How can a Three-Way ANOVA be performed in Python?

Authored by

stats writer

June 27, 2024

A Three-Way ANOVA (Analysis of Variance) is a statistical technique used to examine the differences between three or more groups. It is commonly used to analyze the effects of multiple independent variables on a dependent variable. In Python, a Three-Way ANOVA can be performed using the "statsmodels" library, which provides a function called "ols" (ordinary least squares) to fit linear regression models. This function can be used to specify the dependent and independent variables, as well as any interaction terms between the variables. The "anova_lm" function within this library can then be used to perform the Three-Way ANOVA and obtain the relevant statistical results. Overall, with the use of the "statsmodels" library, a Three-Way ANOVA can be easily conducted in Python to analyze the relationship between multiple variables and their impact on a given outcome.

Perform a Three-Way ANOVA in Python

A three-way ANOVA 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 three factors.

The following example shows how to perform a three-way ANOVA in Python.

Example: Three-Way ANOVA in Python

Suppose a researcher wants to determine if two training programs lead to different mean improvements in jumping height among college basketball players.

The researcher suspects that gender and division (Division I or II) may also affect jumping height so he collects data for these factors as well.

His goal is to perform a three-way ANOVA to determine how training program, gender, and division affect jumping height.

Use the following steps to perform this three-way ANOVA in Python:

Step 1: Create the Data

First, let's create a pandas DataFrame to hold the data:

```
import numpy as np
import pandas as pd

#create DataFrame
df = pd.DataFrame({'program': np.repeat(, 20),
'gender': np.tile(np.repeat(, 10), 2),
'division': np.tile(np.repeat(, 5), 4),
'height': })

#view first ten rows of DataFrame
df
```

program gender division height
0 1 M 1 7
1 1 M 1 7

2 1 M 1 8

3 1 M 1 8

4 1 M 1 7

5 1 M 2 6

6 1 M 2 6

7 1 M 2 5

8 1 M 2 6

9 1 M 2 5

Step 2: Perform the Three-Way ANOVA

Next, we can use the anova_lm() function from the statsmodels library to perform the three-way ANOVA:

```
import statsmodels.apias sm
from statsmodels.formula.apiimport ols

#perform three-way ANOVA
model = ols("""height ~ C(program) + C(gender) + C(division) +
C(program):C(gender) + C(program):C(division) + C(gender):C(division) +
C(program):C(gender):C(division)""", data=df).fit()

sm.stats.anova_lm(model, typ=2)
```

sum_sq df F PR(>F)
C(program) 3.610000e+01 1.0 6.563636e+01 2.983934e-09
C(gender) 6.760000e+01 1.0 1.229091e+02 1.714432e-12
C(division) 1.960000e+01 1.0 3.563636e+01 1.185218e-06
C(program):C(gender) 2.621672e-30 1.0 4.766677e-30 1.000000e+00
C(program):C(division) 4.000000e-01 1.0 7.272727e-01 4.001069e-01
C(gender):C(division) 1.000000e-01 1.0 1.818182e-01 6.726702e-01
C(program):C(gender):C(division) 1.000000e-01 1.0 1.818182e-01 6.726702e-01
Residual 1.760000e+01 32.0 NaN NaN

Step 3: Interpret the Results

The Pr(>F) column shows the p-value for each individual factor and the interactions between the factors.

From the output we can see that none of the interactions between the three factors were statistically significant.

We can also see that each of the three factors (program, gender, and division) were statistically significant with the following p-values:

P-value of program: 0.00000000298P-value of gender: 0.00000000000171P-value of division: 0.00000185

In conclusion, we would state that training program, gender, and division are all significant predictors of the jumping height increase among players.

We would also state that there are no significant interaction effects between these three factors.

The following tutorials explain how to fit other ANOVA models in Python: