

SOCI 3229: Quantitative Data Analysis

Department of Sociology
The Chinese University of Hong Kong

2025-26 Term 2

Fridays 09:30 – 11:15
Chen Kou Bun Building 109

Contact Information

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

Teaching Assistant: Mr. TAN Qiang
E-mail: 115515TAN@link.cuhk.edu.hk
Office hours: By appointment
Office: Sino Building Room 424

Course Description

Data analysis is one of the most powerful ways to investigate social questions and understand human behavior. In an era defined by abundant data and complex social challenges, the ability to work with data is essential for making sense of the world and evaluating claims about it.

This course offers a hands-on introduction to quantitative data analysis for the social sciences. We focus on practical reasoning with data and learn how to pose empirical questions, measure social concepts, assess uncertainty, and build statistical models. Topics include causal inference, measurement and data quality, statistical reasoning and uncertainty, linear and multivariate regression, and introductory tools from machine learning and text analysis.

At the end of this course, students will be able to analyze data, interpret statistical 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%
Assignment	20%
In-class Exam	30%
Final Paper	40%

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.

Assignment (20%)

- There will be one coding and data analysis assignment in this course. It is designed to give you hands-on practice with the tools and concepts covered in class.
- You may discuss ideas and programming techniques with classmates, but all code and written answers must be your own.
- Submit your work on Blackboard,
- Late submissions are not accepted unless you notify the instructor at least 24 hours in advance.

In-class Exam (30%)

- One in-class examination will be held on **March 20**.
- The exam consists of a mix of true/false quizzes, multiple-choice questions, concept-blank/short-answer questions, and essay-style questions.

Final Paper (40%)

- The final paper is a team project (2 – 4 students in a team).
- 3 components:
 - Proposal: Submission by **April 2**
 - Presentation (10%): Class on **April 17**
 - Final Paper (30%): Submission by **May 12**
- 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 and Nora Webb Williams. 2022. *Quantitative Social Science: An Introduction in Tidyverse*. Princeton University Press, 2022.

- The original book is also fine: Imai, Kosuke. 2018. *Quantitative Social Science: An Introduction*. Princeton University Press.
- Loan: Course Reserve 4 hours at the Chung Chi College Library.

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>

Schedule

Lectures

The schedule below is tentative and subject to minor changes.

Week	Date	Topic	Reading
1	Jan-09	Course Introduction	
2	Jan-16	Causality	QSS 1.1 ~ 1.4; 2.1. ~ 2.4
3	Jan-23	Measurement (1)	QSS 2.5 ~ 2.7; 3.1. ~ 3.4
4	Jan-30	Measurement (2)	QSS 3.4 ~ 3.8
5	Feb-06	Inference (1)	OpenIntro 5,6
6	Feb-13	Inference (2)	OpenIntro 7; QSS 7.1 ~ 7.2
7	Feb-20	LUNAR NEW YEAR	
8	Feb-27	Regression (1)	OpenIntro 7.5, 8
9	Mar-06	READING WEEK	
10	Mar-13	Regression (2)	OpenIntro 9; QSS 4.3.
11	Mar-20	Regression (3) In-class exam	QSS 4.3.4
12	Mar-27	Consultation Week*	
13	Apr-03	GOOD FRIDAY	
14	Apr-10	Text Data	QSS 5.1
15	Apr-17	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.

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](#).

AI Use Policy

This course recognizes that AI tools are increasingly integrated into data analysis, programming, and academic work. At the same time, the goal of this course is to develop students' own skills in statistical reasoning, coding, and empirical interpretation. The following guidelines clarify how AI tools may be used responsibly throughout the term:

- *Grammar Check and Writing Polish* (Permitted): Students may freely use AI tools to improve grammar, clarity, and presentation of their writing. No prior permission or formal acknowledgement is required. However, students are expected to critically evaluate AI suggestions to ensure accuracy and appropriateness. This supports language development while maintaining academic standards.
- *Literature Search and Background Information* (Permitted): AI tools may be used to help identify relevant literature, clarify concepts, or locate background information. Students must verify the accuracy and credibility of any information before using it in their work. All cited sources must be real, retrievable, and properly referenced.
- *Programming Assistance* (Permitted with Restrictions): AI tools may be used to support learning and problem-solving in R, including: understanding syntax, debugging errors, clarifying how functions work, exploring alternative approaches. Students must, however, write their own code and be able to explain and reproduce any code they submit. Submitting AI-generated code without understanding it constitutes academic dishonesty, and copy-pasting full solutions from AI tools is not permitted. The purpose of coding and data analysis assignments is to help you develop your own computational skills; relying on AI as a shortcut undermines both your learning and the expectations of the course.
- *Data Analysis, Interpretation, and Substantive Reasoning* (Not Permitted): AI tools may not be used for data analysis, interpretation, or substantive reasoning. Students may not rely on AI to generate interpretations of statistical results, write analytical answers to assignment questions, produce explanations of findings, generate full code solutions for assignment problems, or create any substantive

content for the final paper. All analytical reasoning, interpretation of output, and written explanations must be entirely the student's own work.

- *Idea Generation and AI-Generated Content* (Not Permitted): AI tools may also not be used for idea generation or the creation of original intellectual content. Students may not use AI to generate research questions, hypotheses, arguments, substantive written material, full assignment answers, or any portion of the final paper text. All conceptual and analytical work must reflect the student's own independent thinking and effort.