LPO 8800: Statistical Methods in Education Research

Sean P. Corcoran

Fall 2021

E-mail: sean.corcoran@vanderbilt.edu
Web: **seanpcorcoran.org**
Office: Payne Hall 205C
Classroom: Payne 007
Office Hours: Wednesdays 9am-10:30pm (or by appt.)
Class Hours: TTh 9:35am-10:50pm
https://calendly.com/sean-p-corcoran
Phone: (615) 322-8021

Course Description

This course is an introduction to the theory, methods, and practice of statistics. It is intended as a foundational prerequisite for graduate students who intend to complete the core quantitative methods sequence in the Leadership & Policy Studies program (Regression I/II, Causal Inference). Topics will include probability theory, descriptive statistics, population distributions, hypothesis testing and confidence interval estimation, correlation, and regression. While concepts will be introduced with some mathematical rigor, the primary emphasis is the practical application and conceptual understanding of statistics. The course will be taught using the Stata statistical software package and large-sample datasets commonly used in education policy and social science research.

Prerequisites

A prior undergraduate or graduate course in statistics is recommended, though not required. Note that the class begins at a very introductory level but advances quickly. If you have concerns about your prior preparation for this class, please see me.

Books

The following textbook is recommended, not required:

- *Statistical Methods for the Social Sciences, 5th edition* by Alan Agresti, Pearson, 2018.

If you have no experience with Stata, you may benefit from the following simple introductory text (not required). I will provide other resources for learning Stata.

- *A Gentle Introduction to Stata, Revised 6th edition* by Alan C. Acock, Stata Press, 2018.

Other readings listed in the course schedule below will be made available via Github.

Course Structure

The class will meet twice weekly, in person. Class meetings will be a mix of lecture and in-class exercises. In some cases, supplemental material will be provided online.

Please note this is a graduate course designed for students at the doctoral and master's level. I expect that students enrolled in the course are motivated by a desire to learn the course material. Please come to class having reviewed the lecture notes and recommended readings. Attendance in class and participation in in-class exercises is extremely important.

Stata

Stata is the statistical software used in this course. I recommend the most recent release (Version 17), but other recent vintages are fine (e.g., 13-16). I presently use Stata 15. Be aware that small differences exist between versions, and that files created in recent versions of Stata may not open in older versions. Purchasing options are available via the following website: https://www.stata.com/order/new/edu/gradplans/student-pricing/. A 6-month license for Stata/BE can be purchased for $48, but if you intend to use Stata in your own work I recommend purchasing a perpetual license of Stata/SE or the more powerful multi-processor (MP) version. Stata is freely available to you in the Wyatt 132 computer lab and elsewhere on campus. Vanderbilt also makes Stata available to you virtually through VMWare: https://anywherevu.vanderbilt.edu/portal/webclient/index.html.

There are many great resources for learning Stata, including the Acock text noted above. UCLA has some nice online resources for learning Stata (https://stats.idre.ucla.edu/stata/) and the Stata YouTube site is also quite helpful. There are some handy Stata "cheat sheets" posted here: https://geocenter.github.io/StataTraining/portfolio/01_resource/. I will also upload some Stata references to Github.

Course Requirements

Your grade for the course will be based on **12** problem sets (40%), a statistical project (20%), a midterm (20%), and final exam (20%). The problem sets will vary in length and points possible, but each will be weighted equally when calculating your final grade, using the percent correct on each. Since 13 problem sets are assigned, you will be permitted to drop your lowest score.

The (tentative) schedule of problem set assignments is shown in the course schedule below. These are subject to change based on the pace of the course. Please submit your problem set solutions to me via email at sean.corcoran@vanderbilt.edu. Include your last name and problem set number in the filename (e.g., *Corcoran_PS1.pdf*). Late assignments will not be accepted, particularly after problem set solutions have been provided or discussed in class.

Unless otherwise indicated, the file you submit to me should be a log of your Stata session, saved as a text file (with the .txt extension) or—better yet—converted to a PDF. Begin by copying the problem set instructions into the Stata do-file editor. Comment out the questions. Insert after each question the commands you used to respond to that question. The resulting log file will

include the instructions (in the form of comments), your commands, and the output. Edit this file as appropriate, for example by adding interpretations of your output and any other commentary that might be asked for. Graphical output can be submitted separately, preferably as a PDF file. You are encouraged to work together on the problem sets, but all work submitted must be that of the individual student. Duplicate assignments will not be accepted.

Other Important Information

1. **Github**: All materials pertaining to this course, including lecture notes, problem sets, and datasets, will be available on Vanderbilt's Github platform (https://github.com/spcorcor18/LPO-8800). Check in frequently for new materials and announcements. Most materials will be posted in advance of class, but occasional delays are to be expected. The course is stored in what Github calls a "repository". You can "clone" (sync) this repository to your local drive using Github Desktop (https://desktop.github.com/). I recommend this easy approach to staying up to date with all of the course materials.

2. **Classroom etiquette**: Please bring your laptop to class. To help promote a productive learning environment, please devote your time and attention to the class itself. Please do not use Facebook/Instagram, text messaging, email, or other digital distractions while in class. Please silence your cell phone as well.

3. **COVID-19**: Vanderbilt is requiring masks indoors, except where physical distancing (6 feet) is possible. Instructors are not required to wear masks as long as they remain 6 feet away from others. I plan to not wear a mask when lecturing—assuming physical distancing is possible—but will have one on hand to wear as necessary. Quarantine is not required if you are vaccinated, have close contact with someone with COVID-19, and are asymptomatic. However, if you test positive for COVID-19, you will be required to isolate for 10 days. If this occurs, I will work with students who miss class on a case by case basis. As noted, all class materials—including lecture notes—will be available on Github. For the latest on Vanderbilt protocols for COVID-19, see: https://www.vanderbilt.edu/coronavirus/.

4. **Names and pronouns**: If you would like to use a different name or pronouns than those provided through YES, please let me know at any time prior to or during the semester. Information is available through the LGBTQI Life offices about how to change either or both of these in YES.

5. **Academic integrity**: All academic work at Vanderbilt is done under the Honor System. Students are expected to conform to the highest standards of academic integrity in this course. Any attempt to pass off someone else's work as your own is a violation of this standard, and there are many ways this can happen beyond blatant cheating. Full details of the Vanderbilt Honor System may be found here: http://www.vanderbilt.edu/student_handbook/the-honor-system/ If you have any doubts about how the Honor Code applies to your work in this class, please ask me—not another student—for clarification. Uncertainty about application of the Honor Code does not excuse a violation.

6. **Accommodations**: Vanderbilt is committed to equal opportunity for students with disabilities, as am I. If you need course accommodations due to a disability, please con-

tact VU Student Access Services to initiate the process: https://www.vanderbilt.edu/student-access/. After SAS has notified me of relevant accommodations, we will discuss how these accommodations may best be approached in this class, and I will facilitate the accommodations.

7. **Mandatory reporter obligation**: All university faculty and administrators are mandatory reports. What this means is that all faculty, including me, must report allegations of sexual misconduct and intimate partner violence to the Title IX Coordinator. In addition, all faculty are obligated to report any allegations of discrimination to the Title IX Coordinator (615-343-9004).

   I am willing to discuss such incidents with you, but I can only do so in the context of us both understanding my reporting obligations. If you want to talk with someone in confidence, officials in the Student Health Center, the University Counseling Center, and officials in the Office of the Chaplain and Religious Life (when acting as clergy) can all maintain confidentiality. In addition, officials in the Project Safe Center (Crisis Hotline: 615-322-7233) have limited confidentiality, in that they have to report the incidents they are told of, but can do so without providing identifying information about the victim(s).

8. **Mental health and wellness**: If you are experiencing undue personal and/or academic stress during the semester that may be interfering with your ability to perform academically, Vanderbilt's Student Care Network offers a range of services to assist and support you. I am available to speak with you about stresses related to your work in this course, and I can assist you in connecting with the Student Care Network. The Office of Student Care Coordination (OSCC) is the central and first point of contact to help students navigate and connect to appropriate resources on and off-campus, develop a plan of action, and provide ongoing support. You can schedule an appointment with the OSCC at https://www.vanderbilt.edu/carecoordination/ or call 615-343-WELL.

   The Student Care Network also offers drop-in services on campus on a regular basis. You can find a calendar of services at https://www.vanderbilt.edu/studentcarenetwork/satellite-services/

   If you or someone you know needs to speak with a professional counselor immediately, the University Counseling Center offers Crisis Care Counseling during the summer and academic year. Students may come directly to the UCC and be seen by the clinician on call, or may call the UCC at (615) 322-2571 to speak with a clinician. You can find additional information at https://www.vanderbilt.edu//ucc/

Statistical Project Instructions

Overview

The statistical project is an opportunity for you to apply your knowledge of statistics to real-world data, and answer a specific research question. You will analyze your data using Stata, interpret your results, and communicate your findings in writing.

Requirements

For the statistical project, you may choose one of the following options: (1) provide a written report on a topic to be determined by me, or (2) provide a written analysis of a research question of your choice, using appropriate data. I will provide data for the former, and guidance on the latter. PhD students are recommended (though not required) to choose option 2.

Your written analysis should range between 12—15 pages, including tables and graphs. You may wish to include some graphs and tables in the main body of the paper, and relegate others to an appendix (appendices do not count toward the page limit).

Deadlines

You must decide which option you are pursuing (and if 2, a topic) by **October 19**. The project itself will be due on or before the last day of class, **December 9**. Please submit your completed project as a PDF document via email, and use your last name and the words "Statistical Project" as the filename.

Guidelines

Think of your statistical project as communicating a story through data analysis. Like any good story, it should have a beginning, middle, and end, and a "storyline" that is clear, logical, and compelling. Be selective about tables, graphs, and statistics you choose to include. Overwhelming your reader with output and statistical results is a surefire way to lose and bore them. It also makes you look like an amateur. Be sure to reference all figures and tables in the text itself.

You may use any technique learned in class in your analysis, unless otherwise specified (relevant in option 1). One of the objectives of this project is a demonstration that you know which techniques to apply in specific situations. In many cases there is more than one technique that can be used. Choose the one that most clearly communicates the result you are trying to convey to the reader. Show some range in your use of methods, and avoid overt repetition.

If you choose option 1, you must fully address all of the questions I assign. These will be available later in the semester. If you choose option 2, you may craft your own research question as you see fit—but the quantity and rigor of the analysis should match or exceed that required under option 1.

A sample project and "model answer" will be posted on Github, as an example of past work. Additional details on the project will be provided later in the semester.

Class schedule

The schedule is tentative and subject to change.

Lecture 1 (Aug 26): Introduction to concepts of probability and statistics

- Agresti chapters 1-2
- *Problem set 1 assigned Aug 26, due Sept 2*

Lecture 2 (Aug 31, Sep 2): Describing univariate distributions (I)

- Agresti chapter 3
- Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). *Descriptive analysis in education: A guide for researchers*. (NCEE 2017-4023). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Available at: https://ies.ed.gov/ncee/pubs/20174023/pdf/20174023.pdf
- *Problem set 2 assigned Sept 2, due Sept 9*

Lecture 3 (Sept 7, 9): Describing univariate distributions (II)

- Agresti chapter 3
- Loeb et al. (2017)—see Lecture 2
- *Problem set 3 assigned Sept 9, due Sept 16*

Lecture 4 (Sept 14, 16): Probability and probability distributions

- Agresti chapter 4
- *Problem set 4 assigned Sept 16, due Sept 23.*

Lecture 5 (Sept 21, 23): Sampling distributions

- Agresti chapter 4
- *Problem set 5 assigned Sept 23, due Sep 30.*

Lecture 6 (Sep 28, 30): Statistical inference—estimation

- Agresti chapter 5
- Romer, D. (2020). "In Praise of Confidence Intervals." *AEA Papers and Proceedings* 110: 55-60. Available at: https://pubs.aeaweb.org/doi/pdfplus/10.1257/pandp.20201059
- *Problem set 6 assigned Sept 30, due Oct 5.*

Review for midterm exam (Oct 5)

MIDTERM (Oct 7)

NO CLASS (Oct 12, 14)

Lecture 7 (Oct 19, 21): Statistical inference—significance tests

- Agresti chapter 6
- Greenland, S., S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman and D. G. Altman (2016). "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations." *European Journal of Epidemiology* 31(4): 337—350. Available at: https://link.springer.com/article/10.1007/s10654-016-0149-3
- *Problem set 7 assigned Oct 21, due Oct 28.*
- Please notify Prof Corcoran about your intended statistical project topic.

Lecture 8 (Oct 26, 28): Statistical power and effect size

- Agresti chapter 6
- Kraft, M. A. (2020). "Interpreting Effect Sizes of Education Interventions." *Educational Researcher*. Available at: https://journals.sagepub.com/doi/10.3102/0013189X20912798
- Hill, C. J., H. S. Bloom, A. R. Black & M. W. Lipsey (2008). "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2(3): 172—177. Available at: https://srcd.onlinelibrary.wiley.com/doi/abs/10.1111/j.1750-8606.2008.00061.x
- *Problem set 8 assigned Oct 28, due Nov 4*

Lecture 9 (Nov 2, 4): Hypothesis testing—two groups

- Agresti chapter 7
- *Problem set 9 assigned Nov 4, due Nov 11.*

Lecture 10 (Nov 9): Bivariate covariance and correlation

- Agresti chapters 8-9
- NO CLASS Nov 11 (APPAM)
- *Problem set 10 assigned Nov 9, due Nov 18.*

Lecture 11 (Nov 16, 18): Bivariate regression (I)

- Agresti chapter 9
- *Problem set 11 assigned Nov 18, due Dec 2 (after Thanksgiving).*

THANKSGIVING BREAK (Nov 23, 25)

- *No class—enjoy!*

Lecture 12 (Nov 30, Dec 2): Bivariate regression (II)

- Agresti chapter 9
- *Problem set 12 assigned Dec 2, due Dec 9.*

TBD (Dec 7)

- TBD

Review for final exam (Dec 9)

FINAL EXAM (Dec 14)

- Official final exam time is 9:00 am on Tuesday, December 14.

Schedule at a glance

Aug 26	Lecture 1: Introduction to concepts of probability and statistics	PS 1 assigned
Aug 31	Lecture 2: Describing univariate distributions (I)	
Sep 2	Lecture 2: Describing univariate distributions (I)	PS2 assigned
Sep 7	Lecture 3: Describing univariate distributions (II)	
Sep 9	Lecture 3: Describing univariate distributions (II)	PS3 assigned
Sep 14	Lecture 4: Probability and probability distributions	
Sep 16	Lecture 4: Probability and probability distributions	PS4 assigned
Sep 21	Lecture 5: Sampling distributions	
Sep 23	Lecture 5: Sampling distributions	PS5 assigned
Sep 28	Lecture 6: Statistical inference—estimation	
Sep 30	Lecture 6: Statistical inference—estimation	PS6 assigned
Oct 5	Review for midterm	
Oct 7	**Midterm**	
Oct 12	**NO CLASS - Prof. Corcoran traveling**	
Oct 14	**NO CLASS - VU Fall break**	
Oct 19	Lecture 7: Statistical inference—significance tests	
Oct 21	Lecture 7: Statistical inference—significance tests	PS7 assigned
Oct 26	Lecture 8: Statistical power and effect size	
Oct 28	Lecture 8: Statistical power and effect size	PS8 assigned
Nov 2	Lecture 9: Hypothesis testing—two groups	
Nov 4	Lecture 9: Hypothesis testing—two groups	PS9 assigned
Nov 9	Lecture 10: Bivariate covariance and correlation	PS 10 assigned
Nov 11	**NO CLASS - APPAM conference**	
Nov 16	Lecture 11: Bivariate regression (I)	
Nov 18	Lecture 11: Bivariate regression (I)	PS11 assigned
Nov 23	**NO CLASS - Thanksgiving**	
Nov 25	**NO CLASS - Thanksgiving**	
Nov 30	Lecture 12: Bivariate regression (II)	
Dec 2	Lecture 12: Bivariate regression (II)	PS12 assigned
Dec 7	TBD	
Dec 9	**Review for final exam**	
Dec 14	**Final**	