Sungkyunkwan University (SKKU) International Summer Semester (ISS) 2024

Statistics in Python

Prof. Kyung Eun Park, Wilkes 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.

READING MATERIALS

Textbook 1: Intro Stats, 5th Edition, Richard De Veaux, Paul Velleman, and David Bock, 2018 Pearson, ISBN 13: 978-0-13-421022-3

Textbook 2: Python for probability, statistics, and machine learning, 2nd Edition, Springer, ISBN: 978-3-030-18545-9

Reference:
 - Think Stats, 2nd Edition, https://greenteapress.com/thinkstats2/html/index.html
 - Python Programming, An Introduction to Computer Science, 3rd Ed., John Zelle, Franklin, Beedle, ISBN-13: 978-1590282755
 - Introduction to Computation and Programming Using Python, 2nd 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 each will count 25%. 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 (1 July) – Day 1

       Introduction to Scientific Python

Tuesday (2 July) – Day 2
Python data handling

Wednesday (3 July) – Day 3
Exploratory data analysis and visualization

Thursday (4 July) – Day 4
Introduction to probability

– WEEK II –

Monday (8 July) – Day 5
Useful distributions

Tuesday (9 July) – Day 6
Exam I

Wednesday (10 July) – Day 7
Information entropy

Thursday (11 July) – Day 8
Normal and sampling distribution

– WEEK III –

Monday (15 July) – Day 9
Regression analysis I

Tuesday (16 July) – Day 10
Regression analysis II

Wednesday (17 July) – Day 11
Sampling distribution models and confidence intervals

Thursday (18 July) – Day 12
Machine Learning with Python libraries

– WEEK IV–

Monday (22 July) – Day 13
Exam II

Tuesday (23 July) – Day 14
Advanced programming in Python

Wednesday (24 July) – Day 15
Project presentation