

## SOSC5090: Statistics in Social Science

Fall 2017 (tentative)

Rm CYTG001

Fridays 1:30-4:20 pm

### Instructor

HAN Li

Contact Information: SOSC Rm. 3390; Tel: 2358-7838; [lihan@ust.hk](mailto:lihan@ust.hk)

Office Hours: Fridays 10:00-11:00 am (or by appointment)

Course Website: Canvas

### Tutorial Instructor

Mr. CHAN, Chi Shing, Bill

([chischan@ust.hk](mailto:chischan@ust.hk))

Office Hours: Wednesdays 2-4 pm @

Rm3005

### 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 methods used in social science research
- Use basic function of Stata (a popular statistical package) to analyze 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 Text

The required textbook for this course is *Statistical Methods for the Social Sciences*, 4e (2009) by Alan Agresti and Barbara Finlay, which is available at the bookstore 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*, 4e (2009) by Jeffrey Wooldridge (reserved in the library too). Other readings (such as research papers) will be uploaded to the course website when needed.

### Software

You are required to use Stata to do statistical work in this course. (Although you may find it useful to learn by yourself other packages such as SAS or SPSS, all the course work must be done using Stata.) Stata12 is installed in all PCs at social science computing lab (Rm. 3001). For many of you who have never used Stata before, the TA will give several lab tutorials to help you get familiar with Stata. 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. (A copy is available for loan at TA's office.)
- UCLA Stata Portal (an extensive resource that leads you to many useful links):  
<http://statcomp.ats.ucla.edu/stata>  
<http://www.ats.ucla.edu/stat/stata/sk> (Starter Kit section for new users)
- Princeton Stata Tutorial:  
<http://data.princeton.edu/stata/>
- UNC Carolina Population Center:  
<http://www.cpc.unc.edu/services/computer/presentations/statatutorial/>

### Course Requirements and Grading

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

- |   |     |
|---|-----|
| • Assignments                                 | 10% |
| • A take-home exam (on <b>December 2nd</b> )  | 40% |
| • An independent term paper/replication paper |     |
| ✓ Final paper presentation                    | 25% |
| ✓ Final paper write-up                        | 25% |

We have six assignments through the semester in total. Five of the six assignments will be counted in the final grade. Note that the weight of assignments in the final grades is fairly small. The main purpose of the assignments is to make you familiar with the statistical concepts and comfortable with processing the data.

For the independent project, you will find a published article that interests you, preferably but not necessarily related to one of the topics covered during the semester. I will suggest some articles. You are also encouraged to find an article in your own field. Your basic task is to replicate that article. Replication means obtaining the original data the author(s) used (or essentially similar data, but **NOT** the data extract accompanying the article that can be downloaded from some journals' websites), reconstructing the sample(s) the author(s) used for statistical/regression analysis, repeating the main results the author(s) obtained, performing some robustness/sensitivity tests the author(s) conducted (for example, those tests only mentioned in the article text but not shown in the tables -- typically for space concerns), and (encouraged but not required) even doing some updating or extension work if you think the article has limitations or room for improvement.

At first, this project may sound like a mechanic exercise, but you will find that it is challenging and something from which you can learn much. You will learn about the problems encountered in empirical research and see that published results are often not nearly as neat and clean as they seem. Indeed, replication of existing findings is where you should usually begin any empirical project.

Your grade of the replication project will be based on an oral presentation in the class (**24 November**) and a written paper due on **8 December**. The paper should be no more than 30 (double-spaced) pages and should at least include an introduction, a brief literature review summarizing the original article's (or/and your replication's) contribution, a description of data and sample construction, a comparison of original results and your replication (including relevant discussions), and references, tables, and figures (if applicable). Only a printed copy of

your paper is accepted. For my records, you are also required to submit your data and Stata do-files along with your paper, although they will not be graded. More guidelines on how to present and write up your work will be provided during the semester.

I expect each of you to meet with me to discuss about your projects, and I will be happy to offer advice and help at any stage of the process. To make sure you will make a good progress and not leave everything till the last minute, please notify me (by email is sufficient except for the final paper) of the following stages by their dates (none of the stages before the final paper will be graded).

- 30 September: 1 paragraph (article to replicate, data to use, and specific aims)
- 27 October: descriptive statistics; in-class **presentation** of preliminary results
- 17 November: replication results for presentation; an outline of the paper
- 24 November: presentation
- 8 December: final paper (with data and do-files) due

### **Course Outline and Tentative Schedule (by week)**

Note: **AB** stands for *Statistical Methods for the Social Sciences* and **JW** stands for *Introductory Econometrics: A Modern Approach*. **Ch** stands for chapters and **Ap** stands for Appendix.

1. Motivation and Preliminaries (AB Ch1, JW Ap A)
2. Sampling and Measurement; Research Design (AB Ch2-3, JW Ap B)
3. Descriptive Statistics (AB Ch2-3, JW Ap B)
4. Probability Distributions (AB Ch4, JW Ap B)
5. Statistical Inference I: Estimation (AB Ch5, JW Ap C)
6. Statistical Inference II: Hypotheses Testing & Comparison of Two Groups (AB Ch6-7, JW Ap C)
7. Student presentation of preliminary results of replication paper
8. [Association between Categorical Variables]; Simple Regression (AB Ch8-9, JW Ch2)
9. Multiple Regression I (AB Ch10, JW Ch3)
10. Multiple Regression II (AB Ch11, JW Ch4)
11. Multiple Regression III (AB Ch12-14, JW Ch5-6)
12. Student Presentations: 24 Nov.