

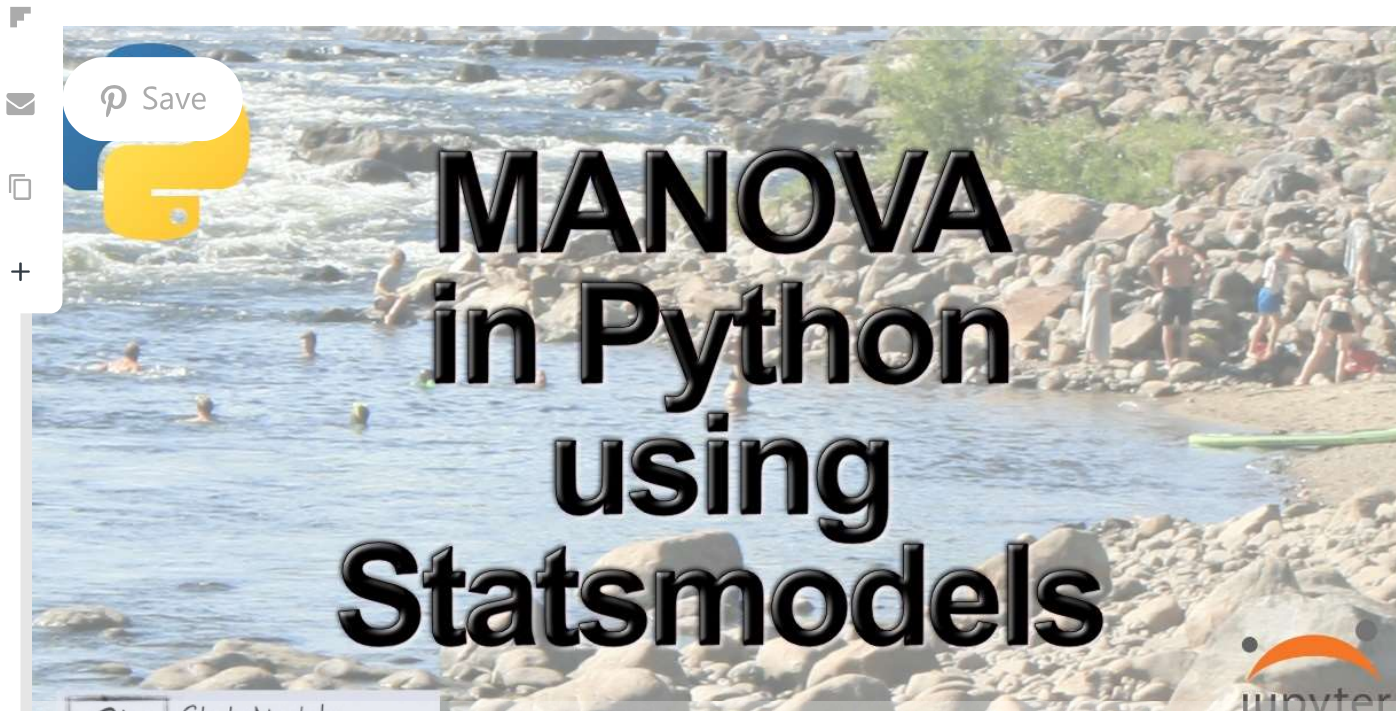


me » Programming » Python » Python MANOVA Made Easy using Statsmodels

Python MANOVA Made Easy using Statsmodels

Erik Marsja  August 26, 2019  Leave a Comment



Last updated on November 12, 2019

Share via:

3
Shares



In previous posts, we learned how to use Python to detect group differences on a single dependent variable. However, there may be situations in which we are interested in several dependent variables. In these situations, the simple ANOVA model is inadequate.

One way to examine multiple dependent variables using Python would, of course, be to carry out multiple ANOVA. That is, one ANOVA for each of these dependent variables. However, the more tests we conduct on the same data, the more we inflate the family-wise error rate (the greater chance of making a Type I error).

This is where MANOVA comes in handy. MANOVA, or Multivariate Analysis of Variance, is an extension of Analysis of Variance (ANOVA). However, when using MANOVA we have two, or more, dependent variables.

MANOVA and ANOVA is similar when it comes to some of the assumptions. That is, the data have to be:

- normally distributed dependent variables
- equal covariance matrices)

In this post will learn how carry out MANOVA using Python (i.e., we will use Pandas and Statsmodels). Here, we are going to use the Iris dataset which can be downloaded [here](#).

Table of Contents

- What is MANOVA?
 - MANOVA Example
 - Assumptions of MANOVA
- How to Carry out MANOVA in Python
 - Python MANOVA Example
- Conclusion
- Resources
 - Analysis of Variance using Python:
 - Repeated Measures Analysis of Variance using Python:

What is MANOVA?

First, we going to have brief introduction to what MANOVA is. MANOVA is the acronym for Multivariate Analysis of Variance. When analyzing data, we may encounter situations where we have there multiple response variables (dependent variables). As mentioned before, by using MANOVA we can test them simultaneously.

MANOVA Example

Before getting into how to do a MANOVA in Python, let's look at an example where MANOVA can be a useful statistical method. Assume we have a hypothesis that a new therapy is better than another, more common, therapy (or therapies, for that matter). In this case, we may want to look at the effect of therapies (independent variable) on the mean values of several dependent variables.

For instance, we may be interested in whether the therapies help for a specific psychological disorder (e.g., depression), at the same time as we want to know how it changes life satisfaction, lower suicide risk, among other things. In such an experiment a MANOVA lets us test our hypothesis for all three dependent variables at once.

Assumptions of MANOVA

In this section, we will briefly discuss some of the assumptions of carrying out MANOVA. There are certain conditions that need to be considered.

- The dependent variables should be normally distributed within groups. That is, in the example below the dependent variables should be normally distributed within the different treatment groups.
- Homogeneity of variances across the range of predictors.
- Linearity between all pairs of dependent variables (e.g., between depression, life satisfaction, and suicide risk), all pairs of covariates, and all dependent variable-covariate pairs in each cell

How to Carry out MANOVA in Python

In this section, we will focus on how to conduct the Python MANOVA using Statsmodels. First, the first code example, below, we are going to import Pandas as pd. Second, we import the MANOVA class from statsmodels.multivariate.manova.

```
import pandas as pd
from statsmodels.multivariate.manova import MANOVA
```

Before carrying out the Python MANOVA we need some example data. This is why we use Pandas. In the next code chunk, we are going to read a CSV file from a URL using Pandas read_csv. We are also going to replace the dots (".") in the column names with underscores ("_"). If you need to find out more about cleaning your data see post [data cleaning in Python with Pandas](#).

```
url = 'https://vincentarelbundock.github.io/Rdatasets/csv/datasets/iris.csv'
df = pd.read_csv(url, index_col=0)
df.columns = df.columns.str.replace(".", "_")
df.head()
```

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa

Dataframe

Learn more about working with Pandas dataframe:

- [Pandas Read Excel Tutorial](#)
- [Read CSV files using Pandas](#)
- [How to Group Data using Pandas Groupby.](#)
- [Pandas Dataframe Tutorial](#)

Python MANOVA Example

Now that we have read a data file (i.e., a CSV file) using Pandas we are ready to carry out the MANOVA in Python. In the Python MANOVA example below we are going to use the `from_formula` method. This method takes the formula as a string object.


In this MANOVA example, we are going to use the width and length columns as dependent variables. Furthermore, the species column is used as independent variable. That is, we are carrying out a one-way MANOVA here:

```
maov = MANOVA.from_formula('Sepal_Length + Sepal_Width + \
                             Petal_Length + Petal_Width ~ Species', data=df)
```

Finally, we have used Python to do a one-way MANOVA. The last thing to do is to print the MANOVA table using the `mv_test` method:

```
print(maov.mv_test())
```

Multivariate linear model

 Save

Intercept	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.0170	4.0000	144.0000	2086.7720	0.0000
Pillai's trace	0.9830	4.0000	144.0000	2086.7720	0.0000
Hotelling-Lawley trace	57.9659	4.0000	144.0000	2086.7720	0.0000
Roy's greatest root	57.9659	4.0000	144.0000	2086.7720	0.0000

Species	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.0234	8.0000	288.0000	199.1453	0.0000
Pillai's trace	1.1919	8.0000	290.0000	53.4665	0.0000
Hotelling-Lawley trace	32.4773	8.0000	203.4024	582.1970	0.0000
Roy's greatest root	32.1919	4.0000	145.0000	1166.9574	0.0000

Python MANOVA table

[Here's a link](#) to a Jupyter Notebook containing the MANOVA Statsmodels example in this post.

Conclusion

In this post, we learned how to carry out a Multivariate Analysis of Variance (MANOVA) using Python and Statsmodels. More specifically we have:

- used Pandas to load a dataset from a CSV file.
- cleaned column names of a Pandas dataframe
- learned multivariate analysis by a MANOVA statsmodels example

Resources

Analysis of Variance using Python:

- [One-Way ANOVA in Python](#)
- [Two-Way ANOVA in Python](#)

Repeated Measures Analysis of Variance using Python:

- [One-Way Repeated Measures ANOVA in R and Python](#)
- [Two-Way Repeated Measures ANOVA using Python](#)

Share via:

3
Shares**Categories:** *Programming Python*

ANOVA

Data Analysis

MANOVA

Pandas

Statsmodel

Previous Post

The Easiest Data Cleaning Method using Python & Pandas

Next Post

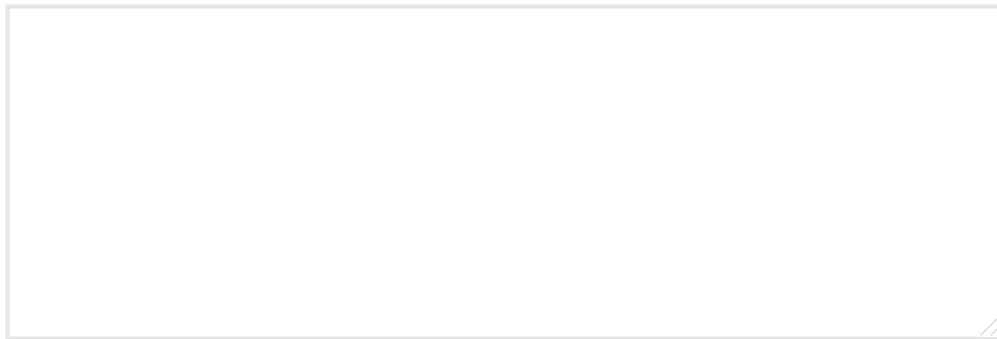
How to Read & Write SPSS Files in Python using Pandas

Be First to Comment

Leave a Reply

Your email address will not be published. Required fields are marked *

Comment



Name*



Jane Doe



Email*



name@email.com



Website



http://google.com



Post Comment

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)

Shift WordPress Theme by Compete Themes.

