Understanding Statsmodels in Python

As its name implies, Statsmodels is designed for advanced statistics and makes it possible to work with statistics in a way that no other software provides. Statsmodels is an excellent statistical analysis tool that is better matched with R, making it easier to use for R users who wish to transition to Python.

In this tutorial, we learn how to use the Statsmodels module. Let’s get started with an introduction to the model and later on proceed to the implementation of the module.

## Introduction to Statsmodels

Statsmodels is a popular Python module that provides estimation and analysis of numerous statistical models. It is composed of mathematical and scientific libraries such as NumPy and SciPy.

The module contains different linear regression models, such as ordinary least squares, generalized least squares, weighted least squares, and others. It also provides efficient time series analysis functions.

Need some datasets? Well, have no fear, as the module includes test and example datasets. Even models based on survival analysis are included in this module. All the statistical tests imaginable for large-scale data are available.

Let's now explain the steps required to install Statsmodels on our PC. To accomplish this, we can open the Command Prompt and enter the command below.

pip install statsmodels

Now that the installation is complete, proceed to the following phase to begin implementing the same in other programs.

## Implementation of Statsmodels in Python

Let's begin by examining a basic example to better comprehend the package. The program begins by importing the appropriate modules, followed by the loading of the dataset. The Guerry dataset is a compilation of historical data utilized to support Andre-Michel Guerry's 1833 Essay on the Moral Statistics of France. The subsequent step is to fit the model using the module and obtain the model's summary after fitting.

Now that we are acquainted with the module, we will move on to something more complex. StatsModels will be utilized to implement Linear Regression on our dataset. Examine the code fragment shown below.

## Conclusion

This lesson has demonstrated that StatsModels simplifies statistical analysis. We have examined examples of developing statistical models using Linear Regression.

The Python StatsModels module makes it simple to develop models with minimal effort and minimal lines of code. Additionally, the output is presented in a form that is easy to read and comprehend.

Have a look at the following tutorials:

1. Python statistics module – 7 functions to know!
2. How to Calculate Summary Statistics in Python?
3. Data Analysis in Python with Pandas

Happy Learning!