UNIVERSITY OF MARYLAND

GVPT 622: Quantitative Methods for Political Science
Fall 2022

Friday 11:00 am to 1:45 pm (EST) in Susquehanna Hall 1117
(There will be a short break at 12:00pm.)

Instructor: Jacob Aronson
Email: jacobaro@umd.edu
Office hours: **Friday** 2:00 pm to 3:00 pm and by appointment
Office: 2117H Chincoteague Hall (in the CIDCM section, back and to the right)

Course Overview

This course will teach you how to think scientifically about data—its development, manipulation, and use. This is an important skill for students who wish to pursue academic, government, or private-sector careers. Very few career paths these days do not intersect with data in some way. Specifically, students will learn how to formulate research questions, and develop plans to collect, transform, and analyze data to test these questions.

A central element to this is statistical inference. This course will combine learning the principles of statistics as used in political science along with hands-on data analysis. While many class examples and assignments will focus on domestic politics and international armed conflict, the skills learned in this class are applicable across most topic areas. This class will focus on providing students with the ability to conduct simple univariate and multivariate regression using linear models (Ordinary Least Squares or "lm"). The companion class that students will take next semester, GVPT 729, will introduce the use of more advanced generalized linear models.

Course Description and Goals

While use of case studies is more common in policy analysis, use of statistics is more common in academic political science. The goal of both methods is the same: to provide evidence that supports or disproves a hypothesized relationship between variables. A case study often allows a "deeper dive" into the inner workings of a case but has difficulty disentangling the effects of a variable from its specific context or identifying how findings speak to other situations. Statistical analysis is the equivalent of many case studies. Variables are recorded in a consistent and reproducible way across cases. This allows a researcher to separate the general effect of a variable (its coefficient) from any case-specific idiosyncrasies (captured in the error term). The downside to statistical analysis is that it may be harder to develop a deep understanding of case-specific dynamics and variables must be chosen that can be collected across many cases.

One way to think about this difference is to compare work conducted by social scientists, which this class seeks to prepare students for, to work done by journalists. Journalists usually seek to "describe" data in a way that makes it easier for others to understand. Think of the daily news—a retelling of what happened in a day (the day's "data") in a manner that is easy for the viewer to understand. By contrast, social scientists seek to use data to explain *new* situations. A journalist

may describe crime that occurred in a location. A social scientist seeks to use this data to describe what may happen tomorrow or in another location. Students who take this class will have a better understanding of this process, its conduct, and challenges.

The class is divided into three sections. The first section focuses on the basics of statistical inference and how to develop a good research design. The second section will provide students with the skills to load and explore data. On Week 5 students will learn how to install and use R and a particularly important R package, Tidyverse. The third section will teach students how to conduct basic statistical analysis that relates one variable to another using a linear model. We will finish by examining a well-done application.

Course Reading and Software Requirements

One book listed below is required for this course. You must also have access to (and use) the statistical package R. Students are required to conduct work on their own laptop, which they will need to bring to class when we work with RStudio. R can be downloaded for free at https://cran.r-project.org/. RStudio, an integrated development environment (IDE), can be downloaded for free at https://www.rstudio.com/products/rstudio/download/. The course will walk students through the process of installing and using this software on Week 5.

One book and one PDF on ELMS are required for this course:

- Andrew Gelman, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories*. Cambridge University Press.
- (**PDF is on ELMS so you do *not* need to purchase this book.**) Gary King, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.

While the book introduces students to both classical and Bayesian inference, we will focus on classical inference as it is more common in applied political science. The book *Regression and Other Stories* will also be used for GVPT 729 so hold on to it!

All reading material aside from the textbook will be posted on the course ELMS page.

Simple R tutorials are available at https://www.w3schools.com/r/default.asp and at https://education.rstudio.com/learn/. The back of the textbook (Appendix A) also provides information on how to program in R. Resources on Tidyverse (the R package we will use extensively) is available at https://www.tidyverse.org/learn/. More complicated tutorials on using Tidyverse for data science are available at https://www.tidymodels.org/books/moderndive/. Programming in R is a challenging but beneficial skill to know—it is worth the investment.

Campus Policies

It is our shared responsibility to know and abide by the University of Maryland's policies that relate to all courses, which include topics like:

- Academic integrity
- Student and instructor conduct

- Accessibility and accommodations
- Attendance and excused absences
- Grades and appeals
- Copyright and intellectual property

Please see the University's website for graduate course-related policies at: https://gradschool.umd.edu/course-related-policies.

Course Requirements and Evaluation

Your final grade will reflect the sum of points (100%) earned from the following categories:

- In-class Assignment (September 23rd): 5%
- Homework Assignment #1 (October 21st): 10%
- Homework Assignment #2 (November 4th): 10%
- Homework Assignment #3 (November 18th): 10%
- Homework Assignment #4 (December 2nd): 10%
- Class Participation: 25%
- Final Exam: 30%

Homework assignments

The four homework assignments will be composed of two to four questions pulled from the list of exercises at the end of each chapter in Gelman, Hill, & Vehtari (2020). The homework will include questions only from the weeks specified in the assignment (listed in the week it is due). Students are free to work together to solves these problems. All homework assignments must be turned in BEFORE CLASS BEGINS ON THE DAY THAT THE ASSIGNMENT IS DUE. Late assignments will not be accepted. It is much better to get half credit than no credit! I will go over each homework assignment in class on the day that it is due. If a student has a Covid-related absence, they can finish a makeup homework assignment once they are feeling better.

Class participation

Students must be prepared to develop and run R code during most class periods. This means that students will need to bring a laptop into class with R Studio installed and ready to run. Week 5 will be spent making sure everyone is able to install and run R. Class periods may include in-class exercises that will be included in the class participation grade. Students who successfully conduct in-class exercises and who ask questions as needed will receive full credit for class participation. Do not hesitate to ask questions. Questions help everyone to better understand the material and contribute to your grade!

Final exam

The final exam will involve a question or two from the exercises at the end of each chapter of Gelman, Hill, & Vehtari (2020) for a total of 10-15 questions. All chapters covered in class are fair game for the final exame. These questions will not be identified before the final but will be like those on the homework—think of the final exam as a long in-class homework assignment. The final exam will need to be completed individually in-class during the final exam period at

the end of the semester. Students who thoroughly read each chapter, are familiar with all the exercises at the end of each chapter, and who have done well on homework assignments will be well-prepared for the final exam. You must pass the final exam to pass the class. The final exam will be graded on a curve.

Course Evaluation

Please submit a course evaluation through CourseEvalUM to help faculty and administrators improve teaching and learning at Maryland. All information submitted to CourseEvalUM is confidential. Campus will notify you when CourseEvalUM is open for you to complete your evaluations for fall semester courses. Please go directly to the Course Eval UM website to complete your evaluations.

Grading Scale

Grades will be based on the following scale:

100 to 98 = A+	74 to 76 = C
94 to 97 = A	70 to 73 = C-
90 to 93 = A-	67 to 69 = D+
87 to 89 = B+	64 to 66 = D
84 to 86 = B	60 to 63 = D-
80 to 83 = B-	0 to 59 = F
77 to 79 = C+	

Course Schedule

Section 1: Research Design and Principles of Quantitative Analysis

Week 1 Introduction September 2nd, 2022

This class period will focus on describing the format of the class, identifying the topics covered in the class, listing class goals, and addressing any student questions. Please read the syllabus and ask questions if any portion requires clarification.

Week 2 Social Science and Research Design September 9th, 2022

- King, Keohane, and Verba, *Designing Social Inquiry*, 1994. Chapters 1-2.

Week 3 The Basics of Causal Analysis September 16th, 2022

- King, Keohane, and Verba, *Designing Social Inquiry*, 1994. Chapter 3.

- Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2008. "Misunderstandings Between Experimentalists and Observationalists about Causal Inference." *Royal Statistical Society* 171(2): 481-502.

Week 4 Developing a Research Design September 23rd, 2022

- Jeffry A. Frieden and David A. Lake. 2005. "International Relations as a Social Science: Rigor and Relevance." *The ANNALS of the American Academy of Political and Social Science* 600(1): 136–56.
- Karl Gustaffsson and Linus Hagstrröm. 2018. "What is the Point? Teaching Graduate Students How to Construct Political Science Research Puzzles." *European Political Science* 17: 634-648.

Students will complete an in-class assignment where they will be required to write two research questions relevant to their research interests. This assignment will be turned in through ELMS and graded using the following scale: zero, half, and full credit.

Section 2: Data Exploration

Week 5 Installing R and RStudio September 30th, 2022

The class period will be spent installing R, RStudio, and important packages. We will make sure that the software can be loaded and that it works. This will prepare students for more advanced activities in the future. R is available at: https://cran.r-project.org/. RStudio is available at: https://www.rstudio.com/products/rstudio/download/.

Students will also be introduced to the R packagge Tidyverse. Tidyverse is an R meta-package that makes loading, manipulating, and processing data easier. The package can be installed by typing the command 'install.packages("tidyverse")' into the R console. While learning this may seem like extra work it will DRAMATICALLY improve your productivity as a data scientist/wrangler or graduate student. More information on this package is available here: https://www.tidyverse.org/.

For more information on R see also: Appendix A in Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. (Students are not required to read this Appendix but it does provide valuable discussion about how to install R/RStudio and how to program in R.)

Week 6 Statistics, Data, and Measurement October 7th, 2022

- Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. Chapters 1-2.

Depending on progress in the prior week, the class may also focus on installing and using R and Tidyverse.

Throughout the rest of the class, we will use both actual data and "fake data" that we create ourselves using various distributions.
The benefit of using fake data is that we control the data generation process (we specify how independent variables relate to an outcome). This allows us to demonstrate that statistical models—when correctly used and specified—uncover the data generation process that we specify. This is intended to build an understanding of the relationship between data and statistical models and to build confidence in the process of statistical analysis. A statistical model is just an attempt to build a model of the data generation process in a dataset.

By understanding how a statistical model relates to fake data, students are better equipped to understand how statistical outputs relate to real data. For instance, after running analysis of real data, students can use their model to build a reconstructed ("fake") dataset and compare this reconstructed dataset to the real data. High similarity implies success in recreating the data generation process. High dissimilarity implies failure.

Week 7 Graphing and Basic Probability October 14th, 2022

- Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. Chapter 3.

Week 8 Statistical Inference October 21st, 2022

- Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. Chapter 4.

Homework assignment #1 (covering weeks 6-7) is due.

Week 9 Simulation and Use of the Bootstrap October 28th, 2022

- Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. Chapters 5.

Section 3: Basic Regression Analysis

Week 10 Regression I: Introduction November 4th, 2022

- Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. Chapter 6.

Homework assignment #2 (covering weeks 8-9) is due.

Week 11 Regression II: One Variable November 11th, 2022

- Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. Chapters 7-8.

Week 12 Regression III: Multiple Variables November 18th, 2022

- Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. Chapter 10.

Homework assignment #3 (covering weeks 10-11) is due.

Week 13 **No Class (Thanksgiving)** November 25th, 2022

Week 14 Regression III: Assumptions and Transformations December 2nd, 2022

- Gelman, Hill, and Vehtari, *Regression and Other Stories*, 2020. Chapters 11-12.

Homework assignment #4 (covering week 12) is due.

Week 15 Wrap Up, Applications, and Other Models December 9th, 2022

- Eli Berman, Jacob N. Shapiro, and Joseph H. Felter. 2011. "Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq." *Journal of Political Economy* 119(4): 766-819.

In the first part of the class, we will discuss an article that uses simple statistical analysis to develop important causal findings. We will then go over student questions about the homework assignments and the Final Exam. Questions about other end-of-chapter exercises that may be assigned for the final exam will also be addressed. Finally, we will briefly discuss generalized linear models (Logit, Survival, Count, etc.) that students may desire to incorporate into their Capstone project. Students will learn more about these models next year when they take GVPT729.

*****Final Exam will be held during finals week*****