RSOC 573: Methods of Survey Data Analysis

Penn State University - Spring 2015

Monday 9:30 AM – 12:30 PM, 004 Life Sciences Bldg.

Professor: Shannon Monnat
Email: smm67@psu.edu
Office: Armsby 110-B
Office Hours: by appointment

Teaching Assistant: Raeven Chandler
Email: rfc134@psu.edu

Course Overview and Objectives

This graduate level course is intended to provide you with familiarity with developing and interpreting multivariate regression analyses from survey data. We will review basic descriptive and inferential statistics and then cover ordinary least squares (OLS) regression models for continuous outcomes, the generalized linear model and maximum likelihood estimation for dichotomous, multi-category nominal, ordinal, and event count outcomes, and end with an introduction to multilevel modeling for spatially nested and longitudinal data. The course will also cover where to locate and how to access secondary survey data, preparing data for analysis, the development of indices from multiple survey items, principals of model building, model diagnostics, and dealing with complex survey sample designs. You will become familiar with SAS, one of the most widely used statistical software programs in the social sciences. Because the course is held in a lab, it will be practical and applied in its discussion of various statistical techniques. You will have the opportunity to analyze real survey data, and the course will be focused on selecting proper methodological techniques, accurate interpretation, and clear presentation of results rather than statistical theory. By the end of the course, you should be able to describe the most common techniques currently used for survey data analysis in the social sciences and be able to apply those techniques to your own research.

Required Texts

- Gordon, Rachel A. 2012. *Applied Statistics for the Social and Health Sciences*. New York, NY: Routledge.
- Luke, Douglas A. 2004. *Multilevel Modeling* (Sage Little Green Book).
- Articles posted on Angel

There is no perfect textbook for a course that covers all of the various survey data methods we will cover in this course. However, I selected Rachel Gordon's textbook for five reasons.

1. As indicated by the title, the text is applied, focusing on the basic skills you will need to produce and read quantitatively based research, and it emphasizes analytic questions that will be important for you when you publish your own research (e.g., how do we interpret the size (not just the significance and sign) of coefficients? Why do coefficients change when other variables are introduced to the model, and what do those changes mean? How exactly does a regression model test for an interaction?).
2. While most statistics texts cover either ordinary least squares regression *or* regression for categorical and other limited dependent variables, this text covers both.
3. Though we will use SAS in this course, STATA is also a popular and useful statistical software program. This text provides code and interpretation for both SAS and STATA for the same tasks.

This will not only enable you to see the similarities and differences between the two software languages, but it will also enable you to experiment with STATA and adapt it to your own needs both during and after this course.

4. The text provides excerpts from published social science research, including the published tables, and then Gordon explains how the findings presented in the tables relate to the author(s)' original research question and hypotheses.

5. The examples use real data from the National Survey of Families and Households which allows us to deal with all of the practical issues that arise when you are working with survey data that you did not collect (e.g., understanding the data's design and documentation, including how variables are coded; dealing with missing data; using sample weights; and accounting for complex survey design). Each chapter provides a roadmap for the steps you would take with any secondary survey data. Note that we will use various other data

Course Format

I do not want to talk for 3 hours, and I am betting you do not want to listen to me for that long. Therefore, most courses will be structured so that the first half of class is devoted to the introduction and instruction of new concepts and methods and the application of those methods in SAS and the second half of the class is devoted to practicing those methods through course exercises individually and/or in small groups, with Raeven and I available to assist you if you get stuck. After the first few weeks of the semester, the second half of class will mostly be devoted to you working on the analysis for your research paper that will be due at the end of the semester. This time will allow you to receive assistance with coding, analysis, interpretation, and feedback and suggestions from me and your classmates so that you can produce a publishable quality paper.

Course Expectations

Methods for statistical analyses can be difficult to learn, especially given that you are learning content along with new data sets and software programming. This difficulty is alleviated when you properly invest your time and energy in this course. This includes completing all assigned reading, putting proper thought into your final paper, taking good course notes, and coming to see me if you feel like you are falling behind. Please note the following expectations for this class:
1. Attend all class sessions: I do not take attendance, but part of your final grade is based on participation. More importantly, it is very easy to fall behind on this material. Missing class will negatively affect your grade and your ability to fully take advantage of this class.
2. Arrive on time: Arriving to class late is inconsiderate and distracting to your instructors and classmates. Therefore, plan to arrive early so that you are in your seat and have time to get your computer warmed up by the time the class begins.
3. Cell phones: Let's face it – many of you likely view it as a means of torture to not be able to check your cell phones for three hours. So, be courteous and discrete about it. Put your ringers on silent, do not blatantly text, Facebook, email, etc. while I am lecturing, and do not allow your phone to become a distraction to me or your classmates.
4. You earn your own grade: I do not "give" you a grade. You earn your grade by successfully completing the course requirements. I will provide feedback on exams and your paper proposal. You can make appointments with me to help you with class material and discuss your paper. Simply showing up for class and doing the work does not automatically equate to an 'A'. Before you submit a product, ask yourself: "Is this the best work I can possibly do?" If the answer is no, then do not expect me to award that work with an A.

Grading and Assignments

Class Participation: You are expected to attend all classes. It is extremely difficult to catch up on this type of subject matter, even when missing just one class. You are also expected to participate in class by

answering questions and completing practice examples during the class period. Class attendance and participation are worth 15% of your total grade.

Exams: There will be two take-home exams during the semester. You may use any books, web resources, and class resources you want to complete your exams, but you may not work with classmates, other peers, or faculty members on your exams. All work must be your own. Each exam is worth 20% of your final grade (40% total for both exams).

Final Paper: You are required to complete a publication-quality research manuscript using survey data and the techniques we utilize throughout the course. You are allowed to select any existing data set, but you should be aware that Raeven and I may not have expertise with those data and may not be able to provide you with the same level of feedback and assistance that we can if you use one of the recommended data sets listed below. If you choose to use data that are not listed below, you should meet with me to briefly discuss those data, how you are going to access them, and your previous experiences with those data. Your paper should follow the format of a published research manuscript. I recommend that you find a target journal where you may be interested in submitting your paper, locate the guidelines for authors on the journal's website, and use those guidelines (including the word count) to format your paper.

You must submit a brief paper proposal by **Feb 16**. This proposal should be no more than 2 pages, single spaced, and should indicate your research question, the motivation for this research (i.e., why should we care about the answers to this question?), your hypotheses, the data you will use, the current status of those data (do you have them in your possession in SAS format?), your outcome(s) of interest, your main independent variables, and control variables. The paper proposal is worth 10% of your final grade.

The final paper is due by Monday, May 4 and is worth 35% of your total grade. I will not accept late papers, and you cannot pass this course without submitting a paper, so please plan accordingly.

Recommended Data Sets: Note that you must use survey data for your paper in this course. These are the survey data sets with which I have the most expertise and can provide you with insight about variables, recommended analytic strategies, and data challenges.
- Behavioral Risk Factor Surveillance System (http://www.cdc.gov/brfss/annual_data/annual_data.htm)
- General Social Survey (http://www3.norc.org/GSS+Website/)
- National Survey of Drug Use and Health (https://nsduhweb.rti.org/respweb/homepage.cfm)
- Current Population Survey (http://www.census.gov/cps/)

Class Participation	15%
Exams	40%
Research Paper Proposal	10%
Final Paper	35%
	100%

Grading Scale

A = 100-94	B = 85-83	C = 75-70
A- = 93-90	B- = 82-80	D = 60-69
B+ = 89-86	C+ = 79-76	F = less than 60

Course Schedule and Reading Assignments

Dates	Topics	Reading
Week 1 (Jan 12)	Overview of course Review of statistical concepts Locating secondary survey data	
Week 2 (Jan 19)	NO CLASS - Martin Luther King Day	
Week 3 (Jan 26)	Getting started with SAS Descriptive statistics and graphics	Gordon: Chapters 1-5
Week 4 (Feb 2)	Sample, population, and sampling distributions Bivariate inferential statistics Creating indices from survey items (factor analysis and internal consistency)	Gordon: Chapters 6-7
Week 5 (Feb 9)	Ordinary least squares regression Addressing complex sampling designs Multiple regression	Gordon: Chapters 8-9
Week 6 (Feb 16)	*Multiple regression cont.* Interactions (conditional regression) Nonlinear relationships	Gordon: Chapters 9-12 **Brief proposal due**
Week 7 (Feb 23)	NO CLASS – Work on 1st exam	**Exam 1 due Feb 27**
Week 8 (March 2)	*Multiple regression cont.* Indirect effects and omitted variable bias Confounders, mediators, moderators, suppressors Model diagnostics - outliers, heteroskedasticity, and multicollinearity	Gordon, Chapter 13-14
Week 9 (March 9)	NO CLASS – Spring Break	
Week 10 (March 16)	Introduction to the generalized linear model	Gordon, Chapter 15
Week 11 (March 23)	The generalized linear model for dichotomous outcomes (logit, probit)	Gordon, Chapter 16
Week 12 (March 30)	The generalized linear model for multi-category nominal outcomes (multinomial logit)	Gordon, Chapter 17
Week 13 (April 6)	The generalized linear model for multi-category ordered outcomes (ordered logit)	Gordon, Chapter 17
Week 14 (April 13)	Regression for event count data (poisson and negative binomial)	Gordon, Chapter 18 **Exam 2 due April 17**
Week 15 (April 20)	Introduction to multilevel models for linear and categorical data	Luke (whole book)
Week 16 (April 27)	Introduction to multilevel models for longitudinal data	

EXAM 1 DUE FRIDAY, FEB. 27
EXAM 2 DUE FRIDAY, APRIL 17

University Policies

Statement of Academic Integrity

Academic integrity is the pursuit of scholarly activity free from fraud and deception and is an educational objective of this institution. All Penn State University policies regarding academic integrity apply to this course. Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating of information or citations, facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of another person or work previously used without informing the instructor, or tampering with the academic work of other students. For any material or ideas obtained from other sources, such as the text or things you see on the web, in the library, etc., a source reference must be given. Direct quotes from any source must be identified as such. All exam answers must be your own, and you must not provide any assistance to other students during exams. Any instances of academic dishonesty WILL be pursued under the University regulations concerning academic integrity.

Statement of Nondiscrimination

The Pennsylvania State University is committed to the policy that all persons shall have equal access to programs, facilities, admission, and employment without regard to personal characteristics not related to ability, performance, or qualifications as determined by University policy or by state of federal authorities. The Pennsylvania State University does not discriminate against any person because of age, ancestry, color, disability or handicap, national origin, race, religious creed, sex, sexual orientation, or veteran status.

Direct all inquiries regarding the nondiscrimination policy to:
Affirmative Action Director
The Pennsylvania State University
201 Willard Building
University Park, PA 16802-2801
Telephone: (814) 863-0471

Disability Access Statement:

If you have a documented disability and wish to receive academic accommodations, please contact the campus disability liaison as soon as possible: (name, office, telephone, email). For additional information, check the university web site: http://www.equity.psu.edu/ods/
NOTE: Accommodations require documentation.