# SOCI 3229: Quantitative Data Analysis
Department of Sociology
The Chinese University of Hong Kong

## 2022-23 Term 1

Mondays 09:30 – 11:15
T.C. Cheng Bldg C2

## Contact Information

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

Teaching Assistant: YUE Xiaoqian, M.A.
E-mail: yuexiaoqian@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 the 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, analyses of network, textual, and spatial data.

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.

This course will also expose students to recent data and computational revolutions leveraged by the emerging field of computational social science. This field collectively addresses longstanding sociological questions—many of which have been difficult to be unearthed by conventional survey methods. We will cover, for example, automated methods and visualization tools to identify patterns in network, textual, and spatial forms of data sets or "big data."

There is no prerequisite to enroll 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) | 15% |
| Assignments | 35% |
| Final Paper | 50% |

**Participation** (15%)
- Attendance will be a crucial part for your success, as the material builds on itself cumulatively throughout the course. Do not fall behind. If you start falling behind, see your tutor immediately, in order to catch up.
- To attend or skip is an adult's decision, so I do not want to punish your few times of absence/lateness with a harsh grade. However, I will highly compensate diligent and active students.
- Class starts promptly on time. Arrivals after 5 minutes of the class beginning will be counted as lateness.
- Equal weights will be given for participation in tutorials.

**Assignments** (35%)
- 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 done on Blackboard.
- For all assignments, late submission is not allowed without at least 24 hours prior notice.

**Final Paper** (50%)
- Final paper is a team project of two. You will be paired with another student to work together.
- 3 components:
  - Proposal (10%): Submission by **November 14**
  - Presentation (10%): Class on **November 28**
  - Final Paper (30%): Submission by **December 19**
- Your team is required to submit it in a form of social science research—the product of your own, original 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.
- 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

| Week | Date | Topic | Reading |
|------|------|-------|---------|
| 1 | Sep-05 | Introduction | |
| 2 | Sep-12 | NO CLASS: Mid-autumn Fest | |
| 3 | Sep-19 | Causality | QSS 1,2 |
| 4 | Sep-26 | Measurement I | QSS 3 |
| 5 | Oct-03 | Measurement II | QSS 3 |
| 6 | Oct-10 | Statistical Inference | OpenIntro 5 6 7 ; QSS 7.1. 7.2 |
| 7 | Oct-17 | Regression I | QSS 4.1, 4.2, 7.3; OpenIntro 8 |
| 8 | Oct-24 | Regression II | OpenIntro 9 |
| 9 | Oct-31 | CONSULTATION WEEK* | |
| 10 | Nov-7 | Regression III | QSS 4.3 |
| 11 | Nov-14 | Data Science I | QSS 5.1.,5.2,5.3 |
| 12 | Nov-21 | Data Science II | TBD |
| 13 | Nov-28 | 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.

### Major Dues

- Term Paper Proposal: November 14
- Final Term Paper: December 19