

SOSC 2400 Intermediate Statistics for Social Research

Fall 2019

Tuesday & Thursday, 4:30 - 5:50pm
Room 2404, Lift 17-18, Academic Building

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

TA: Tian Bing (bing.tian@connect.ust.hk)
Office: Academic Building, Room 3001
Office Hours: Monday, 14:00-15:00pm

Course Description and Objectives:

This course introduces and applies analysis of survey and other types of data of interest to social science researchers. The topics of the course can be divided into two parts. The first part deals with statistical inference with survey data. Topics include t-test, contingency table, multiple linear regression, logistic regression, and approaches to causal inference. The focus will be on implementation, including data manipulation, interpretation of results, and challenges posed by endogeneity, in particular, as well as other issues that commonly arise in the analysis of social data.

The second part introduces analytical strategies for “unconventional” data types, such as social network, spatial data, and plain texts. Special attention will be given to challenges to social scientists regarding data extraction, data integration, and visualization.

Organization:

The class will meet twice a week on Tuesday and Thursday, respectively, each lasting for one hour and twenty minutes. Lectures are given on Thursdays, while Tuesdays are usually reserved for computing sessions, in tandem with the lectures (See “Course Schedule” below for details).

Course materials will be regularly posted on [Canvas](#), including lecture notes and assignments. You will need an ITSC account to access these materials. All materials should be used exclusively for the purpose of this course, with all copyrights reserved. Please do not distribute any material for any other purposes.

Students will form groups of 5-6 to collaborate on the group project. Students are also encouraged to help each other with class materials and computing exercises.

Computing:

We will mainly use R as computing tool.

Prerequisite:

SOSC 1100

References:

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

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

Moore, David S., George P. McCabe and Bruce A. Craig. 2014. *Introduction to the Practice of Statistics*. (8e.) New York: W. H. Freeman & Co. [MMC]

YouTube video: <https://www.youtube.com/user/mrdarkstatstudio>

Assessment:

Your grade will be determined as follows:

(1) Attendance and class participation: 20%

Attendance is required for all regular lectures and computing sessions. Students are also expected to participate in class activities. We will take attendance via iPRS. One point will be deducted for each lecture missed. Please note that *you will receive an F automatically on this course if you have missed half or more classes*.

(2) Midterm exam: 20%

The mid-term exam will be closed-book. *Please bring your laptop for the exam*. The midterm exam will draw exclusively on lectures from the first half of the course.

(3) Group project: 60% (Oral presentation, 20%; written report, 40%)

Under the instructor's supervision, 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.

Course Schedule (*Subject to adjustment*)

Calendar Week	Topic	Readings	Deadlines
Week 1: Tuesday Thursday	Introduction Review		
Week 2: Tuesday Thursday	[R] Computing session Data Generating Process		Project Group Finalized (9/19)
Week 3: Tuesday Thursday	[R] Computing session Statistical Inference		
Week 4: Tuesday Thursday	[R] Computing session Inference for Two-way Table		
Week 5: Tuesday Thursday	*NO CLASS Multiple Linear Regression		
Week 7: Tuesday Thursday	[R] Computing session Logistic Regression		
Week 6: Tuesday Thursday	[R] Computing session Causal Inference		
Week 8: Tuesday Thursday	*MIDTERM EXAM (10/22) *Proposal Defense		
Week 9: Tuesday Thursday	[R] Computing session Text Analytics		
Week 10: Tuesday Thursday	[R] Computing session Social Network Analysis		
Week 11: Tuesday Thursday	[R] Computing session GIS		
Week 12: Tuesday Thursday	[R] Computing session Visualization		
Week 13: Tuesday Thursday	*Project Presentation *Project Presentation		
Week 15: Tuesday			Final Report Due (12/10)