LPO 8800: Statistical Methods in Education Research

Sean P. Corcoran

Fall 2024

E-mail: sean.corcoran@vanderbilt.edu
Office: Wyatt Center 401N
Office Hours: Wednesdays 9-11 am (or by appt.)
https://calendly.com/sean-p-corcoran

Web: **seanpcorcoran.org**
Classroom: Wyatt Center 130
Class Hours: Tues 8:10-11 am
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 Education Policy and Leadership 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 valuable as a reference but 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, 2023.

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

Course Structure

The class will meet once per week, 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 and participation in class are extremely important given our one-per-week schedule.

Stata

Stata is the statistical software used in this course. I recommend buying a short-term or perpetual license. Stata has recently moved to a subscription model called StataNow where users pay an annual fee to license a continually-updated version of Stata (the current version is 18). It is rather expensive, $360 per year for Stata BE (handles mid-size datasets) and $510 for Stata SE (handles large datasets). There are discounted rates for group purchases (2 or more). A perpetual license does not expire but also does not entitle you to upgrades. See the Stata website for details and make sure you get the best price for your intended usage: https://www.stata.com/order/educational-license-options/. I am currently using Stata 18, but other recent vintages are fine for this class (e.g., 14-17). Be aware that small differences exist between versions, and that files created in recent versions of Stata may not open in older versions. If you don't wish to purchase Stata, it is freely available to you in campus computer labs. Vanderbilt also makes Stata available to you virtually through VMWare: https://anywherevu.vanderbilt.edu/portal/webclient/index.html.

There are many excellent resources for learning Stata. See this site for a useful starting point: https://www.stata.com/links/. There are some handy Stata "cheat sheets" posted here: https://geocenter.github.io/StataTraining/portfolio/01_resource/. I will also upload numerous other Stata references and links to Github.

Course Requirements

Your grade for the course will be based on **11** problem sets (45%), a statistical project (15%), 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 12 problem sets are assigned, you will be permitted to drop your lowest score. Unfortunately, due to the advent of AI tools, the exams must be administered in person.

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. (Combine all your PDFs if possible). Please clean up your submitted results so they are readable and look professional. 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.

Based on feedback in previous semesters, I will answer questions about the problem sets in class but not devote much class time to the solutions. Detailed solutions will be posted on Github.

Other Important Information

1. **Github**: All materials pertaining to this course, including lecture notes, problem sets, and datasets, will be available on Github (https://github.com/spcorcor18/LPO-8800). Check in frequently for new material and announcements. Lectures will be posted in advance of class, but occasional delays and revisions 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**: To help promote a productive learning environment, please devote your time and attention to the class itself. Please do not use social media, text messaging, email, or other digital distractions while in class. Please silence your cell phone as well.

3. **Artificial intelligence**: Do not use Chat GPT or other artificial intelligence tools in this course, for any purpose. The goal is for **you** to learn the concepts and coding skills taught in this course. I will not accept submitted assignments that have clearly relied on AI.

4. **Health and safety**: Our mutual commitment to health and safety is vital. Toward that end, all students are expected to adhere to Vanderbilt health and safety protocols. Guidance may be updated throughout the semester.

5. **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.

6. **Honor code**: All work submitted in this course is governed by provisions of the Vanderbilt University Honor Code, found in the student handbook: http://www.vanderbilt.edu/student_handbook/the-honor-system/. Any attempt to pass off someone else's work as your own is a violation of the Honor Code, and there are many ways this can happen beyond blatant cheating. If you have any doubts about how the Honor Code applies to your work in this class, please ask me for clarification. Uncertainty about application of the Honor Code does not excuse a violation.

7. **Classroom accommodations**: Vanderbilt is committed to equal opportunity for students with disabilities. If you need course accommodations due to a disability, please contact 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.

8. **Mandatory reporter obligation**: All university faculty and administrators are mandatory reporters. What this means is that all faculty, including me, must report allegations of sexual misconduct and intimate partner violence to the Title IX Coordinator (615-343-9004). In addition, all faculty are obligated to report any allegations of discrimination. 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 the Office of the Chaplain and Religious Life (when acting as clergy) can 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). The Project Safe Center (https://www.vanderbilt.edu/projectsafe/) serves as the central resource for those impacted by sexual misconduct and intimate partner violence and can assist with navigating all facets of the University's resource and support network and other processes

9. **Mental health and wellness**: If you are experiencing undue stress during the semester that may be interfering with your ability to perform academically, Vanderbilt's Student Care Network offers a range of support services. 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 you navigate and connect to appropriate resources. 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 Urgent Care Counseling. Students should call the UCC at (615) 322-2571 during office hours to speak with an urgent care clinician. You can also reach an on-call counselor after hours or on the weekends by calling (615) 322-2571 and pressing option 2 at any time. 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 1**. The project itself will be due on or before **December 3**. 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.

Sample projects 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 27): Introduction to concepts of probability and statistics

Agresti chapters 1-2

Shafer, Leah. (2016). "When Proficient Isn't Good," Harvard Graduate School of Education. Available at: https://www.gse.harvard.edu/news/uk/15/12/when-proficient-isnt-good

Problem set 1 assigned Aug 27, due Sep 3

Lecture 2 (Sep 3): Describing univariate distributions

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 Sep 3, due Sep 10

Lecture 3 (Sep 10): Probability and probability distributions

Agresti chapter 4

Problem set 3 assigned Sep 10, due Sep 17.

Lecture 4 (Sep 17): Sampling distributions

Agresti chapter 4

Problem set 4 assigned Sep 17, due Sep 24.

Lecture 5 (Sep 24): 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 5 assigned Sep 24, due Oct 1.

Lecture 6 (Oct 1): 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

Imbens, G. W. (2021). "Statistical Significance, p-Values, and the Reporting of Uncertainty." *Journal of Economic Perspectives* 35(3): 157-174. Available at: https://www.aeaweb.org/articles?id=10.1257/jep.35.3.157

Wasserstein, R. L., & Lazar, N. A. (2016). "The ASA Statement on *p*-Values: Context, Process, and Purpose." *The American Statistician*, 70(2), 129–133. Available at: https://doi.org/10.1080/00031305.2016.1154108

Problem set 6 assigned Oct 1, due Oct 15 (after midterm).

Please notify Prof Corcoran about your intended statistical project topic.

MIDTERM EXAM (Oct 8)—in class

Lecture 6 (Oct 15): Statistical inference: significance tests (continued)

Problem set 7 assigned Oct 15, due Oct 22

Lecture 7 (Oct 22): Statistical power and effect size

Agresti chapter 6

Romer (2020) - see Lecture 5

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

Lortie-Forgues, H., Sio, U. N., & Inglis, M. (2021). "How Should Educational Effects Be Communicated to Teachers?" *Educational Researcher* 50(6): 345–354. Available at: https://doi.org/10.3102/0013189X20987856

Matthay, E. C., et al. (2021). "Powering Population Health Research: Considerations for Plausible and Actionable Effect Sizes." *SSM - Population Health* 14: 100789. Available at: https://doi.org/10.1016/j.ssmph.2021.100789

von Hippel, P. "Multiply by 37 (or Divide by 0.027): A Surprisingly Accurate Rule of Thumb for Converting Effect Sizes From Standard Deviations to Percentile Points." *Educational Evaluation and Policy Analysis*. Available at: https://journals.sagepub.com/doi/abs/10.3102/01623737241239677

Problem set 8 assigned Oct 22, due Oct 29

Lecture 8 (Oct 29): Hypothesis testing: two groups

Agresti chapter 7

Problem set 9 assigned Oct 29, due Nov 5.

Lecture 9 (Nov 5): Bivariate covariance and correlation

Agresti chapters 8-9

Problem set 10 assigned Nov 5, due Nov 12.

Lecture 10 (Nov 12): Bivariate regression

Agresti chapter 9

Problem set 11 assigned Nov 12, due Nov 19.

Lecture 11 (Nov 19): Multiple regression: introduction

Agresti chapters 10-11

Problem set 12 assigned Nov 19, due Dec 3 (after Thanksgiving).

THANKSGIVING BREAK (Nov 26)

No class—enjoy!

Lecture 11 (Dec 3): Multiple regression: introduction (cont.)

Agresti chapters 10-11

Review for final exam

Statistical project due

FINAL EXAM (Dec 10)

Schedule at a glance

Aug 27	Lecture 1: Introduction to concepts of probability and statistics	PS1 assigned
Sep 3	Lecture 2: Describing univariate distributions	PS2 assigned
Sep 10	Lecture 3: Probability and probability distributions	PS3 assigned
Sep 17	Lecture 4: Sampling distributions	PS4 assigned
Sep 24	Lecture 5: Statistical inference: estimation	PS5 assigned
Oct 1	Lecture 6: Statistical inference: significance tests (and review for midterm)	PS6 assigned
Oct 8	**Midterm**	
Oct 15	Lecture 6: Statistical inference: significance tests (cont.)	PS7 assigned
Oct 22	Lecture 7: Statistical power and effect size	PS8 assigned
Oct 29	Lecture 8: Hypothesis testing: two groups	PS9 assigned
Nov 5	Lecture 9: Bivariate covariance and correlation	PS10 assigned
Nov 12	Lecture 10: Bivariate regression	PS11 assigned
Nov 19	Lecture 11: Multiple regression: introduction	PS12 assigned
Nov 26	**NO CLASS - Thanksgiving**	
Dec 3	Lecture 11: Multiple regression: introduction (cont.) (and review for final)	Project due
Dec 10	**Final exam**	

Notes:

- Due dates and the number of problem sets are subject to change.
- VU Fall Break is Thursday and Friday, October 10-11.
- I will be traveling on **Tuesday October 8** (midterm exam date) and **Thursday November 21** for APPAM. I will also be in NYC on **Monday, November 18**, which may require us to find a new day/time for our November 19 class meeting. I will have more details soon.