SOCIAL STATISTICS I (SOCI 328/SECTION 202)
(3 credit course)

JAN – APR 2020

Class Meeting: Mondays 2-5pm in ANSO 207
Class format: Lecture, discussion and in-class activities.

Instructor: Guy Stecklov
Office: ANSO 2115
Email: guy.stecklov@ubc.ca
Office Hours: Wednesdays 9-10am
(please email me/TA for additional options)

Teaching Assistant: Jiaxin Gu
Office: TBD
Email: TBD

Course Description

This course aims to provide students in sociology with an introduction to quantitative data analysis. You will learn as sociologists to use to identify and substantiate patterns and causal pathways in social and behavioural relationships. The language and structure in this course builds from statistics, but the emphasis in this class is on *application* of basic statistical tools to explore and understand quantitative data. The data we use are often drawn from real data sources relevant to sociologists and we focus on learning how these methods can be utilized to investigate social processes.

> *Acknowledgement*
> *UBC's Point Grey Campus is located on the traditional, ancestral, and unceded territory of the xwməθkwəy̓əm (Musqueam) people. The land it is situated on has always been a place of learning for the Musqueam people, who for millennia have passed on in their culture, history, and traditions from one generation to the next on this site.*

Who am I

I am a Professor and Head of the Department of Sociology at UBC since 2017. Prior to that I spent nearly two decades as a Professor of Sociology at the Hebrew University in Jerusalem. My research analyzes how demographic outcomes such as births and migration are affected by social and economic changes in very poor and under-resourced sections of the world. My work has been most focused on sub-Saharan Africa, and often in the rural sector. I primarily use quantitative research methods and often rely on unique and rich data sources, partly because I am strongly driven to understand the "causal" effects of social changes on behavior. This drive has pushed me over the past decade to get involved in designing various types of experiments - a relatively novel approach in our discipline – and one that can be very informative and offer more robust "causal" insights.

Computers and software

This course emphasizes the acquisition of basic statistical understanding. Statistics is best digested through application. In this course, students will be constantly using data of sociological relevance to apply their statistical learning. It is important to note that students are <u>required</u> to bring laptops (tablets may work but laptops are preferred) to class in order to complete the necessary lab work which forms a core part of this training throughout the semester. Laptops are also required in order to take the quizzes.

The statistical software we use in this section of this course is a key ingredient. There are a great many available statistical software programs that can be used for a course on introductory statistics. In this section of Soci 328 we will be using <u>R and RStudio</u>. R is arguably the most important <u>free</u> software today for exploring data, statistical analysis and data visualization. R is a programming language environment, thus it requires more investment to gain a basic level of comfort, particularly for students with less background in computing. However, the rewards from this investment are enormous. One is that it enables a richer approach to understanding data. It also develops deep thinking and reproducibility in scientific work. The R programming environment's popularity has grown tremendously and has made it an excellent tool for sociologists and other social researchers keen on enhancing their job market potential, partly as it has become a dominant tool in the world of Big Data. Most importantly, it is free software and can be installed in any computer or used from the cloud at no charge. RStudio is a free software package that works on top of R and makes it more friendly and less time consuming to use R. The RStudio option enables us to work on a cloud version requiring no software installation.

<u>Prerequisite and Anti-requisites</u>
Three credits of 100-level SOCI 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 degrees: STAT 200, 203, BIOL 300, COMM 291, ECON 325, EPSE 482, 483, FRST 231, GEOG 374, KIN 371, POLI 380, PSYC 218, 278, 366, SOCI 328.

<u>Learning Objectives</u>[1]
The overall aim of this course is to develop student's ability to apply basic statistical tools to real data sources and to uncover relationships between those data. At the same time, it also means learning to distinguish sound and faulty statistical reasoning. Specifically, objectives for this course include:
1. Remember key terms in statistics that are necessary for understanding dataset construction and data analysis.
2. Interpret descriptive statistics by looking at summary statistics (such as mean, median, mode, variance and IQR) as well as using common graphical tools (such as boxplots, histograms).
3. Ability to explain and apply the concepts relating to the sampling distribution of a statistic including types of error, behavior of the sample mean, and statistical significance
4. Describe and explain key issues in basic causal analysis and problems that may arise to complicate any causal interpretation.

---

[1] Adapted from McKinney et al. 2004 and Schutt, Blalock, and Wagenaar 1984

5. Demonstrate basic understanding of application of R/RStudio to data to explore and interpret statistical relationships with sociological data.
6. Apply basic statistical tools to bivariate and multivariate datasets.
7. Analyze various sociological datasets to provide insight into behavioral relationships.
8. Interpret critically quantitative analyses involving regression from peer-reviewed research studies.

Evaluating Student Performance

There are two main components to the grade for this course: quizzes and problem sets. Your grades will be based on the sum of the grades on these two components. The learning over the course of the semester is cumulative, so you can expect both the quizzes and the problem sets to force you to accumulate an understanding. Thus, while I will focus on each quiz on the main items learned since the last quiz, you cannot expect to forget what you learned in earlier weeks.

a. Quizzes: A quiz will be administered every 2-3 weeks during the semester and will be distributed at the start of the class. Because of the nature of the course material, the content is inherently cumulative. My focus in each quiz will be on new material since the prior quiz, although it may require you to work with material learned earlier in the semester. The quizzes will be taken on your laptops or tablets (Ipads only) through canvas. The quizzes will be primarily multiple choice and short answer questions and will take between 20 and 30 minutes depending on the quiz. They will be closed book and without notes.
   In order to succeed, students will have to study both notes and material from class as well as the textbook. Grading will be done so that students receive 60 points in total from the quizzes (Total is 60% of course grade). Each quiz is worth 15 points and I will take your 4 top grades out of 5 quizzes for calculating the quiz portion of your grade for the course over the semester. This approach allows you to have an off-day or to miss a quiz without adding additional stress to your semester.
b. Problem Sets: Much of the remainder of your grade will be based on two problem sets. Each of the two problem sets will be worth 20 points so that the sum will provide the remaining 40% of the course grade. The problem sets will push you to engage with statistical concepts learned in class along with real data and to use R to solve problems and present your results. These problem sets can be completed in groups of 2 or 3 persons but all names must be written in at time of submission.

Calendar of Dates for Evaluation Material	
January 27	Quiz 1
February 24	Quiz 2
March 2	Assignment 1 Distributed
March 9	Quiz 3
March 16	Assignment 1 Due
March 23	Quiz 4
March 30	Assignment 2 Distributed
April 6	Quiz 5
April 13	Assignment 2 Due

Textbook

1. Frankfort-Nachmias and Anna Leon-Guerrero. 2018. Social Statistics for a Diverse Society. Eighth Edition. Los Angeles: Sage Press. Available at the UBC Bookstore and in Kindle format at Amazon.ca. **(SSDS). (**Note, there are many editions of this book. Unfortunately, there are too many revisions for me to advise you to proceed and purchase any edition prior to the 8th)
    ➔ *This is the main textbook and is required for the course. It is a good introduction and not too difficult. The first chapter will be made available as a pdf. Two more advanced chapters* are also freely available online by the publisher.
2. Wheelan, Charles. 2013. *Naked Statistics*: Stripping the dread from the data. Norton Press: New York, NY. **(NS)**
    ➔ *This book is offered as optional reading. It is not technical and takes a more chatty, popular perspective to attempt to clarify many of the concepts we will cover over the course of the semester.*

Course Materials

All the material presented in the syllabus is tentative as events and the pace of learning may necessitate changes. In that case, I am committed to keeping you updated and providing an updated syllabus. I will be using canvas in this course. Powerpoint slides for the lectures will be posted on the canvas site in pdf versions at least one hour before class. In the past, students have found that using the slides as a tool for note taking can be an effective manner to concentrate on concepts and material learnt and spend less time frantically writing notes. It should also be absolutely clear that my own interpretations and definitions and teaching may differ at times from that presented in the textbook. Where differences arise, my own explanations are the ones that you are responsible for unless otherwise stated.s

TITLE	DATE	TOPICS	READINGS
1. INTRODUCTION	Jan 6	a. Statistics in sociology b. Best practices for learning statistics c. The research cycle d. Course overview and course evaluation e. Math refresher f. Our tools: paper/pencil/head & R/RStudio g. Application: Installing and understanding R and RStudio setup; First steps and basic calculations	SSDS Ch. 1 SSDS App F
2. THE ORGANIZATION AND GRAPHIC PRESENTATION OF DATA	Jan 13	a. Frequency Distributions b. Rates c. Interval Boundaries d. Reading Research with Statistical Tables e. Graphic Presentation of Data f. Application: Creating and modifying data and variables in R	SSDS Ch. 2
3. MEASURES OF CENTRAL TENDENCY	Jan 20	a. Mode b. Median c. Mean d. Application in R: Calculating central tendency measures e. Choosing a Measure of Central Tendency f. The shape of distributions g. Application in R: Examining distributions	SSDS Ch. 3
4. MEASURES OF VARIABILITY	Jan 27	*QUIZ 1* a. Why Focus on Variability b. Index of Qualitative Variation c. The Range and IQR d. Box Plot e. Variance and Standard Deviation f. Choosing a Measure of Variation g. Application with Data: Measures of variability	SSDS Ch. 4
5. THE NORMAL DISTRIBUTION	Feb 3	a. Properties of the Normal Distribution b. Application of the Normal Curve c. The Standard Normal Distribution d. The Standard Normal Table e. Application with Data	SSDS Ch. 5
6. SAMPING AND SAMPLING DISTRIBUTIONS	Feb 10	a. Aims of Sampling b. Basic Probability Principles c. Probability Sampling d. The Concept of the Sampling Distribution e. The Sampling Distribution of the Mean f. The Central Limit Theorem g. Application: Randomness and discovering the CLT	SSDS Ch. 6
7. ESTIMATION	Feb 24	*QUIZ 2* a. Point and Interval Estimation b. Confidence Intervals for Means	SSDS Ch. 7

		c. Confidence Intervals for Proportions d. Application: Testing	
8. TESTING HYPOTHESES	Mar 2	a. Assumptions of Statistical Hypothesis Testing b. Stating the Research and Null Hypotheses c. Probability Values and Alpha d. Five Steps in Hypothesis Testing e. Errors in Hypothesis Testing f. Hypothesis Testing with One Sample and Unknown Population Variance g. Hypothesis Testing with Two Sample Means h. The Sampling Distribution of the Difference Between Means i. Application with Data: Testing Hypotheses	SSDS Ch. 8
9. BIVARIATE TABLES	Mar 9	*QUIZ 3* a. Constructing Bivariate Tables b. Properties of Bivariate Relationships c. Causality and d. Elaboration e. The Limits of Elaboration f. Application: Constructing tables and elaboration in R	SSDS Ch. 9
10. THE CHI-SQUARE TEST AND MEASURES OF ASSOCIATION	Mar 16	a. The Concept of Chi-Square as a Statistical Test b. The concept of Statistical Independence c. The Structure of Hypothesis Testing with Chi-Square d. Proportional Reduction in Error e. Additional Measures of Association f. Application	SSDS Ch. 10
11. ANALYSIS OF VARIANCE	Mar 23	*QUIZ 4* a. Understanding Analysis of Variance b. Hypothesis Testing with ANOVA c. The F-Test d. Application	SSDS Ch. 11
12. REGRESSION AND CORRELATION	Mar 30	a. The Scatter Diagram b. Linear Relationships and Prediction Rules c. Finding the Best-Fitting Line d. Assessing Accuracy of Predictions e. The concept of statistical control f. Introduction of Multivariate Regression Models	SSDS Ch. 12
13. CONCLUSIONS	Apr 6	*Quiz 5* Description of Final Assignment Course Summary and Review	