Jeff Gill
1737 Cambridge St. N409

Government 52: Models:
Date/Time: TBD. Location: TBD.

- Course Description:

  This course covers statistical models useful for investigating political and social phenomena. We will address the theory and principles behind these methods, their applications, and their limitations. The course will be useful for those undertaking a quantitivative thesis or wanting to have data science modeling skills for other applications. Specific models to be studied include the standard linear regression model in detail, and Generalized Linear Model forms such as logit, probit, log-linear (Poisson), Gamma, ordered choice, multinomial, tobit, and more. In addition Bayesian estimation, Monte Carlo simulation, multilevel, and nonparametric models will be covered.

  We will make extensive use of the software environment **R**. Students are expected to have proficiency, but all new models will be extensively illustrated in **R**.

  The structure of the course is focused on active learning. The first meeting of the week will be a traditional lecture where the statistical modeling concepts are explained and shown how they are estimated in **R**. The second meeting of the week will be in virtual break-out rooms where students will jointly solve one of the weekly exercises in small teams with consultation from the teaching fellows and faculty member.

- Course Requirements:
  - *Regular Attendance: 20%.*
    Reading Prior to Class, Active Class Participation.
  - *Weekly Exercises: 60%.*
  - *Project: 20%.*
    Each participant will do an original data analysis using the models presented in the course with data of their own choosing. These will be presented at the final class meeting. This is not a research paper: it is the middle of a quantitative social science research paper meaning an explanation of the data, the model used, and a description of the results.
- General Policies:

  A student requiring academic adjustments or accommodation is requested to present his/her official letter from the Accessible Education Office and speak with me by the end of September. All such discussions will remain confidential, although there may be a need to consult the AEO. A student who has a disability that may require some modification in seating or class configuration should contact me as soon as possible in September. The best place to obtain further information is the Student Disability Center at 20 Garden Street (496-8707).

  Special Note for Spring 2021: Spring Break 2021 is replaced by individual wellness days on Feb. 5, Mar. 1, Mar. 16, Mar. 31, and Apr. 15. Courses and related course meetings will not meet on these days.

- Office Hours: TBD.

- Teaching Fellows: TBD.

- Text:
  Gelman and Hill, "Data Analysis Using Regression and Multilevel/Hierarchical Models (Cambridge University Press 2007).

- Course Content, By Topic:
  1. **Introduction to the Course**
     - Remarks and explanation of the goals for the course.
     - Motivating Example
     - Reading: Gelman & Hill Chapter 2
  2. **Review of the Linear Model**
     - Linear Regression, Correlation, Estimation, and Inference
     - Reading: Gelman & Hill Chapter 3
  3. **Linear Modeling in Matrix Notation**
     - Introducing a More Efficient Notation for Calculations and Descriptions
     - Reading: Gelman & Hill Chapter 4
  4. **Regression Models for Dichotomous Outcomes**
     - Modeling Expected Outcomes
     - Logit, Probit, cloglog
     - Reading: Gelman & Hill, Chapter 5
  5. **Introducting Generalized Linear Models**
     - Poisson Regression, Exposure, and Overdispersion
     - Logistic-binomial model
     - Multinomial regression
     - The Exponential Family Form
     - Deviances
     - Reading: Gelman & Hill, Chapter 6.1 to 6.5
  6. **More on Generalized Linear Models**
     - Robust Regression
     - Ordered Regression
     - Multinomial Models
     - Reading: Gelman & Hill, Chapter 6.6 to 6.9
  7. **Simulation of Probability Models and Statistical Inferences**
     - Introduction to Simulation Tools
     - Summarizing Linear Regressions Using Simulation
     - Simulation for Nonlinear Predictions
     - Predictive Simulation for Generalized Linear Models
     - Reading: Gelman & Hill, Chapters 7, 8
  8. **Causal Inference**
     - Causal Inference and Predictive Comparisons
     - The Fundamental Problem of Causal Inference
     - Randomized Experiments
     - Understanding Causal Inference in Observational Studies
     - Intermediate Outcomes and Causal Paths
     - Imbalance and Lack of Complete Overlap
     - Matching: Subsetting the Data to Get Overlapping and Balanced Groups
     - Instrumental Variables in a Regression Framework
     - Reading: Gelman & Hill, Chapters 9, 10
  9. **Multilevel Models**
     - Notation
     - Varying-Intercept and Varying-Slope models

- Time-Series Cross Sections, and Other Non-Nested structures
- Indicator Variables and Fixed or Random Effects
- Partial Pooling with No Predictors
- Partial Pooling with Predictors
- Group-Level Predictors
- Model Building and Statistical Significance
- Predictions for New rOservations and New Groups
- Reading: Gelman & Hill, Chapters 11, 12

10. **Handling Missing Data**
    - The Types of Missing Data
    - Why Case-wise Deletion is Evil
    - Common Ways to Deal with Missingness
    - Random Imputation
    - Maximum Likelihood
    - Multiple Imputation (Rubin 1979)
    - Applying mice() To Actual Mice
    - Old-Style Hot-Deck Imputation
    - Multiple Hot-Deck Imputation (Cranmer Gill 2012)
    - What About Missingness On the Outcome Variable?
    - Multiple Imputation Then Deletion (MID)
    - Expectation-Maximization
    - Bayesian Estimation of Missing Data
    - Multiple Imputation, Then Prediction (MIP)
    - Reading: Handout
11. **Nonparametric Modeling**
    - Smoothing, Starting Vocabulary
    - Linear Regression as a Smoother
    - Interpolation as a Smoother
    - Bin Smoother
    - Running Mean and Line Smoothers
    - The Lowess Smoother
    - Regression Splines
    - Cubic Regression Splines
    - Penalized Splines
    - Thin Plate Splines
    - Tensor Product Smoothing
    - Generalized Additive Models
    - Predictions From GAM Output
    - Reading: Handout