SOSC5090: Quantitative Methods to Social Science (Tentative)

Fall 2025

Tuesdays 3:00 pm - 5:50 pm

Rm 2126A, Lift 19

Instructor

WEI Jinlin

Contact Information: Rm. 2363; jinlinwei@ust.hk

Office Hours: Tuesdays 2:00 pm -- 3:00 pm (or by appointment)

Course Website: Canvas

Tutorial Instructor

Brian Ying Yeung HO

Contact Information: Rm. 2359; brianyyho@ust.hk.

Office Hours: Tuesdays 12:30 pm – 2:00 pm (or by appointment)

TA Session: Fridays 1:00 pm – 2:00 pm (location TBA)

Course Description

The purpose of this course is to introduce basic statistical concepts and applications that are heavily used in (quantitative) social science research. The course serves as a foundation for anyone who is interested in empirical research and also as a prerequisite to taking more advanced methodology courses (such as SOSC5340 and above).

After completing this course, you will be able to:

- Understand basic statistical concepts and quantitative methods used in social science research
- Use basic functions of R and (or) Stata to analyse real data
- Given a research question, use relevant data to test hypotheses, conduct statistical inference, and interpret regression results
- Clearly present your research work, in both oral and written formats

Required Textbook

The required textbook for this course is *Quantitative Social Science: An Introduction* by Kosuke Imai, which is available at the textbook service and also reserved in the library. Supplementary readings for the course come from several chapters and appendices (see the list of chapters in the course

outline) in *Introductory Econometrics: A Modern Approach, 4th edition (2009)* by Jeffrey Wooldridge (reserved in the library too, the electronic 7th edition published in 2018 is also reserved in the library). Other readings (such as research papers) will be uploaded to the course website when needed.

Software

You are required to use R and/or Stata to do statistical work in this course. While using Stata is acceptable for some exercises, we emphasize how to use R to answer quantitative social science questions in lectures. R is freely available for download and runs on Macintosh, Windows, and Linux computers. Students are strongly encouraged to use Rstudio, another freely available software package that has numerous features to make data analysis easier. Install package swirl and find some course exercises there. There are detailed descriptions about swirl in the textbook. Some of questions in problem sets are based on the exercises.

Additional resources:

• Download R: https://www.r-project.org/

• Rstudio: https://www.rstudio.com/

• An Introduction to R: https://cran.r-project.org/doc/manuals/R-intro.pdf

• UCLA IDRE: https://stats.idre.ucla.edu/r/

• Princeton R Tutorial: https://exploringr.princeton.edu/self-learning-resources-for-r/

• You can combine the Swirl exercises for this course with the exercises for another course named R programing.

• R for beginners. https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

You have access to STATA18 through Virtual Barn Desktop provided by HKUST. For many of you who have never used Stata before, the TA will help you get familiar with Stata in tutorials. It is also easy to train yourself given the rich self-learning resources available.

Here are some recommended resources:

• Hamilton, Lawrence C. (2006). Statistics with STATA. Belmont, CA: Duxbury Press.

• Princeton Stata Tutorial: https://www.princeton.edu/~otorres/Stata/

After all, the best way to learn software is **learning by doing**. The most useful tutorials are the official help files created by authors, search engines, and generative AIs.

Course Requirements and Grading

Your final grade in the course will be based on your performance on the following:

1. Assignments 10%
    a. We have four assignments through the semester in total. We pick the best two of the four. Each of the two accounts for 5% of the final grade.
    b. The main purpose of the assignments is to make you familiar with the statistical concepts and comfortable with processing the data in R.
    c. Submit via Canvas.
2. Quizzes 10%
    a. We have five quizzes throughout the semester in total and only count 4 with highest grades. Each accounts for 2.5% of the final grade.
    b. The main purpose of the quizzes is to make you familiar with the basic concepts and improve your skills in assessing empirical research projects.
3. Presentations 5%
    a. You are required to form a two-person group to present the classical projects included in the textbook (or provided via Canvas). You will be assessed based on a set of criteria which will be distributed in a separate document. Both people have the same score for one single presentation.
    b. Presentations will be recorded to facilitate feedback after lecture. Please tell the instructor before your presentation if you do not want to be recorded.
4. Class Participation 5%
    a. You can contribute to the presentations and discussions in the lectures. Participation in one lecture earns you 0.5 points. A maximum of 1.5 points can be earned in the student presentations in the last week (November 25th).
    b. You can participate in all the practice quizzes in one single lecture to earn 0.5 points.
5. Final exam (Time to be confirmed) 30%
    a. The exam will be made up of two sections. One in **written** and one in **coding** (R language).
    b. The coding exam allows students to bring printed and electronic materials, but does not allow students to access the Internet.
6. An independent term paper (replication and extension)
    a. Final paper presentation 5%
    b. Meeting the deadlines for writing the paper 5%
    c. Final paper write-up 30%

For the independent term paper, you will find an article that meets the following criteria:

1. It was published in the top journals in your field in or after 2015 and interests you.

2. It should be an article that enables you to learn more about countries which are very different from where you are from. For example, students who are from East Asia should find a paper based on other continents, like Europe, North America, or Africa.
3. It must provide replication materials on the website of the journal or on that (those) of the author(s). Please submit the title of the paper for approval before you proceed. You need to replicate using a different software from what the author(s) used.

Journal List:

Economics: American Economic Review, Quarterly Journal of Economics, Econometrica, Journal of Political Economy, Review of Economic Studies. **Not** AER: Insights, AEA Papers and Proceedings, JPE Micro, or JPE Macro.

Political Science: American Political Science Review, American Journal of Political Science, The Journal of Politics, British Journal of Political Science

Psychology: Nature, Science, Proceedings of the National Academy of Sciences of the United States of America, Psychological Review, Psychological Bulletin, Psychological Science, Annual Review of Psychology

Sociology: American Journal of Sociology, American Sociological Review, Social Forces, Demography

Your basic task is to replicate that article. You are encouraged to extend it, but an extension of the paper is not necessary to make you pass this course. The criteria of replication and extension will be distributed in a separate document. Only an electronic copy of your paper is accepted.

The aim of the replication project is to encourage you to read the good articles published in top journals in recent years and think of how you can contribute to the literature (and finally publish if needed in the future). In this way, you know what people worked on in the fields in which you have interests. By replicating the paper, you understand the motivations for the research question and why academia pays attention to the question. You will also learn a few strands of literature to which the paper contributes and how the author(s) figured out their contribution to the literature. You would also learn how good social scientists build causal relationships and use various research tools in their research papers.

The paper should first include (1) an introduction that includes the research question, motivations for the research question, what the author(s) did, a brief literature review summarizing the original article's contribution, and how the literature developed after this paper was published.

Next, I want you to (2) summarize the empirical method(s) that the author(s) used, explicitly discuss the assumptions, and how the author(s) argued that their paper is convincing.
Then, I would like you to (3) replicate **five** tables or figures that show summary statistics, the main results, causal relationships, and mechanisms (if there are still slots).
Lastly, please (4) briefly discuss the results that you replicate and the results provided in the article. You do not need to discuss the data collected and can just skip.
To show that you finish the replication process, you are also required to submit your data and codes along with your paper, although they will not be graded. The data and codes should be organized in such a way that I can **run the codes directly** if I just change the directory of your data and codes.
To make sure you will make a good progress and not leave everything till the last minute, please submit relevant materials via Canvas before the following deadlines. None of the materials submitted before the final paper will be graded, but please feel free to come to my office hours to discuss them.
1. Confirming the paper to replicate (1%)
Deadline: Submit before the lecture on **September 16th**

2. Basic replication results (1%)
Deadline: Submit before the lecture on **October 14th**
Submit the codes showing that you have replicated the main results, which is **the table** showing how x affects y, and the codes showing **the tables or figures** that **support the causal relationship**.

3. Wrap up your replication results, compare and discuss your results and the author(s)' results. (1%)
Deadline: Submit before the lecture on **November 11th**
Submit the draft for the replication results without the introduction. You should only correct grammatical errors after this deadline.

4. Find a related well-defined research question that you can use the replication data and other publicly available data (if any) to answer (1%)
Deadline: Submit before the lecture on **November 25th**
It can be a simple extension, or a different research question. Report your research question, your contribution to the literature, the additional data that you will use (if any), and your empirical

methods. It is a simplified version of a proposal, but I would suggest you start looking for relevant data from the start of the semester.
Note: Publicly available data includes multiple sources, like public database, replication data of research papers, and data shared by scholars on their websites. This is not an exhaustive list.

5. Wrap up your own empirical results, and write the introduction (1%)
Deadline: Submit before the **final deadline (to be confirmed).** Missing this deadline will make me assess your replication paper based on your last submission (hopefully the one submitted by the fourth deadline).

The 5 deadlines are checkpoints. If you miss one deadline, you lose the point that corresponds to the specific deadline. The final deadline is not confirmed yet. I will revise according to the examination dates.

Problem set collaboration policy
Discussions with other people (whether via the Internet or in person) are permitted and strongly encouraged.
The use of generative AIs (HKUST has its own generative AI. Try it!) is also permitted in helping you understand the materials and writing codes.
However, you must write your own homework by yourself. Do not, under any circumstances, copy another person's code. Incorporating someone else's code into your program in any form is a violation of academic regulations. Abetting plagiarism or unauthorized collaboration by sharing your code is also prohibited. Do not e-mail your code to ANYONE!!!
Novices often have the misconception that copying and mechanically transforming a program (by rearranging independent code, renaming variables, or similar operations) makes it something different. Actually, identifying plagiarized source code is easier than you might think. For example, there exists computer software that can detect plagiarism. If you have any questions about these matters, please consult the course instructor.
Please remember that we are having an exam that makes up of 30% of the assessment. It will include a coding section. You need to understand concepts, do maths, analyse data, and write your own codes to earn your points.

Course Outline and Tentative Schedule

Note: **QSS** stands for Quantitative Social Sciences: An Introduction and **JW** stands for Introductory Econometrics: A Modern Approach. **Ch** stands for chapters and **Ap** stands for Appendix.

1. Week 1: Motivation, Overview and a brief introduction to R (QSS Ch1, JW Ap A)
2. Week 2-3: Causality (QSS Ch2)
    a. An example of randomized trials
    b. Observational studies
3. Week 4-5: Measurement and Research Design (QSS Ch3, JW Ap B)
    a. Survey sampling and its potential biases. e.g., Measuring public opinions through sample surveys
    b. Clustering

==========Holiday on October 7th==========

4. Week 6-7: Prediction (QSS Ch 4, JW Ch2-5)
    a. Prediction and loop
    b. Linear regression models
5. Week 8-9: Probability theory, random variables, and probability distribution (QSS Ch 6, JW Ap B)
    a. Probability and conditional probability
    b. Random variables and their distributions, large number theorems
6. Week 10-11: Uncertainty (QSS Ch 7, JW Ap C)
    a. Estimation
    b. Hypothesis testing and confidence intervals
    c. Regression with uncertainty
7. Last week: student presentations