SOC 271C Methods of Sociological Research

Topic:	Course Description (Syllabus)
Course:	SOC 271C
Course Title:	Methods of Sociological Research
Semester:	Spring 2020
Units/Credits:	3.0
Class #:	
Course convener:	Trond Petersen
Time:	Monday 09:00am-12:00pm
Place:	402 Barrows Hall
Office Hours:	Monday 01:00-02:00pm
	Monday 03:30-04:30pm
Office:	458A Barrows
Phone:	(510) 642-6423
E-MAIL:	Trond.Petersen.Courses@GMail.com

Graduate Student Instructor (GSI):	Robert Pickett
Computer Section:	Tuesday 02:00pm-04:00pm
Room for Section:	402 Barrows Hall
Office Hours:	Wednesday 10:00-11:00am
	Thursday 12:00-01:00pm
Room Office Hours:	483 Barrows Hall
Office:	Demography Department, Attic
E-MAIL:	Robert Pickett <rpickett@berkeley.edu>
More Office Hours:	There will be more office hours by others as needed

Date: January 01, 2020

Description of the course

The focus of the course is on techniques for basic statistical analysis of social science data.

Part I covers analysis of Binary and Multinomial Dependent Variables.

First, I will discuss why linear regression analysis is not always used for analyzing binary dependent variables. An example of such a variable is whether a person is in the labor force or not. This discussion serves to motivate the use of various nonlinear models for dichotomous and other categorical dependent variables.

Second, I will cover methods for analyzing binary and multinomial dependent variables by means of logit and probit models.

Part II covers analysis of event history data

Event Histories: These are data that arise when one records the dates at which units enter and leave social states (e.g., employment, marriage, democracy). The goal of the analysis is to assess the determinants of how long units stay in the state entered.

Part III covers analysis of panel data and data with a nested or a group structure.

Panel Data: One observes several individuals (or other units, such as organizations, nations, etc) at several points in time, say every year. The higher level (or the group) is the individual, that is, the group is all the observations over time on the individual. The lower level or the observations on the group are the observations in each year.

Other data with Group Structure: Observations of several schools and several students within each school. The higher level (or the group) is the school. The lower level or the observations on a group are the observations of each student in the school.

Part IV covers recent advances in Causal Analysis:

Causal Analysis: Introduction to the Counterfactual Framework.

Causal Analysis: Natural Experiments and the use of Instrumental Variables and Regression Discontinuity designs.

Causal Analysis: Observational Data and the use of Matching methods relying on Propensity Scores and Weighting of these.

Part IV New Opportunities in Quantitative Data Analysis: Big Data, Machine Learning, Prediction.

This part covers some new opportunities in quantitative data analysis, techniques which require knowledge of what has been taught up to this point.

The course relies heavily on examples. We will use computer outputs to illustrate the techniques covered by the readings. There will also be some computer exercises in class, where we will formulate models that we are to estimate. Then we will estimate these, and finally interpret the estimates produced.

The primary objective of the course is to teach the participants how to apply these methods. In order to do so, some theory must be covered. But, this is not a course in the theoretical aspects of these types of models.

Course prerequisites

You should have a firm grounding in linear regression analysis, with knowledge of how to deal with categorical independent variables and how to deal with interactions.

Format

The course will rely heavily on lecture notes and class handout prepared by the course convener. The classes themselves might be compact, but the materials will then be considered in a fair amount of detail in the lecture notes. The lecture notes will be made available prior to the course. The lecture notes will also contain various technical materials not gone through in class.

Goals of Course

The class has three goals.

The first is to teach you four important methods heavily used in sociology that go beyond linear regression analysis. The goal here is that eventually you will be able to read quantitative articles published in professional journals.

The second goal is to teach you some ability to use statistical software for doing quantitative research. To this end, we use the software package STATA, a popular package for undergraduate and graduate classes and popular among quantitative researchers.

A third goal is to teach you how to access publicly available datasets, such as the GSS, Current Population Survey, PUMS. You will also use some of these in class.

Course requirements

These are the requirements for the course:

(1) Do homework assignments, theoretical and computer related
(2) Complete 2 in-class tests (during lab) and one take-home exam
(3) Write a final paper using the methods, 10-15 pages

Course Grading

Your grade will consist of six parts:

	In %	In Points
10 Graded homework assignments:	20%	200 Points
Test 1:	15	200 (In-class)
Test 2:	15	200 (Take home)
Test 3:	10	200 (In-class)
Final paper:	40	400
Sum	100%	1,000 Points

The tests are open note and open book. See below.

Test 1 will cover materials on Categorical Dependent Variables
Test 2 will cover materials on Panel Data, Data with Group or Nested Structure

Before each test, you will be given a Trial Test. If you solve this, we will correct it, and give you feedback. You will then be prepared for the actual test. Without solving the Trial Test, you will have difficulty solving the actual test in the time allotted.

The final grade is determined from the total number of points you earn. No intermittent grade is given.

Software for Class

We will use the software package STATA.

The Manual for STATA comes in several volumes. The software has good online help.

A good alternative to the manual is:

Hamilton, Lawrence G. 2012. *Statistics with STATA, Version 12*. Belmont, CA: Duxbury Press. Amazon.com lists the book for $57.06.

Class WEBsite

The course has a WEBsite:

bCourses: Soc 271C

Section handouts and other class documents are also uploaded. Check back regularly for updates.

Here you will find the syllabus, assignments, solutions to assignments (to be posted), and data sets. The WEBsite will be updated weekly.

On the Readings

The textbooks for the class are:

Agresti, Alan and Barbara Finlay. 2009. (4th ed.). *Statistical Methods for the Social Sciences*. Prentice Hall. Only Chaps. 15.1-15.3 are assigned.

or

Agresti, Alan and Barbara Finlay. 2018. (5th ed.). *Statistical Methods for the Social Sciences*. Global Edition. Prentice Hall. Only Chaps. 15.1-15.3 are assigned.

There are minor differences between the two editions. They key one being that the 5th edition lacks chapter 16 included in the 4th edition. The course does not use chapter 16. The 4th edition can be rented for one semester from Amazon.com for $20.49. The 5th edition can be bought from Amazon.com for $43.68.

Morgan, Stephen L., Christopher Winship. 2015 (2nd Ed.). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press. Parts of Chapter 4, and Chapters 5-6.

The course will rely heavily on a set of type-written Lecture Notes.

I will also use a few published articles that use and illustrate the techniques covered in the course. These will be handed out during the course.

Some of the readings are difficult. Do not despair. The lectures will present water-downed versions of the readings, with an occasional baby-proof of technical materials inserted. Mostly, I will not follow the notation in the readings.

Note that in the detailed outline below I have indicated precisely which pages to read and not to read. Most of the materials in the assigned readings are fairly useless from the viewpoint of the practicing social scientist. I explicitly ask you therefore to focus on those page selections indicated in the detailed outline.

The Reader

All the course materials will be posted on the class WEBsite, as PDF files. You may download them from there.

The “reader” can be downloaded from the WEBsite.

There is one “reader” for the course, containing the following eleven parts:

I. Lecture Notes I: Categorical Dependent Variables
II. Lecture Notes II: Panel Data Analysis
III. Lecture Notes III: Event History Analysis
IV. Lecture Notes IV: Causal Analysis
V. Statistical Tables for z- and t-tests, binomial distribution, F-Table
VI. Homework assignments
VII. Instructions for Final Paper
VIII. Instructions for Tests
IX. Description of data (numbered on Website as VIIII)
X. Description and Manuals for Computer Software
XI. Various readings

Weekly Schedule

Week 0:	Tue 01/14	Spring Semester Begins
Week 1:	Mon 01/20	Holiday MLK Day
	Tue 01/21	Instruction Begins
Week 2:	Mon 01/27	First Day of Classes Categorical Dependent Variables Calculus: Exponents, Logarithms, Odds, Odds-Ratios
Week 3:	Mon 02/03	Categorical Dependent Variables
Week 4:	Mon 02/10	Categorical Dependent Variables
Week 5:	Mon 02/17	Holiday Presidents' Day
	Tue 02/18	Event History Data (Note: Lecture Not Lab)
	Fri 02/21	Due Date for Trial Test 1
Week 6:	Mon 02/24	Lab (Note: Not Lecture, but Lab)
	Tue 02/25	In Class Test 1 During Lab
Week 7:	Mon 03/02	Panel Data: Data Structures and Models
Week 8:	Mon 03/09	Panel Data: Fixed Effects Models, Selection Versus Treatment Effects
Week 9:	Mon 03/16	Panel Data: Random Effects Models
	Fri 03/20	Due Date for Trial Test 2
S Break	Mon 03/23 to Fri 03/27	
Week 10:	Mon 03/30	Causal Analysis: Counterfactual Framework
	Fri 04/03	Due Date for Test 2 (Take Home)
Week 11:	Mon 04/06	Causal Analysis: Natural Experiments, Instrumental Variables
Week 12:	Mon 04/13	Causal Analysis: Natural Experiments, Regression Discontin.
Week 13:	Mon 04/20	Causal Analysis: Matching, Propensity Scores, Weighting
Week 14:	Mon 04/27	Last Day of Class New Opportunities: Big Data, Machine Learning, Prediction
	Fri 05/01	Due Date for Trial Test 3
	Fri 05/01	Formal Classes End
	Sat 05/02	Year-end Party for First-year PhD Students
Week 15:	Mon 05/04	Reading/Review/Recitation Week
	Mon 05/04	Test 3 (In Class)
	Fri 05/08	Last Day of Instruction
Week 16:	Mon 05/11	Final Examination
	Fri 05/15	Final Paper Due at Hours 14:00
	Fri 05/15	Spring Semester Ends
	Sat 05/16	Commencement

Summary on Number of Hours of Instruction:

There are 12 Mondays available for Lectures, 24-36 hours, depending on the number of hours I lecture each Monday, with hours 9-12 available. But I will not teach more than 28 hours, students are busy with other classes.

For the 15 weeks of classes, there will be 3 weeks when the class does not meet, resulting in the class meeting 12 times. If the class meets 3 hours per week, there will be 36 hours of instruction.

A graduate class of 3.0 units/credits should have 28 hours of instruction plus some hours during RRR Week. To reach the goal of 28 hours of instruction, the class will have 9 weeks where it meets 2 hours a week (for 18 hours of instruction) and another 3 weeks where it meets 3 hours a week (for 9 hours of instruction, in Weeks 10-12), resulting in

- 27 hours of in-class instruction in Weeks 1-14

I may add one hour of meetings in one of the weeks, if need arise. That will bring the hours instruction to 28.

Agenda:

Note that the number to the far left indicates the topic covered. They correspond to the chapters in the Lecture Notes and to the Homework Assignments.

Part I: Categorical Dependent Variables

Week 01 No Class, Classes Start on Tue 01/22

Week 02 Binary Dependent Variable: Preliminaries, Probabilities, Odds

M 01/27 Lecture Starts at 10:00
Readings:
Binary Dependent Variable
Readings:
Binary Logit and Probit Models: Probabilities, Odds, and Odds-Ratios
Readings:
Petersen, Lecture Notes on Logit Models, Chap. 1, Chap. 2 (pp. 8-10)
Agresti and Finlay 1997 or 2018, Chaps. 15.1-15.3
Petersen, Class Handout: Probabilities, Odds, and Odds-Ratios
Examples of applications of Logit Models in high-stakes contexts:
Beckett and Hearns 2016, esp. Tables 5-9 (on Death Penalty)
Card 2018 (on Affirmative Action at Harvard University)
Esp. Exhibits 17 (p. 63), 18 (p. 65), 19 (p. 70), 21 (p. 74)

T 01/28 LAB, Computing Odds, Estimating Logit Models etc

Week 03 Binary Dependent Variables

M 02/03 Lecture Starts at 10:00
Binary Logit and Probit Models: Probabilities, Odds, and Odds-Ratios
Readings:
Petersen, Lecture Notes on Logit Models, Chap. 2
Petersen 1985
Agresti and Finlay 1997 or 2018, Chaps. 15.1-15.3

T 02/04: LAB: Logit Model, Exponents
Estimation of Logit and Probit Models

Week 04 Binary Dependent Variables

M 02/10 Lecture Starts at 10:00
Binary Logit and Probit Models: Probit, MNL
Readings:
Petersen, Lecture Notes on Logit Models etc, Chap. 2
Petersen, Lecture Notes on Logit Models etc, Chap. 3
Petersen, Saporta, and Seidel 2005 (footnotes 11, 12, 13)
Other Methods: Simultaneous Equations Models

T 02/11 LAB: Estimation of Logit and Probit Models

Part II Event History Data

Week 05 Event History Analysis

M 02/17 Presidents' Day, No Class

T 02/18 Class Starts at 02:00pm
Discrete-time Event History Analysis
Readings:
Petersen 1995a, Sections 1-3

Allison 1982, pp. 61-66, 70-81
F 02/21 Due Date for Trial Test 1

Week 06 **Event History Analysis**
M 02/24 Class Starts at 10:00
LAB: Event History Analysis
T 02/25 Lab: In class Test 1

Part III **Panel Data and Nested Data Analysis**

Week 07 **Panel Data: Data Structures and Models**
M 03/02 Class Starts at 10:00
Readings:
Andress et al 2013 (chap. 3)
Petersen 2004 (pp. 331-336)
T 03/03 LAB: Panel Data

Week 08 **Panel Data: Fixed Effects Models, Selection Versus Treatment**
M 03/09 Class Starts at 10:00
Readings:
Petersen 2004 (pp. 336-340)
Andress et al 2013 (chap. 4.1)
Petersen, Penner, Høgsnes 2014 (Table 2, pp. 1454-)
Petersen, Penner, Høgsnes 2011 (Table 3, p. 297, Table 5, p. 299)
T 03/10 LAB: Estimation with Fixed Effects

Week 09 **Panel Data: Random Effects Models**
M 03/16 Class Starts at 10:00
Readings:
Petersen 2004 (pp. 340-343)
Andress et al 2013 (chap. 4.1)
Other Methods: Comments on Multilevel Methods
T 03/17 LAB: Estimation with Random Effects
F 03/20 Due Date for Test 2

Week **Mon 03/23 to Fri 03/27 Spring Break**

Part IV **Causal Analysis**

Week 10 **Causal Analysis: Counterfactual Framework**
M 03/30 Class Starts at 09:00
Guest Lecturer: David Harding
Readings:
Morgan and Winship 2015 [Chap. 2 (Sects. 2.1-2.7), Chap. 3 (Sects. 3.1-3.3), Chap. 4 (sects. 4.1-4.2, 4.4)]
T 03/31 LAB: Casual Analysis: Counterfactual Framework
F 04/03 Due Date for Test 3

Week 11 **Causal Analysis: Natural Experiments and Instrumental Variables**
M 04/06 Class Starts at 09:00
Guest Lecturer: David Harding

Readings:
Morgan and Winship 2015 [Chap. 9 (Sects. 9.1-9.3.2, 9.4)]
T 04/07 Lab: Estimation with Instrumental Variables

Week 12 Causal Analysis: Natural Experiments and Regression Discontinuity
M 04/13 Class Starts at 09:00
Guest Lecturer: David Harding
Lee and Lemieux 2010 (Sects. 1-2, 3.0-3.4.1, 4.0, 4.1, 4.6)
Kirk 2009
T 04/14 Lab: Estimation with Regression Discontinuity

Week 13 Causal Analysis: Matching with Propensity Scores and Their Weighting
M 04/20 Class Starts at 10:00
Morgan and Winship 2015 (Chap. 5-6)
Morgan 2001
T 04/21 Lab: Estimation with Matching

Part V New Opportunities: Big Data, Machine Learning, Prediction

Week 14 New Opportunities: Big Data, Machine Learning, Prediction
M 04/27 Class Starts at 10:00
Readings:
Molina and Garip 2018, Gives an Overview
Ayres 2007 (Introduction), The Use of Data for Decisions
Agrawal et al 2018 (Chap. 4), Machine Learning and Prediction
Applications:
Stephens-Davidowitz 2017 (Chap. 4), Use of Big Data
Kleinberg et al 2018, Machines "modelling" of data with focus on Machines versus Humans in Decision Making
Applications of Machine Learning to Images and Text:
Wang and Kosinsky 2018, Analysis of Images
Fligstein et al 2017, Analysis of Text

For an interesting project at Princeton University using machine Learning to study poor families, see
https://www.fragilefamilieschallenge.org/#publishing

T 04/28 Lab

F 04/30 Due Date for Trial Test 3

S 05/02 Year-end Party for First-year PhD Students, Starts at 18:00
Venue: TBA

Week 15: Reading/Review/Recitation Week, No Formal Classes
M 05/04: Class as Needed
T 05/05 Lab: In class Test 3

Week 16: Final Exam Week, Mon Dec 10 through Fri Dec 14

F 05/15 14:00 Final Paper Due
F 05/15 Fall Semester Ends

Detailed Readings Lists for Each Part of Course

Part I Categorical Dependent Variables

Agresti, Alan and Barbara Finlay. 2009 or 2018. (4^{th} or 5th ed.). *Statistical Methods for the Social Sciences*. Prentice Hall.
Note: Chapter references are the 4^{th} Ed. Most of the chapter numbers in the 5^{th} Ed. are the same.

Beckett, Katherine, and Heather Evans. 2016. Race, Death, and Justice. Sentencing in Washington State 1981-2014. Columbia Journal of Race and Law 6(2): 77-114.

Card, David. 2017. Report of David Card. December 17, 2017. For Court Case on Harvard University undergraduate admissions. 198 pages.

Petersen, Trond. 1985. "A Comment on Presenting Results From Logit and Probit Models." *American Sociological Review* 50(1): 130-131 (February 1985).

Petersen, Trond, Ishak Saporta, Marc-David Seidel. 2005. "Getting Hired: Gender Differences." *Industrial Relations* 44(3): 416-443 (July). Read footnotes 11, 12, 13.

Petersen, Trond. No Date. *Lecture Notes on Analyzing Categorical Dependent Variables*. Chapters 1-3.

Additional Readings:

Kuha, J., & Mills, C. 2018. "On Group Comparisons with Logistic Regression Models." *Sociological Methods and Research*

Part II Event History Data

Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories." *Sociological Methodology* 13: 61-98.

Petersen, Trond. 1995a. "The Statistical Analysis of Failure Time Processes." Pp. 453-517 in *A Handbook of Statistical Modelling for the Social and Behavioral Sciences*. Edited by G. Arminger, C. C. Clogg, and M. E. Sobel. New York: Plenum Press.

Additional Readings:

Allison, Paul D. 1984. *Event History Analysis. Regression Models for Longitudinal Event Data*. Bevelerly Hills, CA: Sage.

Part III Panel Data and Nested Data

Andress, Hans-Jurgen, Katrin Golsch, and Alexander W. Schmidt. 2013. *Applied Panel Data Analysis for Economic and Social Surveys*. New York: Springer Verlag.

Petersen, Trond. 2004. "Analyzing Panel Data: Fixed- and Random-Effects Models." Chapter 15 (pp. 331-345) in Alan Bryman and Melissa Hardy (Eds.), *Handbook of Data Analysis*. London: Sage Publications Inc. 2004.

Petersen, Trond, Andrew Penner, and Geir Høgsnes. 2011. "The Male Marital Premium: Sorting versus Differential Pay." *Industrial and Labor Relations Review* 64(2): 283-304.

Petersen, Trond, Andrew Penner, and Geir Høgsnes. 2014. "From Motherhood Penalties to Fatherhood Premia: The New Challenge for Family Policy. The Case of Norway." *American Journal of Sociology* 119(5): 1434-1472.

Part IV Causal Analysis

Kirk, Davis S. 2009. "Natural Experiment on Residential Change and Recidivism: Lessons from Hurrican Katrina." *American Sociological Review* 74(3): 484-505.

Lee, Davis S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (June): 281-355.

Morgan, Stephen L. 2001. "Counterfactuals, Causal Effect Heterogeneity, and the Catholic School Effect on learning." *Sociology of Education* 74 (October): 341-374.

Morgan, Stephen L., Christopher Winship. 2014 (2nd Ed.). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Additional Readings:

Blundell, R., & Dias, M. C. 2009. "Alternative Approaches to Evaluation in Empirical Microeconomics." *Journal of Human Resources 44*(3): 565-640.

Part V New Opportunities: Big Data, Machine Learning, Prediction

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018. *Prediction Machines. The Simple Economics of Artificial Intelligence*. Boston, Mass.: Harvard Business Review Press.

Ayres, Ian. 2007. *Super Crunchers. Why Thinking-by-Numbers Is the New Way to Be Smart*. New York: Bantam Books.

Fligstein, Neil, Jonah Stuart Brundage, and Michael Schultz. 2017. "Seeing like the Fed: Culture, cognition, and framing in the failure to anticipate the financial crisis of 2008." *American Sociological Review* 82(5): 879-909.

Kleinberg, Jon, Himabindu Lakkaraju, Jre Leskovec, Jens Ludwig, Sendhil Mullainathan. 2018. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics* 133(1): 237-293 (February).

Molina, Mario, and Filiz Garip. 2018. "Machine Learning for Sociology." Annual Review of Sociology 45: 27-45.

Stephens-Davidowitz, Seth. 2017. *Everybody Lies. Big Data, New Data, and What the Internet Can Tell Us about Who We Really Are*. New York: Harper Collins.

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology* 114*(2)*: 246–257 (February). https://doi.org/10.1037/pspa0000098

Note: This article also generated several newspaper write ups as well as controversy. Here are two of the critiques:

https://afog.berkeley.edu/2018/03/21/how-might-the-history-of-ai-help-us-think-about-and-critique-wang-and-kosinskys-gaydar-study/

https://www.theregister.co.uk/2019/03/05/ai_gaydar/

(FILE: M: Soc 271C Categorical Dependent Variables, Panel Data, and Causal Analysis Syllabus Spring 2020.doc)