DA503 Applied Statistics

Lecture 00
Course Essentials

- **DA503 Applied Statistics / Fall 2020**

- Course duration: Nov 20th to Jan 3rd

- Instructor
  - H. Sait Ölmez
  - Email: olmez@sabanciuniv.edu
  - Phone: (216) 483 9578

- Teaching Assistant: Yalçın Can Kılıç

- Course materials on SUCourse+
  - Slides, code + data, HW assignments, solution sets and other reading materials

- **Recommended textbooks**
  - Python for Data Analysis by Wes McKinney
  - An Introduction to Statistical Learning
    - G. James, D. Witten, T. Hastie and R. Tibshirani
    - www-bcf.usc.edu/~gareth/ISL/
  - Introduction to Statistics with Python: With Applications in the Life Sciences
    - Thomas Haslwanter
  - See the complete list in the syllabus
- **Prerequisites**
  - Linear Algebra and Calculus
  - Basic Probability
  - Programming in Python (all of DA501)
  - Curiosity and will to exercise on your own

- What do I expect from you?

    1. Academic honesty

    2. Will to learn

    3. Plus (reasonable) success in:

        - 3-to-4 x Homework assignments (%25)

        - 1 x MT (35%)

        - 1 x final or take-home exam (40%)

- What will you gain in return?

    – More coding experience in Python

    – Data manipulation and analysis skills

    – Ability to infer from data

Python installation and setup

- ## Which version of Python?

  - We'll use Anaconda distribution (3.7)

- ## Why Anaconda?

  - OS support (Windows, MacOS, Linux)

  - One package that has it all (no additional installations)

    - Python shell

    - IPython shell, **IPython Notebook**

    - Spyder IDE

    - Many modules/libraries that come with the distribution

- ## Download from: http://continuum.io/downloads

- Libraries we'll use frequently (they all come pre-installed with Anaconda)

- **Statistics**:
  - Numpy and Pandas (DA501)
  - Scipy
  - Statsmodels
  - Scikit-learn

- **Visualization**:
  - Matplotlib
  - Seaborn
  - Plotly