.**

**This tutorial explains how to conduct a two-way ANOVA in Python.**

### Example: Two-Way ANOVA in Python

**A botanist wants to know whether or not plant growth is influenced by sunlight exposure and watering**

frequency. She plants 30 seeds and lets them grow for two months under different conditions for sunlight exposure and watering frequency. After two months, she records the height of each plant, in inches.

Use the following steps to perform a two-way ANOVA to determine if watering frequency and sunlight exposure have a significant effect on plant growth, and to determine if there is any interaction effect between watering frequency and sunlight exposure.

**Step 1: Enter the data.**

First, we'll create a pandas DataFrame that contains the following three variables:

**water:** how frequently each plant was watered: daily or weekly  
**sun:** how much sunlight exposure each plant received: low, medium, or high  
**height:** the height of each plant (in inches) after two months

```
import numpy as np
```

```
import pandas as pd
```

```
#create data
```

```
df = pd.DataFrame({'water': np.repeat(, 15),
```

```
'sun': np.tile(np.repeat(, 5), 2),  
'height': })
```

```
#view first ten rows of data  
df
```

```
water sun height
```

```
0 daily low 6
```

```
1 daily low 6
```

```
2 daily low 6
```

```
3 daily low 5
```

```
4 daily low 6
```

```
5 daily med 5
```

```
6 daily med 5
```

```
7 daily med 6
```

```
8 daily med 4
```

```
9 daily med 5
```

**Step 2: Perform the two-way ANOVA.**

Next, we'll perform the two-way ANOVA using the `from` from the `statsmodels` library:

```
import statsmodels.api as sm
```

```
from statsmodels.formula.api import ols
```

### #perform two-way ANOVA

```
model = ols('height ~ C(water) + C(sun) +  
C(water):C(sun)', data=df).fit()  
sm.stats.anova_lm(model, typ=2)
```

```
sum_sq df F PR(>F)
```

```
C(water) 8.533333 1.0 16.0000 0.000527
```

```
C(sun) 24.866667 2.0 23.3125 0.000002
```

```
C(water):C(sun) 2.466667 2.0 2.3125 0.120667
```

```
Residual 12.800000 24.0 NaN NaN
```

### Step 3: Interpret the results.

We can see the following p-values for each of the factors in the table:

```
water: p-value = .000527 sun: p-value =  
.0000002 water*sun: p-value = .120667
```

Since the p-values for water and sun are both less than .05, this means that both factors have a statistically significant effect on plant height.

**Note:** Although the ANOVA results tell us that watering frequency and sunlight exposure have a statistically

**significant effect on plant height, we would need to perform to determine exactly how different levels of water and sunlight affect plant height.**

**The following tutorials explain how to perform other common tasks in Python:**

ARABPSYCHOLOGY.COM