SOCI 328 SOCIAL STATISTICS I (SECTION 101)
SEP – DEC 2017

Instructor: Dr. Gerry Veenstra
Office: ANSO 1321
Phone: (604) 822-4351
Email: gerry.veenstra@ubc.ca

Purpose

This course introduces students to elementary techniques of quantitative data analysis common in sociological research. It emphasizes selection of appropriate statistical techniques, examination of assumptions associated with them and interpretation of the results provided by them. The course will not emphasize complex calculations or memorization of complex formulae. It might even be fun on occasion.

Calendar Entry

The testing of sociological theories using quantitative data analysis techniques on numerical data from social surveys, experiments and official statistics. SOCI 328 excludes credit for a number of other statistics courses in various departments. Please consult the Science Credit Exclusion List before registering. Prerequisite: One of SOCI 100, SOCI 101, SOCI 102.

Prerequisite and Anti-requisites

Three credits of 100-level sociology is the only prerequisite for this course. The Science Credit Exclusion List in the UBC Calendar indicates that UBC students cannot apply credits from any two of the following introductory statistics courses to their UBC degree: STAT 200, 203, BIOL 300, COMM 312, 291, ECON 325, EPSE 482, 483, FRST 231, GEOG 374, KIN 371, POLI 380, PSYC 218, 366, SOCI 328.

Format

We will meet on Mon/Wed from 12:30–1:50 pm in ANSO 207. Classroom time will entail a mixture of lectures, quizzes, problem-based exercises and tutorials.

Textbook (required)

Garner, Roberta. 2010. *The Joy of Stats. Second Edition*. Toronto: The University of Toronto Press. This soft-cover textbook is available (new or used) at the UBC Bookstore, in ebook format at the University of Toronto Press website and in Kindle format at Amazon.ca.

Statistical Software

We will use the statistical software package Stata. Stata 12 is available in Buchanan computer labs B101, B121, B123, B125 and B126. You can also purchase Stata 15 for yourself from Statacorp (the IC version will do); this might be a worthwhile investment for students who also intend to take SOCI 380 Sociological Methods: Survey Research and/or SOCI 418 Social Statistics II. You are also welcome to use the free statistical software JASP (available for download at https://jasp-stats.org) which will allow you to work on the assignments at home; please note that we will not spend much time working with JASP in class.

Evaluation

Evaluation will be based on attendance (5%), quizzes and exercises (20%), two assignments (10% each), two midterm exams (15% each) and a final exam (25%). Attendance, quizzes and exercises will be implemented in *Top Hat* (join code 099694). The assignments involve implementing statistical techniques in Stata using real data. The exams will be comprised primarily of multiple choice and short answer questions. There will be no opportunities for extra credit. Final grades may be scaled up but will not be scaled down.

Attendance

Students are expected to regularly attend classes and to write examinations at the scheduled times. Students who encounter medical, emotional or personal problems that affect their ability to attend classes or write examinations should contact the Faculty of Arts Academic Advising Services. **Per UBC policy, students who miss a substantial number of classes (in this case, more than five classes) will not be permitted to write the final exam unless Arts Academic Advising formally recommends academic concession in this regard. A make-up midterm exam will be scheduled only if Arts Academic Advising formally indicates that this is appropriate. Students who are unable to write the final examination at the scheduled time should apply to Arts Academic Advising for AEG (aegrotat) or SD (deferred standing); if SD is granted a make-up final exam will be held during the formal SD exam period for the Winter session (typically in late July or early August).** UBC accommodates students with disabilities who have registered with the Disability Resource Centre. The university also accommodates students whose religious obligations conflict with attendance, submitting assignments or completing scheduled tests and examinations. A list of religious holidays involving fasting, abstention from work or study or participation in religious activities is available on the UBC website. Students should let the instructor know in advance (in the first week of the course) if they will require accommodation on these grounds. Students who plan to be absent for varsity athletics, family obligations or other commitments should not assume they will be accommodated and should discuss their commitments with the instructor early in the term. Please refer to the UBC Academic Calendar for a more thorough discussion of academic concession.

Topics and Timeline

Unit 0 – Introduction

- *Topics:* statistics in sociology; numerical literacy

Unit 1 – Basic concepts

- *Topics:* units of analysis; variables; levels of measurement; conditions for causality
- *Required readings:* Garner pp. 17–45
- *Supplementary readings:* Garner pp. 287–317 (math refresher)

Unit 2 – Populations and samples

- *Topics*: descriptive and inferential statistics; parameters and statistics; types of samples
- *Required readings:* Garner pp. 46–53

Unit 3 – Summarizing a categorical variable

- *Topics*: frequency tables; pie charts and bar charts
- *Required readings:* Garner pp. 55–57, pp. 70–73

Unit 4 – Summarizing an interval-ratio variable

- *Topics*: central tendency (mean, median, mode); dispersion (range, interquartile range, standard deviation); shape (histogram, stemplot, boxplot); transformations
- *Required readings*: Garner pp. 58–67, pp. 73–79

First midterm exam (Units 1 – 4) ***Oct 11***

Unit 5 – Basic concepts in statistical inference

- *Topics*: randomness; probability models; normal distributions; central limit theorem; sampling distributions of means and proportions
- *Required readings*: Garner pp. 87–117, pp. 121–127

Unit 6 – Statistical inference (confidence intervals)

- *Topics*: confidence intervals for means and proportions
- *Required readings*: Garner p. 128
- *Supplementary readings*: Garner pp. 130–135

Unit 7 – Statistical inference (tests of significance)

- *Topics*: logic of hypothesis testing; null and alternative hypotheses; test statistics; p-values and alpha levels; Type I and Type II errors
- *Required readings*: Garner pp. 128–129, pp. 135–142

Second midterm exam (Units 5 – 7) ***Nov 6***

Unit 8 – Association between two categorical variables

- *Topics*: cross-tabulations; % differences; Chi-squared test of significance; Cramer's V; Kendall's tau-b
- *Required readings*: Garner pp. 153-155, p. 157, pp. 191–199

Unit 9 – Association between two interval-ratio variables

- *Topics*: scatterplots; Pearson's *r*; OLS regression; coefficient of determination; Spearman's rho
- *Required readings*: Garner pp. 156, pp. 166–186

Unit 10 – Association between a categorical variable and an interval-ratio variable

- *Topics*: comparing central tendencies, dispersions and shapes; one-way ANOVA
- *Required readings*: Garner pp. 157–158, pp. 203–209

Unit 11 – Multiple regression

- *Topics*: multivariate causal scenarios; multiple regression
- *Readings*: Garner pp. 159–165, pp. 186–190
- *Supplementary readings*: Bryson 1996; Christensen & Carpiano 2014

Final exam (Units 1 – 11) held during the formal exam period in December

Lab bookings

The Buchanan Computer Lab B125 has been booked for our use at the following times:

- 9:00-11:00 am Wed Nov 8th
- 12:00-2:00 pm Thurs Nov 9th
- 12:00-2:00 pm Thurs Nov 23rd
- 9:00-11:00 am Mon Nov 27th

Assignments

Assignment 1 (due **Nov 15**): Use Stata (or JASP) to investigate the bivariate association between the categorical variables NOC2006_C10 (National Occupational Classification of the respondent) and incmhsd (household income) in the GSS dataset.

1. Provide a theoretical rationale to guide your analysis. Why would you expect the two variables to be related? How would you expect them to be related? Are they likely to be strongly or weakly related, do you think? Is one likely to cause or influence the other, i.e., does it make sense to explicitly designate one as the independent variable and the other as the dependent variable? (Do this *before* examining the relationship between the two variables.)
2. As far as you can tell, what, exactly, do the variables measure or assess? What are their levels of measurement? What do their distributions look like in this dataset?
3. Using statistical techniques covered in the lectures and textbook, describe and summarize the relationship between the two variables. You may have to recode or transform a variable first to make the relationship more intelligible. How are the

variables related to one another (if at all)? If they *are* related, is the relationship strong or weak (or something in between)?

4. Provide some interpretive insights regarding the relationship between the variables. Were your theoretical expectations met?

Present your analyses and insights in sentence and paragraph form (accompanied by attractive graphs and tables) as if you were writing a formal report for public consumption. Please type your double-spaced assignment using Times Roman 12-point font. You are permitted to work in teams of two or three people and hand in one version of an assignment on behalf of both or all of you. (You can also work alone if you so desire.) Please ensure that co-authored assignments are truly co-authored and alert the instructor if this is not the case. Assignments should be submitted to the course instructor or at the Department of Sociology main office (not to someone else, not by email, etc.). Late assignments will be penalized 10% per day (a weekend counts as one day). Assignments handed in after 3:00 pm on the due date are deemed to be one day late.

Assignment 2 (due **Nov 29**): Use Stata (or JASP) to investigate the bivariate association between the interval-ratio variables suicidert (suicide rate) and richpoor (ratio of share of total income for richest and poorest 10% of population) in the Country dataset.

1. Provide a theoretical rationale to guide your analysis. Why would you expect the two variables to be related? How would you expect them to be related? Are they likely to be strongly or weakly related, do you think? Is one likely to cause or influence the other, i.e., does it make sense to explicitly designate one as the independent variable and the other as the dependent variable? (Do this *before* examining the relationship between the two variables.)
2. As far as you can tell, what, exactly, do the variables measure or assess? What are their levels of measurement? What do their distributions look like in this dataset?
3. Using statistical techniques covered in the lectures and textbook, describe and summarize the relationship between the two variables. You may have to recode or transform a variable first to make the relationship more intelligible. How are the variables related to one another (if at all)? If they *are* related, is the relationship strong or weak (or something in between)?
4. Provide some interpretive insights regarding the relationship between the variables. Were your theoretical expectations met?

Present your analyses and insights in sentence and paragraph form (accompanied by attractive graphs and tables) as if you were writing a formal report for public consumption. Please type your double-spaced assignment using Times Roman 12-point font. You are permitted to work in teams of two or three people and hand in one version of an assignment on behalf of both or all of you. (You can also work alone if you so desire.) Please ensure that co-authored assignments are truly co-authored and alert the instructor if this is not the case. Assignments should be submitted to the course instructor or at the Department of Sociology main office (not to someone else, not by email, etc.). Late assignments will be penalized 10% per day (a weekend counts as one day). Assignments handed in after 3:00 pm on the due date are deemed to be one day late.