GOV 2905
Advanced Quantitative Analysis in Political Science
Spring 2022

MW
2:50-4:15 PM
Kanbar Hazelton Rm 109

Instructor: Michael Franz
Email: mfranz@bowdoin.edu
Phone: 207-798-4318 (office)
Office: 200 Hubbard Hall

Office Hours:
Tuesday, 9-11am
Thursday, 1-2pm
Book an appointment in Blackboard;
Or email about a different time as needed

This course introduces and discusses multivariate regression models used in political science. This includes an extended discussion of linear regression followed by a focus on regression models for dichotomous dependent variables (e.g., did a candidate win or lose), ordered variables (e.g., levels of support for candidates), and event counts (e.g., number of bills introduced by a member of Congress). The course considers and reviews ordinary least squares regression and then introduces maximum likelihood estimation. Students will apply these regression models by analyzing existing datasets using a statistical software package such as Stata or R. The final project will involve some analysis of survey data. Along with reviewing and explaining these statistical techniques, the course will involve extended discussions of the opportunities and challenges of the empirical study of political phenomena.

Learning Goals

In this course, you should:

1. Learn the basics of linear regression models in predicting the distribution of many dependent variables.
2. Learn the basics of maximum likelihood estimation techniques for regression models with dichotomous dependent variables, ordered variables, and event counts.
3. Examine various datasets and design and estimate a variety of regression models.
4. Improve the ability to read and discuss political science articles that use regression models.
5. Consider the broader issues (along with the challenges and opportunities) of using quantitative empirical techniques to study and understand political topics.

Course Requirements

There are four major components to your grade:

1. **Five problem sets** (50 points; each worth 10 points)—Due dates are listed on the syllabus. You will use datasets provided to you, and you will design and discuss statistical models using the covered techniques and estimation strategies. See the end of the syllabus for more information.

2. **Four reaction papers** (8 points; each worth 2 points)—Due dates are listed on the syllabus. These are short reactions of about 2 pages (double-spaced) that discuss assigned

articles. What did you think of the empirical approach in the article? Are there any issues with the data or analysis that stood out to you? There are no right or wrong answers, but I will challenge you to think logically. These papers are due by the start of class. To hand in a reaction paper, see the "Assignments" link on the left-side menu of our Blackboard page. I will grade and return the papers to you via Blackboard.

3. **Class participation** (20 points)—this includes attendance, class participation, and some regular posts on a Discussion Board*. I will spend a lot of time in many classes reviewing the foundations of various statistical models. But I encourage students to ask me questions about the material I cover. There will also be opportunities throughout the semester to discuss application articles that use the statistical models we cover.
*For each Monday until Spring Break we will read short sections of Matt Grossman's book, listed below. We will spend a few minutes on these Mondays discussing the points raised in the book, but in advance of class you should post a few thoughts/reaction to the Discussion Board on that week's reading. This will form the basis of some follow-up questions from me to the rest of the class.

4. **Final Assignment** (22 points)—See the end of the syllabus for more information. You will pick from one of two options on the final assignment.

Readings

There is one book for this course, and several outside articles. All the outside readings can be accessed through Blackboard.

- *How Social Science Got Better: Overcoming Bias with More Evidence, Diversity, and Self-Reflection*, by Matt Grossman, 2021: Oxford University Press

Other Issues

1. I expect all students to abide by the Bowdoin Academic Honor Code, which can be accessed online at: https://www.bowdoin.edu/dean-of-students/ccs/community-standards/the-codes.html. If you have any concerns or questions about how to cite work appropriately, please consult a reference librarian or me.

2. If you have chosen to take the class as Credit/D/F, I will only grant a Credit grade if the student has completed all the work for the class.

3. It is possible (maybe even likely) that you might need to miss a class this semester because of illness or a quarantine related to COVID-19. If that happens, I will schedule a time to meet with you via Zoom to review the missed course work and establish a plan for completing the work.

A Note on Software

Students are permitted to use any statistical software that they like, but I am proficient in Stata and R and can help students most if they use these options. My preference is for students to use R, and in particular R Studio Server accessed via the Bowdoin network (which you can use on your iPad or personal computer without any need to download software). Alternatively, I have used Stata for over 20 years, and it is available on most campus computers. I should be able to

help students who use Stata to diagnose any issues with data loading, recoding, or model estimation.

For the problem sets, I encourage students who use R to embed their code, output, and write-up in an R Markdown file, but it will also be sufficient to "copy and paste" output over to a word processing document. For students using Stata, since there is no easy Markdown equivalent (although there are work-arounds, which you can talk with me about if you like), I am ok if you "copy and paste" Stata output into a word processing document.

Ideally, in cases where students "copy and paste" output from R or Stata, I would strongly discourage screenshots of output. I recommend instead creating tables within the word processor where you can add the results in manually. Another option is to copy and paste output into an Appendix but put a subset of your results—your final model, for example—into a more pleasing-looking table that is embedded alongside the analysis/write-up. I will push you this semester to practice your skills in the presentation of statistical results.

Class Schedule

Monday, January 24—Introductions and Expectations: Models and Data [meet via Zoom]

Wednesday, January 26—Limits and Derivatives: the Baseline of Model Optimization [meet via Zoom]

Monday, January 31— Linear Regression in Scalar and Assumptions for Making Inferences

- Grossman, Preface and Chapter 1
- "Understanding Regression Analysis: An Introductory Guide," pp.11-36

Wednesday, February 2—Linear Regression in Scalar and Assumptions for Making Inferences, continued

- "Understanding Regression Analysis: An Introductory Guide," pp.36-53

Monday, February 7—Linear Regression in Scalar and Assumptions, continued

- Grossman, Chapters 2-3
- "Understanding Regression Analysis: An Introductory Guide," pp.53-80

Wednesday, February 9— Linear Regression in Scalar and Assumptions, continued

- Review also of data analysis software
- Appendix B, "Rudiments of Matrix Algebra," *Basic Econometrics*, Damodar Gujarati

Monday, February 14—Linear Regression in Matrix

- Grossman, Chapters 4-5
- "The Matrix Approach to Linear Regression Model," *Basic Econometrics*, Damodar Gujarati

Wednesday, February 16—Linear Regression in Matrix, continued

- No reading

- Review also of data analysis software and code

Monday, February 21—Example of Linear Regression
- Grossman, Chapters 6-7
- Jared S. Rosenberger and Rick Dierenfeldt. 2021. "Media and Confidence in the Police: Differences across Race/Ethnicity," *American Journal of Criminal Justice*.
- *Reaction paper 1 due by Class*

Wednesday, February 23—no class
- *Problem Set 1 due @ 11pm*

Monday, February 28—Maximum Likelihood Estimation and limitations of Least Squares Regression
- Grossman, Chapters 8-9

Wednesday, March 2—Logistic Regression
- "Logistic Regression: A Primer," pp. v-vii and 1-18

Monday, March 7—Logistic Regression, cont.
- Grossman, Chapters 10-11
- "Logistic Regression: A Primer," pp. 39-70

Wednesday, March 9—Logistic Regression, cont.
- *Problem Set 2 due @ 11pm*

March 14-25—Spring Break!

Monday, March 28—Predicted Probabilities in Logistic Regression
- No readings, but I will supply code for estimating predicted probabilities, which I will ask you to review before class

Wednesday, March 30—Predicted Probabilities in Logistic Regression, cont.
- No readings, but we will spend time in class on Problem Set 3

Monday, April 4— Examples of Logistic Regression in Political Science
- John Griffin and Michael Keane. 2006. "Descriptive Representation and the Composition of African-American Turnout," *American Journal of Political Science*.
- *Problem Set 3 due @ 11pm*
- *Reaction paper 2 due by Class*

Wednesday, April 6—Ordered Logits and Probits
- "Learn About Ordered Logit Regression in R With Data from the General Social Survey," Sage Research Methods

Monday, April 11—Ordered Logits and Probits, cont.

- Review reading from April 6

Wednesday, April 13—Ordered Logits and Probits, cont.
- No reading

Monday, April 18—Ordered Logits and Probits, cont.
- Larry Bartels. 2020. "Ethnic Antagonism Erodes Republicans' Commitment to Democracy," Proceedings of the National Academy of Sciences.

Wednesday, April 20—Examples of Ordered Logits and Probits in Political Science
- Stephen Nicolson and Gary Segura. 2012. "Who's the Party of the People? Economic Populism and the U.S. Public's Beliefs About Political Parties," *Political Behavior*.
- *Reaction paper 3 due by Class*

Monday, April 25—Event Counts
- Stefany Coxe, Stephen G. West, And Leona S. Aiken. 2009. "The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives," *Journal of Personality Assessment.*
- *Problem Set 4 due @ 11pm*

Wednesday, April 27—Event Counts, cont.
- Review reading from April 25

Monday, May 2—Event Counts, cont.
- Reading TBA

Wednesday, May 4—Examples of Event Counts in Political Science
- Morgan Johnstonbaugh. 2020. "Standing Up for Women? How Party and Gender Influence Politicians' Online Discussion of Planned Parenthood," *Journal of Women, Politics & Policy*
- *Reaction paper 4 due by Class*

Monday, May 9—Discussion of Final Assignment
- Reading TBA
- *Problem Set 5 due @ 11pm*

Wednesday, May 11—Discussion of Final Assignment, cont.

Final Assignment due by 5pm on May 21

Assignments

I will grade these according to the following general rubric:
- **Clarity** in the discussion and presentation of the results: I recommend working on presenting tables that look like what you might see in an academic paper. This goes also for any graphs you choose to include in these problem sets.

- **Competence** in the discussion of results/effects: Did you demonstrate that you understand the underlying estimates of each model? Did you consider multiple model specifications and discuss the results of these?
- **Code**: For all assignments, I will ask also that you submit your R script or Stata do-file. If you are using R Markdown, your code can be displayed and embedded alongside your write-up. Is your code clear and efficient? These are important skills to practice, and I will provide feedback on ways to clean up and annotate your code.

Problem Set 1: Linear Regression and Explaining the 2020 Presidential Election
I will provide you a dataset of county-level election results from the 2020 presidential election. There are over 3,000 counties in the United States. The data also contain a variety of columns with information on various characteristics about the county. The dependent variable is labeled "**biden20**." Consider the variables in the dataset and run some regression models, reporting the results. This should include discussion of the coefficient estimates for the independent variables, changes in estimates across models, and the R-squared values of the models. This assignment should be completed in 5-6 pages, including any tables and graphs. (Graphs are not required, but sometimes they are illustrative in showing a scatterplot of y on x. I'll leave it to you to decide whether a graph is worth including.) There is no time limit for completing the assignment once you access it.
*See the "Assignments" link on the left-side menu of our Blackboard page. I will grade and return the assignment to you via Blackboard.

Problem Set 2: Logistic Regression using the 2021 Bowdoin Polar Poll
I will provide to you the data from the Fall 2021 poll of Bowdoin students run as part of GOV 2070 (Quantitative Analysis in Political Science). The poll is referred in that class, and here, as the "Polar Poll." Consider the variable "**cancel1**," which is from a question asked of students: "Do you favor or oppose the following action sometimes taken by colleges and universities? Disinviting speakers because some students perceive their message as offensive or biased against certain groups of people." Favor=1 and Oppose=0. Run a few logistic regression models that predict the probability of a student favoring this action. I would recommend a few models with 5-7 independent variables. In this assignment, it is likely necessary (and encouraged) to recode some variables (i.e., creating dummy variables). This assignment should be completed in 5-6 pages, including any tables and graphs. There is no time limit for completing the assignment once you access it.
*See the "Assignments" link on the left-side menu of our Blackboard page. I will grade and return the assignment to you via Blackboard.

Problem Set 3: Predicted Probabilities from the 2021 Bowdoin Polar Poll
Start with the results from Problem Set 2. Take a few of the models you used in that Problem Set and estimate predicted probabilities from your models for various categories of students. Interpret and discuss the results. This assignment should be completed in 4-5 pages, including any tables and graphs.
*See the "Assignments" link on the left-side menu of our Blackboard page. I will grade and return the assignment to you via Blackboard.

Problem Set 4: Ordered Probits and Logits using the 2021 Bowdoin Polar Poll

Again using the "Polar Poll", there are a number of variables in the dataset that are ordered variables. For example:

- "**corona**" asks for students' level of support of the College's COVID-19 policies.
- "**jboard_1**" asks for students' evaluation of the change in the name of the Judicial Board to the Conduct Review Board
- "**yikyak1**" asks students to express their opinion of YikYak
- "**DEI**" asks students their opinion on how often efforts to enhance diversity, equity, and inclusion on campus conflicts with free speech.
- "**mental**" asks students their opinion on the sufficiency of mental health resources on campus.

Choose one as a dependent variable. Run a few ordered probit or logit regression models that predict the probability of being in the various categories of the dependent variable. I would recommend a few models with 5-7 independent variables. This assignment should be completed in 5-6 pages, including any tables and graphs. There is no time limit for completing the assignment once you access it.
*See the "Assignments" link on the left-side menu of our Blackboard page. I will grade and return the assignment to you via Blackboard.

Problem Set 5: Event Counts of Presidential Vetoes
I will provide to you a count of the number of presidential vetoes in each congressional session from 1947 to 2021. Construct an event count model predicting the number of vetoes in each term using the set of independent variables also provided in the dataset. First consider a Poisson regression model and evaluate the results. Next consider whether you should run a negative binomial regression model. Run that model and report and compare the results with the Poisson results. This assignment should be completed in 5-6 pages, including any tables and graphs.
*See the "Assignments" link on the left-side menu of our Blackboard page. I will grade and return the assignment to you via Blackboard.

Final Assignment: You have two options:

1. Download a survey from the American National Election Study. There is an election-year survey available for most elections back to 1948. I'll let you decide which survey to examine. Perhaps you are interested in examining survey results in the 1960 election year, for example. Using the codebook, select a dependent variable. Construct a regression model (OLS, logit/probit, ordered logit/probit, or event count; select the one appropriate to the distribution of the dependent variable) with the independent variables you think are important. Run the model and report and explain the results. This assignment should be completed in 8-10 pages, including any tables and graphs.
2. Download a survey from the American Trends Panel at the Pew Research Center. These surveys extend back to 2014. Follow the instructions above. Using the codebook, select a dependent variable. And so on.

*In either option, explain why you selected that dataset and dependent variable. You should also discuss predicted effects in addition to explaining the coefficient estimates from the models you run.