
Rutgers University
Department of Sociology
Sociology 541
Analysis of Sociological Data I
Spring 2020

Professor: Quan D. Mai
Office: 049 Davison Hall
Email: quan.mai@sociology.rutgers.edu
Office Hours: Monday 12:30 PM - 1:30PM or by appointment

Teaching Assistant: Tsai-Yen Han
Email: tyhan@sociology.rutgers.edu
Office Hours: **Wednesday 9:00AM – 11:00AM**

Class Meets: Mondays 9:30 AM -12:10 PM

Course Description

This introductory course is the first part of a two-semester sequence (541 and 542). It is geared toward graduate students who are training to become professional sociologists. The course introduces students to basic concepts and techniques in quantitative analysis and illustrates how these core statistical ideas can be applied to social science questions and real world problems. The course covers multiple topics such as descriptive statistics, probability, statistical inference, significance test, and ordinary least square regression. Content covered in SOC 541 serves as the foundation for the second course in the sequence. This course assumes that students have little or no statistical background.

Learning Goals

1. Introduce students to basic concepts, terminology, and process of data analysis.
2. Teach students to intelligently read published quantitative research and creatively design their own quantitative projects
3. Acquire statistical literacy and be able to determine when, why, and how statistical tests are used.
4. Familiarize students with a statistical software package to perform analyses of quantitative data.
5. Foster the ability to critically engage with media or scientific report of studies based on quantitative data.

Required Texts

The following two books will be used throughout the semester.

1. Agresti, Alan and Barbara Finlay. 2018. *Statistical Methods for the Social Sciences: Fifth edition*. Pearson.
2. Acock, Alan C. 2018. *A Gentle Introduction to Stata: Sixth Edition*. College Station, TX: Stata Press.

Other References

Devore, J. and R. Peck. Statistics: The Exploration and Analysis of Data.

Freedman, D., R. Pisani, and R. Purves. Statistics. (Less mathematical presentation)

Tanur, J. Statistics: A Guide to the Unknown. (Full of examples and a useful source for projects)

Wonnacott, T. and R. Wonnacott. Introductory Statistics. (More mathematical presentation)

Computing

There are several powerful statistical packages. For the purpose of this class, we will use STATA. If you are interested in other packages such as R, feel free to let me know. The TA, is your go-to person for all STATA-related requests. Other helpful links can be found below

<https://data.princeton.edu/stata/>

https://dss.princeton.edu/online_help/stats_packages/stata/

https://www.cpc.unc.edu/research/tools/data_analysis/statatutorial

<https://stats.idre.ucla.edu/>

Diversity Statement

The Rutgers Sociology Department strives to create an environment that supports and affirms diversity in all manifestations, including race, ethnicity, gender, sexual orientation, religion, age, social class, disability status, 3 region/country of origin, and political orientation. We also celebrate diversity of theoretical and methodological perspectives among our faculty and students and seek to create an atmosphere of respect and mutual dialogue. We have zero tolerance for violations of these principles and have instituted clear and respectful procedures for responding to such grievances.

Course Requirements

1. **Problem Sets:** Complete 10 homework or problem sets. Problem sets will include [1] applying statistical theory that lectures cover and [2] producing and interpreting statistical softwares' outputs. Problem sets are due in class one week after they are assigned, and constitute **25%** of your final grade.
2. **Class presentations:** Students are expected to participate fully in this class. We will typically begin each meeting by reviewing homework assignments and I will ask students to present solutions to assigned problems at that time. In addition, we will often break up into small groups to review certain statistical concepts, and I will ask a representative from the groups to present solutions and conclusions to the class. Your participation in these presentations constitutes **15%** of your final grade.
3. **Mid-Term Examination:** There will be a midterm exam addressing concepts covered in the first half of the course. The exam makes up **25%** of your final grade
4. **Final Oral/Written Report:** You will be required to acquire your own secondary data set and apply the techniques learned in this course to analyze the data. You have multiple options with regard to the data set you choose to use. I will make available three data sets that you can analyze: [1] the 2010 General Social Survey (GSS), and [2] 2013 Global Dataset, and the [3] 2010 European Working Conditions Survey (EWCS). The data and codebooks for these datasets are available on the course's website. These data sets

contain a series of variables and you can construct a smaller data set from this information, focusing on a topic that is of greatest interest to you. If you prefer, you may analyze another data set with which you are familiar. The TA and I can help you with obtaining data from other sources such as ICPSR.

I ask that you each turn in a one-page proposal of your topic (see last page of syllabus for details), due on **March 2**. The last class will be devoted to oral reports of these findings. Each class member will be allotted approximately ten minutes toward this end. A short paper (10 pages) detailing the analysis will also be required. This report and paper account for **35%** of your grade.

→ Summary:

1. 10 Problem Sets: 25%
2. Class Presentations/Participations: 15%
3. Mid-Term Exam: 25%
4. Final Oral and Written Report: 35%

Tentative Course Schedule¹

Date	Week		Content
27-Jan	1	Introduction	Lecture 1 Agresti Chapter 1; Acock Chapter 1 and 4
			Introduction to Statistics, Syllabus Walkthrough. Samples and Populations
3-Feb	2	LAB 1	<ol style="list-style-type: none"> 1. Introduction to Stata 2. Creating a do-file, copying your results to a word processor 3. Check out Stata learning modules (especially Fundamentals of Using Stata) both before and after the lab session at: https://stats.idre.ucla.edu/stata/
10-Feb	3	Sampling and Measurement	Problem Set 1 due
			Lecture 2 Agresti Chapter 2
			Variables and measurements, randomization, sampling
			Lab 2 Acock Chapter 3
			Creating and modifying variables, reverse-code variables
17-Feb	4	Descriptive Statistics	Problem Set 2 due
			Lecture 3 Agresti Chapter 3
			Description by tables and graphs, variability, measures of position, bivariate, population parameters
			Lab 3 Acock Chapter 5.1-5.3
			Distribution centers, distribution dispersion

¹ This schedule is subject to change at the instructor's discretion.

24-Feb	5	Probability	Problem Set 3 due
			Lecture 4 Agresti Chapter 4
			Probability distribution, normal probability, sampling distribution
			Lab 4 Acock Chapter 5.4-5.6
			Graphs for nominal, ordinal, and numeric variables
2-Mar	6	Review Session	Problem set 4 due Review Session
9-Mar	7	MIDTERM	
16-Mar	8	SPRING BREAK	
23-Mar	9	Statistical Inference: Estimation	Lecture 5 Agresti Chapter 5
			Point and interval estimates, confidence intervals, sample size
			Lab 5 Acock Chapter 2
			Creating and Checking a dataset using the Data Editor and Variables Manager
30-Mar	10	Significance Test	Problem Set 5 due
			Lecture 6 Agresti Chapter 6-7
			Theories, hypotheses, test statistics, p-values, type I and type II errors, independent and dependent t-tests
			Lab 6 Acock Chapter 7
			One-sample test of means, two-sample test of group means, one-sample and two-sample test of a proportion, unequal variance, repeated measures tests
6-Apr	11	Bivariate Analysis	Problem Set 6 due
			Lecture 7 Agresti Chapter 8-9
			Contingency tables, expected frequencies, correlations
			Lab 7 Acock Chapter 6

			Cross-tabs, chi-squared tests, degrees of freedom, associations
13-Apr	12	Linear Regression	Problem Set 7 due
			Lecture 8 Agresti Chapter 10-11
			Regression, sum of squares, mean square errors, standard deviations
			Lab 8 Acock Chapter 8
			Scattergrams, regression lines, Spearman's Rho
20-Apr	13	Analysis of Variance	Problem Set 8 due
			Lecture 9 Agresti Chapter 12
			Regression with dummy variables, ANOVA, two-way ANOVA.
			Lab 9 Acock Chapter 9
			One way and two-way ANOVA, repeated measures
27-Apr	14	Multiple Regression	Problem Set 9 due
			Lecture 10 Agresti Chapter 13-14
			Association and causality, control for another variable, spuriousness, alternative explanations, interaction, dummy variables, F-tests
			Lab 10 Acock Chapter 10
			Multiple regression, regression diagnostics, categorical predictors, interactions
4-May	15	Presenta-tion	

Advice for Succeeding in the course

Advice given to students (Inspired by the A. Rossman's Instructor Notes to *Workshop Statistics*):

1. Ask Questions!
2. Use all your resources (the book, the labs, the lectures, class discussion, office hours).
3. Continually review and integrate the material throughout the semester - don't fall behind or get overconfident. Don't leave early topics behind, but build on them.
4. Work together, ask questions of me and other students.
5. Practice the material in new situations.
6. Thoughtfully read the assigned sections before coming to class.
7. Always provide clearly written explanations of your answers.
8. Have fun.
9. Be responsible for your own learning.

Reading is fundamental, re-reading is inevitable, and even then ultimate comprehension is not guaranteed. Other readings that supplement content covered in the textbooks or that clarify a particularly difficult or important concept will be distributed during class or made available via

Sakai. Further, I will put several introductory statistics textbooks for social scientists on reserve in my office, just stop by and borrow one if you feel the need to see material addressed in another way. If you are struggling with the readings, talk to other students in the class, and then set up a time to talk with the teaching assistant and/or the professor.

Short Paper

The final oral and written report makes up 35% of your grade. You are required to write a short paper (no shorter than 10 written double-spaced pages, plus tables/graphs) on a topic of your choosing. I recommend that you use one of the three following datasets:

- [1] the 2010 General Social Survey (GSS)
- [2] 2013 Global Dataset
- [3] 2010 European Working Conditions Survey (EWCS).

The data and codebooks for these datasets are available on the course's website.

If you have access to other data sets that are more relevant to your own research interest, please feel free to use them. The primary goal of the paper is for you to learn the statistical methods, and if you can couple that with advancing your interest in a relevant sociological subject, even better! Feel free to come speak with me if you need help with finding an appropriate data set.

A one-page summary of your topic is due on **March 2, 2020** in class. The summary should appropriately address the following questions:

1. What is your research question? What is the central relationship you are interested in? Identify your outcome variable and the explanatory or independent variables. You want to identify 4-5 variables with which you'll work. You can select one outcome variable and 4-5 independent variables or you can select multiple outcome variables and 2-3 independent variables.
2. What hypotheses can you draw about the relationship? Be sure to link these hypotheses to sociological theory. What guides your thinking and leads to these hypotheses?
3. How will you operationalize your hypotheses? What data set will you use? Which variables will you use from the data set to test your hypotheses?

Please feel free to come see me with any questions about this project. I'd like to see you early on in the semester, so that I can make sure you are on the right track with the assignment.