

A Non-Programmer Guide for Python for Data Analysis

Ling Liu

2018 年 11 月 29 日

目录

Chapter 1

Data Analytica Models

In this lecture we will discuss how to use Python to build data analytic models.

1.1 Statistical Models

Statsmodels is probably the most common Python package for statistical models. Statsmodels(<http://www.statsmodels.org/stable/>) is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. An extensive list of result statistics are available for each estimator. The results are tested against existing statistical packages to ensure that they are correct.

1.1.1 Simple OLS

在调用相关的 package 之前，我们需要先理解模型的意思，比如下面这个多元回归方程：

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Reference