



Sungkyunkwan University (SKKU) International Summer Semester (ISS) 2023

## Statistics in Python

Prof. Kyung Eun Park, Graceland University

### SHORT COURSE DESCRIPTION

---

This course is mainly focusing on statistical thinking concepts that are essential for learning from data and communicating insights. This course will introduce Python basics for Data Science and the primary concepts in probability distributions with Python. Students can learn skills to perform exploratory analysis, principles of sampling, regression analysis, statistical modeling, and inference. As a general-purpose language, Python is used to help students learn integrating analytics functionalities with well-structured analytics system. Students will experience beginner level machine learning algorithms based on the topics covered in the class.

### READING MATERIALS

---

**Textbook 1:** Intro Stats, 5<sup>th</sup> Edition, Richard De Veaux, Paul Velleman, and David Boeck, 2018 Pearson, ISBN 13: 978-0-13-421022-3

**Textbook 2:** Python for probability, statistics, and machine learning, 3<sup>rd</sup> Edition, José Unpingco, Springer, ISBN: 978-3-030-18545-9

#### Reference:

- Think Stats, 2<sup>nd</sup> Edition, <https://greenteapress.com/thinkstats2/html/index.html>
- Python Programming, An Introduction to Computer Science, 3<sup>rd</sup> Ed., John Zelle, Franklin, Beedle, ISBN-13: 978-1590282755
- Introduction to Computation and Programming Using Python, 2<sup>nd</sup> Ed., The MIT Press, ISBN-13: 978-0262529624

Additional reading materials will be provided.

Class materials will be provided.

Download and install Python in your laptop:

<https://www.anaconda.com/distribution/>

### COURSE REQUIREMENTS AND GRADING

---

**Attendance (20%):** It is important that you attend every lecture. SKKU regulations require students to attend at least 80% of all classes. Bring your laptop for in-class exercises. Please do not use your cellphone in class.

**Assignments (30%):** Homework will be assigned regularly. No late submission will be accepted without a valid excuse in advance. You are welcome to consult each other on assignments, but you should submit your own work. In order to get full credit on homework and exam problems, you must show everything necessary to establish the validity of your answer and the approach that you use to get it. If you use Python to answer, you must include the relevant code and outputs.

**Exams (50%):** There will be two exams and two quizzes. No make-up exams will be given unless you have a proper reason not to present in the exam (should notify the instructor in advance). You will earn a pass if you receive a grade of 60% or above.

Academic dishonesty will not be tolerated. A fail will be given if academic honesty is violated.

## **COURSE SCHEDULE**

---

### **– WEEK I –**

Monday (26 June) – Day 1

Introduction to Statistics and Python Programming Environment

Tuesday (27 June) – Day 2

Introduction to Python Programming Language

Wednesday (28 June) – Day 3

Python Sequence Data Handling

Thursday (29 June) – Day 4

Basic Statistics with Python NumPy Library

### **– WEEK II –**

Monday (3 July) – Day 5

Random Variables and Probability Function

Tuesday (4 July) – Day 6

Sampling

Wednesday (5 July) – Day 7

Python Pandas Complex Data Structures: Series and Data Frames

Thursday (6 July) – Day 8

Continuous Random Variables

### **– WEEK III –**

Monday (10 July) – Day 9

Probability Measure

Tuesday (11 July) – Day 10

Inferential Statistics

Wednesday (12 July) – Day 11

Sampling Methods and Distributions

Thursday (13 July) – Day 12

Regression Analysis and Hypothesis Testing

### **– WEEK IV –**

Monday (17 July) – Day 13

Introduction to Machine Learning

Tuesday (18 July) – Day 14

Classification Supervised Machine Learning in Scikit Learn Library

Wednesday (19 July) – Day 15

Final exam