

How can you perform an ANCOVA in Python?

Authored by
stats writer

April 17, 2024

RECOMMENDED CITATION

stats writer (2024). *How can you perform an ANCOVA in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136203>

ANCOVA (Analysis of Covariance) is a statistical technique used to compare the means of two or more groups while controlling for the effect of one or more continuous variables. This technique can be performed in Python by using the statsmodels library. The first step is to import the necessary modules and load the data into a pandas dataframe. Next, the model is set up by specifying the dependent variable, independent variables, and the covariates. The model is then fitted using the OLS (ordinary least squares) function. The results can be interpreted by examining the ANCOVA table, which displays the effect of the covariate on the dependent variable, as well as the adjusted means for each group. Additionally, post-hoc tests can be performed to determine any significant differences between the groups. Overall, performing an ANCOVA in Python allows for a comprehensive analysis of the relationship between categorical and continuous variables, while controlling for the effect of covariates.

Perform an ANCOVA in Python

An ANCOVA ("analysis of covariance") is used to determine whether or not there is a statistically significant difference between the means of three or more independent groups, after controlling for one or more covariates.

This tutorial explains how to perform an ANCOVA in Python.

Example: ANCOVA in Python

A teacher wants to know if three different studying techniques have an impact on exam scores, but she wants to account for the current grade that the student already has in the class.

She will perform an ANCOVA using the following variables:

**Factor variable: studying technique
Covariate: current grade
Response variable: exam score**

Use the following steps to perform an ANCOVA on this dataset:

Step 1: Enter the data.

First, we'll create a pandas DataFrame to hold our data:

```
import numpy as np
import pandas as pd

#create data
df = pd.DataFrame({'technique': np.repeat(, 5),
'current_grade': ,
'exam_score': })
#view data
df

technique current_grade exam_score
0 A 67 77
1 A 88 89
```

2 A 75 72

3 A 77 74

4 A 85 69

5 B 92 78

6 B 69 88

7 B 77 93

8 B 74 94

9 B 88 90

10 C 96 85

11 C 91 81

12 C 88 83

13 C 82 88

14 C 80 79

Step 2: Perform the ANCOVA.

Next, we'll perform an ANCOVA using the [ancova\(\)](#) function from the pingouin library:

```
pip install pingouin
```

```
from pingouin import ancova
```

```
#perform ANCOVA
```

```
ancova(data=df, dv='exam_score',  
covar='current_grade', between='technique')
```

```
Source SS DF F p-unc np2
0 technique 390.575130 2 4.80997 0.03155 0.46653
1 current_grade 4.193886 1 0.10329 0.75393 0.00930
2 Residual 446.606114 11 NaN NaN NaN
```

Step 3: Interpret the results.

From the ANCOVA table we see that the p-value (p-unc = "uncorrected p-value") for study technique is 0.03155. Since this value is less than 0.05, we can reject the null hypothesis that each of the studying techniques leads to the same average exam score, *even after accounting for the student's current grade in the class.*