

**NATIONAL UNIVERSITY OF SINGAPORE**  
**NUS Business School**  
**Department of Analytics & Operations**

**DAO2702 Programming for Business Analytics**

**Lecturers:** Associate Professor WANG Tong (Co-lecturer)  
Dr. TUNG, Yi-liang (Co-lecturer)

**Session** : Semester 1, 2018/2019

**Description**

Analytics involves both a theoretical foundation of Statistics and practical capabilities of implementation via programming. The two are typically covered in different courses in an isolated and uncoordinated way. This module, as an introductory course to Data Analytics, aims to bring the theory and practice together and offer a holistic and organically connected view of both sides.

The module starts with basic Python programming. It then walks through Statistics topics from visualizing and summarizing data, to estimating model parameters and hypothesis testing, and to linear regression analysis. For each topic, Python illustrations and experimentations are interwoven inside so as to help students better appreciate statistical theory and also understand how it works in practice. The module finishes with practical issues on acquiring, cleaning, and organizing data using Python. This completes the cycle of data analysis and the students are able to independently execute a basic Data Analytics project.

**Objectives**

Students are expected to grasp good understanding of fundamental Statistics and become proficient in the extensive use of Python in Data Analytics. Moreover, many useful Python packages are also discussed to equip students with technical skills to conduct real business analytics projects in the future.

**Course Outline**

**1) Basic Python Programming**

- a) Data Structures
- b) Flow Control

- c) Functions and Packages
- d) Data Import and Output

## **2) Understanding Data and Visualization**

- a) Summary Statistics and Empirical Cumulative Distribution Function
- b) Data Visualization: Histogram, Scatterplot, Boxplot, Line plot
- c) Python Implementation: Matplotlib and NumPy packages

## **3) Statistical Concepts and Inference Techniques**

- a) Sampling and Population
- b) Parameter Estimation
- c) Confidence Intervals
- d) Hypothesis Testing

## **4) Linear Regression Analysis**

- a) Model Assumptions and Interpretations
- b) Categorical Variables and Interaction Effects
- c) Model Selection

## **5) Advanced Python**

- a) Pandas Package
- b) Obtaining data from the Internet
- c) Data Cleaning

As well, several ethical issues will be discussed throughout the whole semester. The specific topics are as follows:

- a) Ethics for Data Visualization
- b) Ethics for Data Collection and Analysis
- c) Ethics for Making Generalization based on Sample Data

## **Syllabus**

*Week 0.*

**Setup of Python environment** (Spyder and Jupyter notebook). To be done offline.

*Week 1.*

**Python Basics I** — Data Structures: Variable, List, Dictionary.

**Python Basics II** — Flow Control: If-else, For/while loop, Iterators, List comprehension.

*Week 2.*

**Python Basics III** — Functions and Packages. lambda functions.

**Python Basics IV** — File Operations: reading and writing txt files.

*Week 3.*

**Statistics I:** Visualizing data. Histogram, Scatterplot, Boxplot, line plot of time series.

**Python Package I** — Visualization with matplotlib. Basic visualization and plotting tools.

*Week 4.*

**Statistics II:** Summarizing data with summary statistics. Mean/Median/Mode, Min/Max, Stdev/Variance, Quantile.

**Python Package II:** Scientific Calculations with NumPy: NumPy multi-dimensional array, standard scientific functions such as logarithm.

*Week 5.*

**Statistics III:** Parameter estimation and confidence intervals.

**Python Practice:** Simulating data I: random number generator, seed;

*Week 6.*

**Statistics IV:** Hypothesis Testing and Power Function of a Test

**Python Practice:** Simulating data II

*Week 7.*

**Linear Regression I:** predictive modelling, interpretation of regression outputs (coefficient, confidence interval, p-value,  $R^2$ )

**Python Practice:** Linear regression with statstools I.

*Week 8.*

**Linear Regression II:** categorical variables, modeling nonlinearity (transformation, interaction).

**Python Practice:** Linear regression with statstools II.

*Week 9.*

**Linear Regression III:** model selection, missing data, outliers.

**Python Practice:** Linear regression with statstools III.

*Week 10.*

**Advanced Python I:** organizing data with pandas. Data frame structures and data manipulation

tools.

*Week 11.*

**Advanced Python II:** obtaining data from the Internet: crawling data from API (json/XML), crawling data from HTML page.

*Week 12.*

**Advanced Python III:** data cleaning I: tidying data (subset, transformation)

*Week 13.*

**Advanced Python IV:** data cleaning II: merging data.

### **Reading List**

#### **Compulsory reading:**

“Business Analytics: Data Analysis & Decision Making” 6th Edition by Albright and Winston.  
(Only chapters 7,8,9,10 and 11 are discussed)

#### **Supplementary reading:**

“Python Crash Course: A Hands-On, Project-Based Introduction to Programming” 1st Edition  
by Eric Matthes (**Freely Downloadable from Internet**)

“Guide to NumPy” 2nd Edition by Travis E. Oliphant (**Freely Downloadable from Internet**)

“Python Data Science Handbook: Essential Tools for Working with Data” 1st Edition by Jake  
VanderPlas

### **Course Package**

### **Prerequisites**

DAO1704 Decision Analytics using Spreadsheets

### **Weightage of Assessment**

#### **Continuous Assessment:**

Class Participation	10%
Group Project	20%
Midterm Quiz	10%

Assignments	20%
<b>Final Examination</b>	<b>40%</b>