

SOSC 2400 **Quantitative Data Analysis for Social Research II**

Fall 2020

Monday & Wednesday, 10:30-11:50am

Zoom

Instructor: Dr. WANG, Hongbo (hbwang@ust.hk)
Office: Academic Building, Room 2372 (Ext. 7804)
Office Hours: Thursday, 2-3:00pm

TA: Shen, Minghong (mhshenaa@connect.ust.hk)
Office: Academic Building, Room 3001
Office Hours: Tuesday, 10:30-12:00nn (via Zoom, by appointment)

Course Description and Objectives:

This course introduces applied multiple linear regression (MLR) for social scientific research. The focus will be on the specification of models including choice of variables, incorporation of variables in different forms, and the interpretation of results. We leave the issue of inference almost entirely to more formal statistics courses.

This course will also address the challenges posed by omitted variable bias, endogeneity, and other issues commonly arising in the analysis of social data. We provide a brief discussion on selected solutions to some of the issues.

Organization:

The class will meet twice a week on Monday and Wednesday, respectively, each lasting for one hour and twenty minutes. The lectures are given on Mondays, while Wednesdays are reserved for computing sessions, in tandem with the lectures (See “Course Schedule” below for details).

The instructor will regularly post course materials on [Canvas](#), including lecture notes and computing sessions files. You will need an ITSC account to access these materials. All materials should be used *exclusively* for this course, with all copyrights reserved. Please do not distribute any material for any other purposes.

Students are encouraged to form groups of 5-6 individuals to collaborate on a group project, the main assessment component of this course. Each group member must contribute adequately to the project. You will fail the project, and the whole course as well, if the rest of the group reports lack of adequate contribution from you.

Computing:

We will mainly use R as computing tool.

Prerequisite:

SOSC 1100

References:

Baumer, Benjamin S., Daniel T. Kaplan, and Nicholas J. Horton. 2017. *Modern Data Science with R*. Chapman and Hall/CRC. [BKH]

Imi, Kosuke. 2018. *Quantitative Social Science: An Introduction*. Princeton University Press. [I]

Navarro, Danielle. [Learning statistics with R: A tutorial for psychology students and other beginners](#) (Version 0.6). [N]

Treiman, Donald J., 2009. *Quantitative Data Analysis: Doing Social Research to Test Ideas*. Jossey-Bass. [T]

Wooldridge, Jerrey, 2013. *Introductory Econometrics: A Modern Approach* (5th Ed.). South-Western. [W]

Assessment:

Your grade will be determined as follows:

(1) Attendance and class participation: 30%

Attendance is required for both regular lecture and computing session. You will get one point deducted for each *instance of lack of participation in class* (e.g. missed lecture, non-response to questions).

(3) Group project: 70% (presentation, 20%; written report, 50%)

Under the supervision of the instructor, each group will choose a topic of their own, locate appropriate data sources, carry out data analysis, present the findings, and, finally, submit a written report. Detailed guidelines will be provided in a separate document.

Please keep a diary of your work on the project which describes all related activities, including exploration of literature and data, downloading and processing of data including data that you decided not to use, any calculations you are doing, including ones that didn't work out. This should document that you have spent the expected amount of time on the project, including on data and calculations that are not reflected in the final report. The records should be specific on who did what.

Each group should submit a brief report on progress via Canvas *every two weeks starting in Week 6* (See the deadlines below).

Course Schedule (*Subject to adjustment*)

Calendar Week	Topic	Readings	Deadlines
Week 1: Monday Wednesday	Introduction NO CLASS!		
Week 2: Monday Wednesday	Correlation vs Causality [R] Computing session		Project Group Finalized (9/14)
Week 3: Monday Wednesday	Motivations for MLR [R] Computing session		
Week 4: Monday Wednesday	Choice of Variables [R] Computing session		
Week 5: Monday Wednesday	Logarithmic Transformation [R] Computing session		Proposal Due (10/5)
Week 6: Monday Wednesday	Incorporating Curvilinear Relationship [R] Computing session		<i>Progress report (10/14)</i>
Week 7: Monday Wednesday	Dummy Variable [R] Computing session		
Week 8: Monday (10/26) Wednesday	NO CLASS! [R] Computing session		<i>Progress report (10/28)</i>
Week 9: Monday Wednesday	Categorical Variable as X [R] Computing session		
Week 10: Monday Wednesday	Interaction Effect [R] Computing session		<i>Progress report (10/11)</i>
Week 11: Monday Wednesday	Model Goodness-of-Fit [R] Computing session		
Week 12: Monday Wednesday	Pitfalls in Regression Analysis [R] Computing session		<i>Progress report (11/25)</i>
Week 13: Monday Wednesday	*Project Presentation *Project Presentation		
Week 14:			
Week 15:			Final Report Due (12/18)