

Introduction to Applied Social Statistics (SOC 504)

Class Schedule

Lecture: M & W 1:30-3:20PM
Zoom:
URL <https://canvas.uw.edu/>

Professor

Name: Zack W. Almquist
Office:
Office Hours: By appointment
Email: zalmquist@uw.edu

Teaching Assistant

Name:
Office:
Office Hours:
Email:

Course Description

This first course in Sociology's statistics-methods sequence is designed to introduce basic statistical techniques used in the social sciences and prepare students for more advanced statistics courses. Modern social sciences depend heavily on computation and statistical procedures. The content of the course can be divided into three sections: (1) In the first three weeks we will cover the basics of statistics through modern computational social sciences with an emphasis on the R statistical programming language. The focus will be on learning R, data management in R and basic programming. (2) In weeks four to seven we will cover the basics of statistics from a small and large sample perspective with an eye towards the standard data

forms found in the social sciences. (3) In weeks nine to ten will cover the basics of correlation analysis with a focus on the simple linear model.

The course will be taught under the assumption that registered students have little or no statistical background but is also designed to bolster the skills of those with past statistical training.

Course Objectives

The course is designed to:

- provide students with an introduction to the basic concepts, terminology and procedures of data analysis;
- foster general statistical literacy and the ability to critically assess the application and interpretation of core statistical tools;
- increase the ability to calculate, interpret, and explain basic descriptive and inferential statistics;
- introduce the use of a statistical software package (R) to access and analyze data;
- and practice the development of analytic strategies to most effectively test theoretical arguments.

Prerequisites

- Pre-quarter Mathematics and R bootcamp
 - Review basic mathematical notation
 - Review mathematics concepts for social sciences
 - Introduction to R for social sciences
- Covers material equivalent to
 - [CSSS 508 | UW CSSS508](#)
 - [Math Camp | Center for Statistics and the Social Sciences](#)

Text Books

- Standard Free Introduction Textbooks in Statistics
 - [Online Statistics Education: A Free Resource for Introductory Statistics](#)
onlinestatbook.com
 - [HyperStat Online: An Introductory Statistics Textbook and Discussion of whether most published research is false](#)
 - [Concepts and Applications of Inferential Statistics-a free, full-length, and occasionally interactive statistics textbook](#)
 - [The Little Handbook of Statistical Practice](#)
 - [Sportscience](#)
 - [OpenIntro Statistics](#)

- Textbooks in R
 - Golemund, Garrett, and Hadley Wickham. "R for data science." (2018).
 - <https://r4ds.had.co.nz/>

Research Articles

- Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, 69(4), 371-386.

Lecture Notes

- TBD

R Help and activities

- <https://www.datacamp.com/enterprise/soc-504-introduction-to-applied-social-statistics>
- Over the course I and the TA will recommend modules for self-study and help.

Homeworks

There will be 5 homeworks for the quarter. Homeworks is meant to provide experience with both the theoretical and practical side of the course material we will cover in class.

Note 1: Homeworks are meant to achieve three results: (1) provide practice with the computational and statistical programming language R, and (2) provide practice with the statistical concepts discussed in class and (3) provide a chance to demonstrate your mastery of material and highlight areas where more work is needed. You may work in a group, but all write-ups must be done independently. All collaborators should be appropriately cited in your write up and any R code from a given source should also be cited as you would a journal article.

Note 2: Your lowest homework score will be dropped.

Project Proposal

In the second week of class you are expected to submit a research proposal of no more than 2 pages. The proposal should consist of (1) the data set you will use for your final project and (2) a brief description (written) about the data set. Options for the data set will be provided.

Project

The final project will be due on Wednesday of Finals week. The final project will be an analytic report with the following structure

1. Introduction
 - a. Including research problem
2. Data and Methods
3. Results
4. Discussion

Grading

Homeworks: 70%
Project Proposal: 10%
Project: 20%

Letter grade assignment

% Points Earned	Number grade	Letter Grade
100-97	4.0-3.9	A
96-90	3.8-3.5	A-
87-89	3.4-3.2	B+
86-84	3.1-2.9	B
83-80	2.8-2.5	B-
79-77	2.4-2.2	C+
76-74	2.1-1.9	C
73-70	1.8-1.5	C-
69-67	1.4-1.2	D+
66-64	1.1-0.9	D
63-60	0.8-0.7	D-
59-0	0	F

Course Schedule

Note that the dates listed below are subject to change. Changes will be announced in class.

Boot Camp		Data Examples	Handout	Weekly Project Activity	Weekly R Lab
1 Day	Introduction to R	Data camp and Chuck's lecture			
1 Day	Mathematics Review	Algebra, Basic Calculus, Vectors, Matrices			
Week 0					
Lecture 1	--		Vocabulary for Univariate Statistics		
Lecture 2	Introductions, Overview, Set expectations	GSS, ANES, ACS, UN Country data (alr3 package)	Mathematical equations for Univariate Statistics	Pick Dataset for Project - Write up introduction to the data and research question	Install GSS package
Week 1					
Lecture 1	Univariate Statistics - Graphical Representation of Distributions	ACS	R Code for visualizing Univariate Distributions	Print Dataset for project using R - Write up description	Histogram, Scatter Plot, Boxplot
Lecture 2					
Week 2					
Lecture 1	Univariate Statistics - Numerical Representation of Distributions	ACS	R code for mean, variance, standard deviation	Plot two columns of data - Write up discussion of these plots	mean, variance, standard deviation
Lecture 2					
Week 3					

Lecture 1	Univariate Statistics - Summarizing distributions: Percentiles, Standardized Scores	GSS	R code for percentile ranks from normal distribution and z-scores	Tabulate categorical grouping for project using R - Write up what you learned from computing these marginals	Calculate z-score; calculate percentile ran; introduce r code for normal distribution
Lecture 2					
Week 4					
Lecture 1	Discrete Random Variables	GSS	R code for binomial distribution, poisson distribution; Mathematical and vocabulary handout for discrete random variables	Compute Z-scores for two columns of data - plot histogram. Write up discussion.	R code for binomial distribution, poisson distribution
Lecture 2	Continuous Random Variables	ACS	R code for uniform distribution, normal distribution, Chi-square distribution, t-distribution, F-distribution; Mathematical and vocabulary handout for continuous random variables	Compute percentiles for two columns of data - Write up discussion of what the percentile ranks tell you about the data.	R code for uniform distribution, normal distribution, Chi-square distribution, t-distribution, F-distribution
Week 5					
Lecture 1	Sampling Distributions	ACS, GSS	R code for sample(); r code for monte-carlo bootstrap; Mathematics and vocabulary handout for sampling distribution	Write up a hypothesis for your data - plain language, should relate to your research question from week 0 and be testable with the data you	Simulate Sample Distribution: R code for sample function; R code for monte carlo bootstrap

				have.	
Lecture 2	Confidence Intervals	ACS, GSS, UCI Machine Learning Repository	R code for percentile CI from mc bootstrap and normal approximation for CI; Mathematics and vocabulary handout for confidence intervals		R code for Confidence Interval
Week 6					
Lecture 1	Classic Hypothesis Testing	ACS, GSS, UCI Machine Learning Repository	Mathematics and vocabulary handout for classic hypothesis testing	Estimate Sampling Distribution for two columns of data - Visualize the mean distribution of two variables of interest. Discuss what you learned from this exercise.	R code for t-test; proportion test; Chi-square test
Lecture 2	Classic Hypothesis Testing: R		R code for t-test; proportion test; Chi-square test; calculating p-value for all tests		
Week 7					
Lecture 1	Review	ACS, GSS, UCI Machine Learning Repository	Review handout of all R code used so far; Review of all mathematics and vocabulary handout	Rewrite hypothesis in Statistical Language -- Make your hypothesis testable with the tools just introduced.	R Code Review
Lecture 2	Bivariate Statistics - ANOVA Categorical		R Code for Risk, relative risk, odds-ratio; Mathematics and vocabulary handout bivariate statistics		R Code for Risk, relative risk and odds ratio
Week 8					

Lecture 1	Bivariate Statistics - ANOVA Categorical		R Code for Chi-square test	Test Hypothesis - Perform hypothesis and write up what the outcome is -- discuss what you think is going on with your finding.	R Code Categorical ANOVA and Correlation
Lecture 2	Bivariate Statistics - Correlation		R Code for Kendall Rank Test; Spearman Correlation Test		
Week 9					
Lecture 1	Bivariate Statistics - SLR		R Code - lm, plot lm, anova	Run SLR on data - Run SLR, print out the output; write up discussion.	R Code for SLR and model fit diagnostics
Lecture 2					
Week 10					
Lecture 1	Data Analysis Case Study using SLR	Zillow housing data	R Code example for data analysis	Using SLR in week 9 - perform model adequacy assessment; write up discussion of how reliable you think your model is.	R Code Data Analysis Exercise
Lecture 2					
Week 10					
Finals				Project due at end of week. Take project activities and synthesize them into a single write up.	

Datacamp

This class is supported by [DataCamp](#), the most intuitive learning platform for data science and analytics. Learn any time, anywhere and become an expert in R, Python, SQL, and more. DataCamp's learn-by-doing methodology combines short expert videos and hands-on-the-keyboard exercises to help learners retain knowledge. DataCamp offers 325+ courses by expert instructors on topics such as importing data, data visualization, and machine learning. They're constantly expanding their curriculum to keep up with the latest technology trends and to provide the best learning experience for all skill levels. Join over 5 million learners around the world and close your skills gap.

Vic's Datacamp Recommendations

- [Programming - Part 1 \(Writing code in RStudio\)](#)
- [Introduction to R](#)
- [Intermediate R](#)
- [Introduction to the Tidyverse](#)
- [Reporting with R Markdown](#)
- [Managing - Part 1 \(Projects in RStudio\)](#)
- [Introduction to Writing Functions in R](#)
- [Introduction to Statistics in R](#) (Chapters 1-3)
- [Foundations of Probability in R](#) (Chapters 1-2)

Chuck's R Introduction to R for Social Scientists

- [CSSS 508 | UW CSSS508](#)

Chris Adolph's Visualization Course

- [Chris Adolph :: Visual](#)