The Chinese University of Hong Kong
BSSc Data Science and Policy Studies
DSPS1003: Foundation of Data Science
First Term 2020/2021

Teacher : Dr. HO Chi Pui
Lecture time : Thursday 10:30am - 1:15pm
Lecture venue : Zoom link:

Email : chipuiho@cuhk.edu.hk
Phone : 3943 9835
Office : Room 517B, 5/F, Chen Kou Bun Building, Chung Chi College

1. Course Overview

This introductory course provides an overview on the study of data science, covering the development of data industry and challenges in working with big data. Topics to be covered include the definition of data science, data analytics, data visualization, data science process, development of data industry, data science coding, and the wide applications of data science in social science and policy studies. This course also equips students with the essential quantitative skills and knowledge to prepare them for the advanced data science courses (such as DSPS2102 and DSPS2201). Topics covered include data types, data presentation, data transformations, frequency analysis, descriptive statistics, probability theories, confidence intervals, hypothesis testing, correlation, as well as their applications in social science and management problems.

2. Learning Outcomes

By the end of this course, students should be able to:

a. Explain key terms in data science and the challenges to data analytics;
b. Appreciate the wide applications of data science in social science and policy studies;
c. Recognize the importance of statistics and coding in data analytics;
d. Explain the fundamental probability and statistical concepts;
e. Perform interval estimation and hypothesis testing, and detect correlation in social science data.

3.   Course Materials

Lecture notes will be posted on the course website. The course textbooks are listed below:

- Cao, Longbing (2018). *Data Science Thinking: The next Scientific, Technological and Economic Revolution*. Sydney: Springer.
- Miah, Abdul Quader (2016). *Applied Statistics for Social and Management Sciences*, Singapore: Springer.
- (Optional) Miller, Irwin, and Miller, Marylees (2014). *John E. Freund's Mathematical Statistics with Applications*, 8th Edition. Harlow: Pearson.

4.   Coursework and Assessment

Individual assignments (30%)

There will be six homework assignments that will help students understand course materials. Students will submit their homework through the course website. No late homework assignments will be accepted. Demonstrating a good understanding of lecture materials and the ability to present solutions (with steps) clearly will be rewarded in all forms of assessment.

Group assignment (10%)

There will be one group assignment to give students hands-on experience on applying data science knowledge and skills in social science or policy-related issues. The topic and data will be provided. Two to three students will form a group to conduct data analysis (using Excel or other statistical software) and write a short essay or policy memo on the topic. The details will be announced after the test.

Test and Exam (20%+30%)

The test and exam will be open-notes/textbook and administered in the course website. **The whole test/exam paper with hand-written solutions will need to be scanned and uploaded to the course website while taking the test/exam.** More information will be made available before the test/exam.

The use of online resources (other than course notes and the textbook) or communicating with other students during a test/exam will be considered a gross violation of academic honesty. Such violations seriously undermine the academic integrity, and will be treated severely.

Make-up exams will be given only in the event of a legitimate excused absence. For the request for make-up exam, if it is based on medical reasons, it should be supported by medical documents issued by a doctor. If it is due to other reasons, a written request should be submitted in advance and to be approved by the teacher and endorsed by the DSPS Programme Director.

Assessment scheme

Participation*	10%
Individual assignments	30%
Group assignment	10%
Test	20%
Exam	30%
Total	**100%**

* When joining the zoom lecture, each student should turn on their camera. Please edit your screen name to be your surname, space, and then your given name.

Grade descriptions

A: Outstanding performance on all learning outcomes.

A-: Generally outstanding performance on all (or almost all) learning outcomes.

B: Substantial performance on all learning outcomes, OR high performance on some learning outcomes which compensate for less satisfactory performance on others, resulting in overall substantial performance.

C: Satisfactory performance on the majority of learning outcomes, possibly with a few weaknesses.

D: Barely satisfactory performance on a number of learning outcomes.

F: Unsatisfactory performance on a number of learning outcomes, OR failure to meet specified assessment requirements.

5. <u>Course Schedule</u>

Tentative lecture topics and schedule:

Date	Topics
Week 1	Introduction to Data Science
Week 2	Data Science Foundation
Week 3	Data Presentation
Week 4	Probability Theory
Week 6	Probability Distributions
Week 7	Statistical Inference
Week 8	Testing Hypothesis with Data
Week 9	**Test**
Week 10	Introduction to Coding in Data Science
Week 11	Detecting Correlation in Data
Week 12	Advanced Topics in Applied Statistics
Week 13	Computational Social Science

6. **Honesty in academic work and the University's policies**

  The Chinese University of Hong Kong places very high importance on honesty in academic work submitted by students, and adopts a policy of zero tolerance on cheating in examinations and plagiarism. Any related offence will lead to disciplinary action including termination of studies at the University.

  Attention is drawn to University policy and regulations on honesty in academic work, and to the disciplinary guidelines and procedures applicable to breaches of such policy and regulations. Details may be found at http://www.cuhk.edu.hk/policy/academichonesty/.

  With each assignment, students will be required to submit a signed declaration that they are aware of these policies, regulations, guidelines and procedures. For group projects, all students of the same group should be asked to sign on the declaration.

  For assignments in the form of a computer-generated document that is principally text-based and submitted via VeriGuide (https://www.cuhk.edu.hk/veriguide), the statement, in the form of a receipt, will be issued by the system upon students' uploading of the soft copy of the assignment. Assignments without the receipt will not be graded by teachers. Only the final version of the assignment should be submitted via VeriGuide.