SAMPLE course outline only
Course contents will vary if taught by different teachers.
Please refer to the course teacher for the actual course content offered in each term.

As of January 6, 2024

SOCI 3229: Quantitative Data Analysis

Department of Sociology
The Chinese University of Hong Kong

2023-24 Term 2

Thursdays 16:30 – 18:15
Esther Lee Building 308

Contact Information

Instructor: Professor LEE Jaemin
E-mail: jaeminlee@cuhk.edu.hk
Office hours: By appointment
Office: Sino Building Room 418

Teaching Assistant: Mr. LI Yujie
E-mail: yjli@link.cuhk.edu.hk
Office hours: By appointment
Office: Sino Building Room 421

Course Description

Data analysis is a powerful tool for answering many interesting questions about societies and human behavior. Data-driven approaches are useful for solving the multitude of challenging problems that today's societies confront. The focus of this course is on hands-on data analysis and the practical application of basic statistical methods to real-world social problems. Topics include causality, measurement, probability and statistical theory, linear and multivariate regression models, and basics of machine learning and text analysis.

At the end of this course, students should be able to analyze data, interpret the results, and effectively communicate their empirical findings. The best way to learn about data analysis and new statistical procedures is by doing—not by reading and paper-and-pencil statistics per se. Hence, particular attention throughout the course will be paid to learning and implementing the **R** statistical program.

There is no prerequisite to enroll in this course. Having taken Social Statistics (SOCI 2004 or equivalent elsewhere) or prior computing experiences might be helpful but not required.

Assessment and Grading

The grade for the course will be calculated as a weighted average of the following components:

Participation (Lecture & Tutorial)	10%
Assignments	40%
Final Paper	50%

Participation (10%)

- Attendance will be an important grading component and a crucial factor in your success as the material builds on itself cumulatively throughout the course. I will highly compensate diligent and active students.
- Do not fall behind. If you start falling behind, see your tutor immediately to catch up.
- Class starts promptly on time. Arrivals after 5 minutes of the class beginning will be counted as lateness.
- Equal weight will be given to attendance in tutorials.

Assignments (40%)

- Assignments consist of problem sets, quizzes, and programming exercises on a (roughly) biweekly basis. The goal is to maximize your understanding of the content of the week.
- You are allowed to help each other to understand materials or programming techniques. But do not, under any circumstances, copy another person's code or answers.
- All submissions should be made on Blackboard.
- For all assignments, late submission is not allowed without at least 24 hours prior notice.

Final Paper (50%)

- The final paper is a team project (2 – 4 students in a team).
- 3 components:
    - Proposal (10%): Submission by **April 3**
    - Presentation (10%): Class on **April 18**
    - Final Paper (30%): Submission by **May 12**
- Your team must submit it in the form of social science research—the product of your own quantitative data analysis.
- Specific instructions and assessment criteria will be provided.

Grading

Grade	Descriptors
A	Excellent: Outstanding performance on all learning outcomes.
A-	Very Good: Generally outstanding performance on all (or almost all) learning outcomes.
B	Good: Substantial performance on all learning outcomes, OR high performance on some learning outcomes which compensates for less satisfactory performance on others, resulting in overall substantial performance.
C	Fair: Satisfactory performance on the majority of learning outcomes, possibly with a few weaknesses.
D	Pass: Barely satisfactory performance on a number of learning outcomes.
F	Failure: Unsatisfactory performance on a number of learning outcomes, OR failure to meet specified assessment requirements.

Tutorials

You are required to attend tutorials. Tutorials will offer solutions to problem sets given in assignments, demonstrate R programming tutorials, and catch up/further the content of the lectures.

Textbook

It is your responsibility to do the reading *before* class. Required readings will be based on the combination of the following books (in the order of importance):

Imai, Kosuke. 2018. *Quantitative Social Science: An Introduction*. Princeton University Press.

- E-book: Purchasable at Amazon about HK$290
- Loan: Course Reserve 4 hours at United College Wu Chung Library
- Paperback copy: I've requested adoption at CUHK Bookstore but it might take 4-6 weeks to arrive.

Diez, David M., Christopher D. Barr, and Mine Cetinkaya-Rundel. 2019. *OpenIntro Statistics*. 4th edition. https://www.openintro.org/book/os/

- PDF downloadable in free on Blackboard or a hard copy purchasable here: https://leanpub.com/os

Academic Honesty

Please keep in mind the university's policy on academic honesty. Plagiarism will not be tolerated in the term paper and assignments. The ideas and language should be your own, and any outside sources must be clearly and properly cited. There are severe consequences if you commit any acts of academic dishonesty. In addition to the department's policy and guidelines for citations, please refer to the university-level disciplinary guidelines and procedures. The Faculty of Social Science has also compiled a handout to alert students of the importance of academic honesty and the consequences of violating the University's Rules. To this end, the final term paper should be submitted to VeriGuide.

Schedule

Lectures

The schedule below is tentative and subject to minor changes.

Week	Date	Topic	Reading
1	Jan-11	Course Introduction	
2	Jan-18	Causality	QSS 1.1 ~ 1.4; 2.1. ~ 2.4
3	Jan-25	Measurement (1)	QSS 2.5 ~ 2.7; 3.1. ~ 3.4
4	Feb-01	Measurement (2)	QSS 3.4 ~ 3.8
5	Feb-08	LUNAR NEW YEAR	
6	Feb-15	Inference (1)	OpenIntro 5,6
7	Feb-22	Inference (2)	OpenIntro 7.5,8; QSS 7.1 ~ 7.2
8	Feb-29	Regression (1)	OpenIntro 9
9	Mar-07	READING WEEK	
10	Mar-14	Regression (2)	QSS 4.3.1 ~ 4.3.3
11	Mar-21	Regression (3)	Yu and Kuo 2017; QSS 4.3.4
12	Mar-28	Consultation Week*	
13	Apr-04	CHING MING FESTIVAL	
14	Apr-11	Text Data	QSS 5.1
15	Apr-18	Student Presentation	

* Students are required to schedule a meeting with the instructor to consult their final projects.

Tutorials

TBD. Tutorial times will be determined according to the tutor and students' availability. Around the end of the add/drop period, the tutor will distribute a time survey and students fill the form.