Ranjit
Ku Nayak

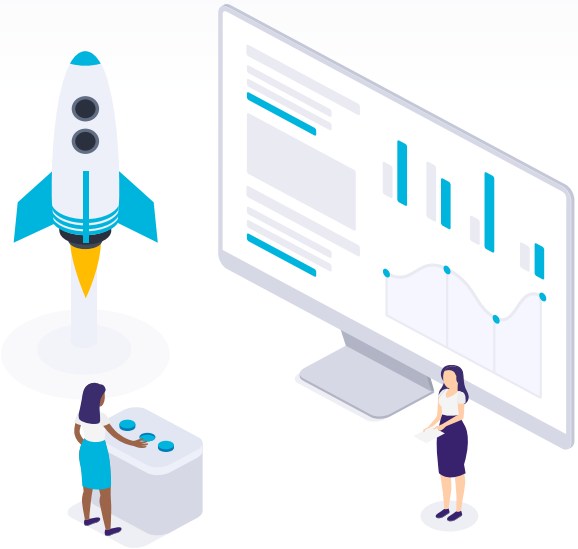
A CASE STUDY ON TWO-WAY ANOVA TEST OF PLANT HEIGHT

Ranjit Ku Nayak

WHAT IS ANOVA

❑Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

❑The t- and z-test methods developed in the 20th century were used for statistical analysis until 1918, when Ronald Fisher created the analysis of variance method. ANOVA is also called the Fisher analysis of variance, and it is the extension of the t- and z-tests.



Ranjit
Ku Nayak

TWO-WAY ANOVA

There are two main types of ANOVA: one-way (or unidirectional) and two-way. There also variations of ANOVA. A two-way ANOVA is an extension of the one-way ANOVA. With a one-way, we have one independent variable affecting a dependent variable. With a two-way ANOVA, there are two independents. For example, a two-way ANOVA allows a company to compare worker productivity based on two independent variables, such as salary and skill set. It is utilized to observe the interaction between the two factors and tests the effect of two factors at the same time.

WHEN TO USE A TWO-WAY ANOVA

Ranjit Ku Nayak

- We can use a two-way ANOVA when we have collected data on a quantitative dependent variable at multiple levels of two categorical independent variables.
- A quantitative variable represents amounts or counts of things. It can be divided to find a group mean.
- A categorical variable represents types or categories of things. A level is an individual category within the categorical variable.

Ranjit
Ku Nayak

HOW DOES THE ANOVA TEST WORK?

- ANOVA tests for significance using the F-test for statistical significance. The F-test is a GroupWise comparison test, which means it compares the variance in each group mean to the overall variance in the dependent variable.
- If the variance within groups is smaller than the variance between groups, the F-test will find a higher F-value, and therefore a higher likelihood that the difference observed is real and not due to chance.
- A two-way ANOVA with interaction tests three null hypotheses at the same time:
- There is no difference in group means at any level of the first independent variable.
- There is no difference in group means at any level of the second independent variable.
- The effect of one independent variable does not depend on the effect of the other independent variable.

Data sheet:

Ranjit Ku Nayak

Water	Sun	Height
Daily	Low	6
Daily	Mid	6
Daily	Low	5
Daily	Mid	6
Weekly	High	7
Daily	Mid	6
Weekly	Low	5
Weekly	Mid	5
Daily	High	6
Weekly	High	7
Weekly	High	6
Daily	Mid	6
Daily	High	8
Weekly	High	8
Daily	High	6

Data sheet is about A botanist wants to know whether or not plant growth is influenced by sunlight exposure and watering frequency. She plants 15 seeds and lets them grow for two months under different conditions for sunlight exposure and watering frequency. After two months, she/she records the height of each plant, in inches.

Data

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.

Ranjit Ku Nayak

Interpretation of Result

jupyter Marketing Analytics Project Last Checkpoint: 3 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

In [2]:
```
import numpy as np
import pandas as pd
import statsmodels.api as statm
from statsmodels.formula.api import ols
```

In [7]:
```
dt=pd.read_excel('E:/CUTM/4th sem/Marketing Analysis/Tree Height.xlsx')
```

In [8]:
```
dt
```

Out[8]:

	Water	Sun	Height
0	Daily	Low	6
1	Daily	Mid	6
2	Daily	Low	5
3	Daily	Mid	6
4	Weekly	High	7
5	Daily	Mid	6
6	Weekly	Low	5
7	Weekly	Mid	5
8	Daily	High	6
9	Weekly	High	7
10	Weekly	High	6
11	Daily	Mid	6
12	Daily	High	8
13	Weekly	High	8
14	Daily	High	6

jupyter Marketing Analytics Project Last Checkpoint: 3 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Trusted

Run Code

In [11]:
```
ano=ols('Height~Water',data=dt).fit()
```

In [12]:
```
first=statm.stats.anova_lm(ano,type=2)
```

In [13]:
```
first
```

Out[13]:

	df	sum_sq	mean_sq	F	PR(>F)
Water	1.0	0.177778	0.177778	0.189091	0.670806
Residual	13.0	12.222222	0.940171	NaN	NaN

In [15]:
```
ano2=ols('Height~Sun',data=dt).fit()
```

In [16]:
```
Second=statm.stats.anova_lm(ano2,type=2)
```

In [17]:
```
Second
```

Out[17]:

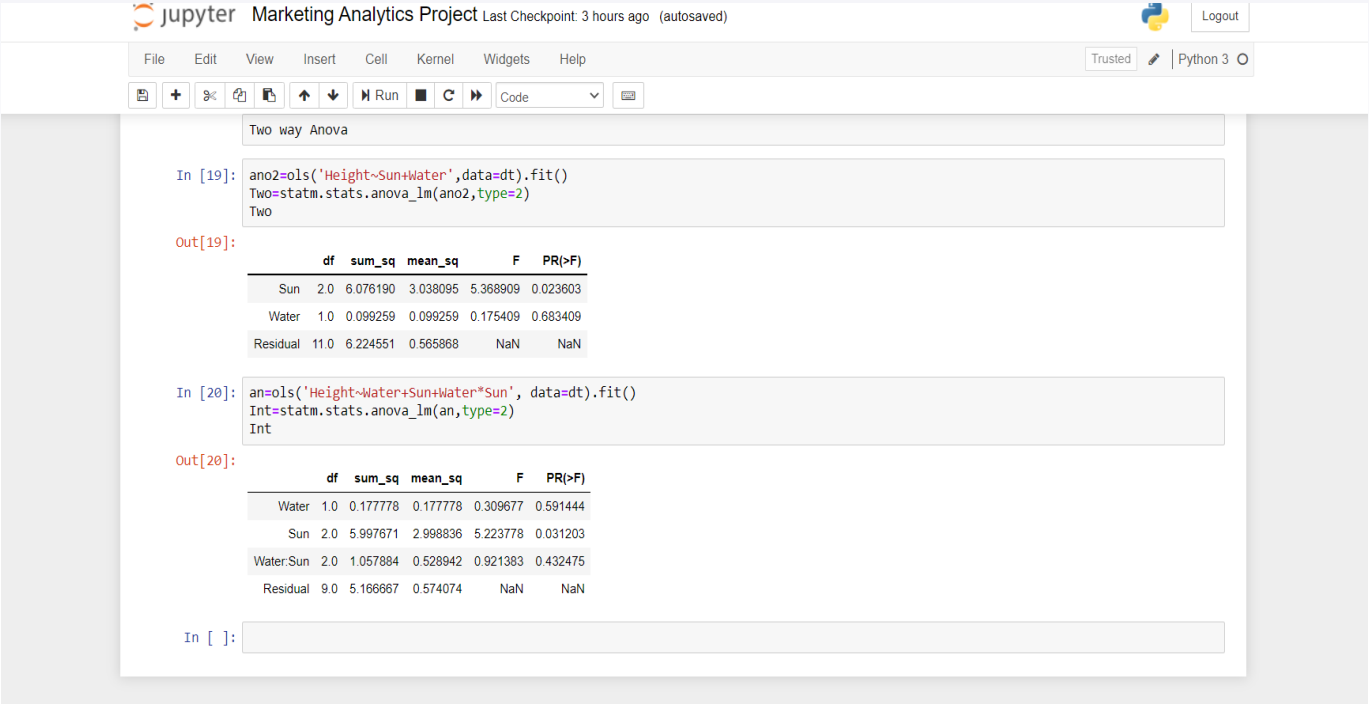
	df	sum_sq	mean_sq	F	PR(>F)
Sun	2.0	6.07619	3.038095	5.76506	0.017593
Residual	12.0	6.32381	0.526984	NaN	NaN

```
Two way Anova
```

In [19]:
```
ano2=ols('Height~Sun+Water',data=dt).fit()
Two=statm.stats.anova_lm(ano2,type=2)
Two
```

Out[19]:

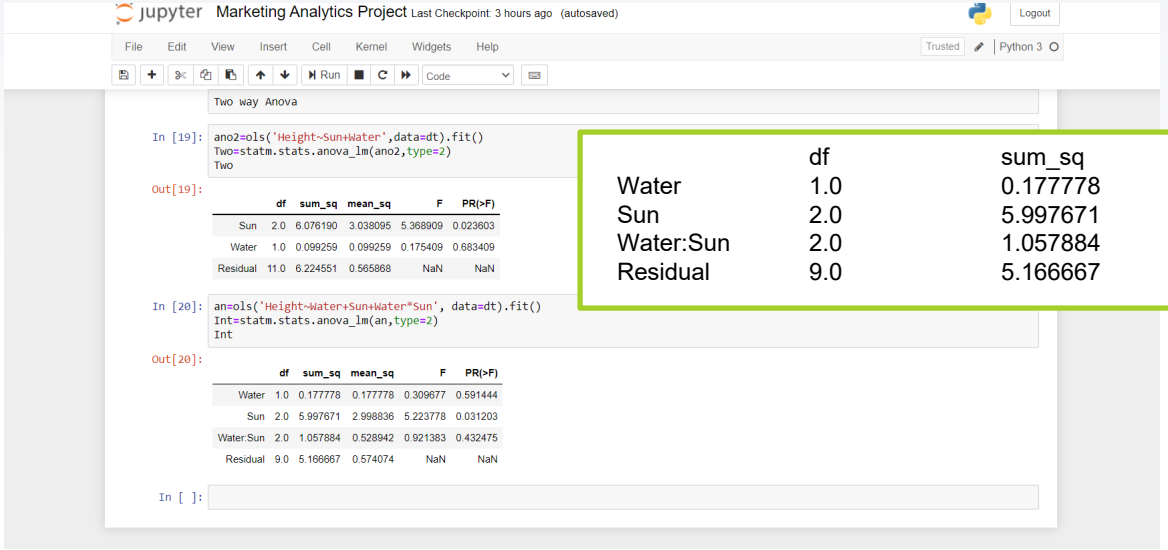
Interpretation of Result



DF: Degree of Freedom
Sum sq: Sum of Squares
Mean sq: Mean sum of Squares
F value: Test Statistic
Pr(>F): P-value

Ranjit
Ku Nayak

Interpretation of Result



	df	sum_sq	mean_sq	F	PR(>F)
Water	1.0	0.177778	0.177778	0.309677	0.591444
Sun	2.0	5.997671	2.998836	5.223778	0.031203
Water:Sun	2.0	1.057884	0.528942	0.921383	0.432475
Residual	9.0	5.166667	0.574074	NaN	NaN

We can see the following p-values for each of the factors in the table.
Water: p-value=0.591
Sun: p-value=0.0312
Water and Sun=0.4324

Ranjit
Ku Nayak

Conclusion:

- Since the p-value for sun less than .05. this means that sun have statistically significant effect on plant height.
- Water p-value is more than .05 this is tells us that there is no significant effect on water.
- And since p-value for the interaction effect is 0.4324 which is not less than .05, this tells us there is no significant interaction effect between sunlight exposure and watering frequency.
- ANOVA result tell us that sunlight effect on plant height.

Ranjit
Ku Nayak