POS 6933: MAXIMUM LIKELIHOOD ESTIMATION – FALL 2025

Department of Political Science,
University of Florida
Monday: Periods 5-7; RNK 0225
Credit Hours: 3

Instructor: Prof. Badredine Arfi
Office: 221 Anderson Hall
Office Hours: M 3:15-4:15pm, W 12:30-2:30pm
Or: by appointment
Phone: (352) 273 2357
Email: barfi@ufl.edu

COURSE DESCRIPTION AND OBJECTIVES

Political science research often grapples with complex data structures and nuanced causal relationships that exceed the capabilities of the Classical Linear Regression Model (CLRM). While CLRM remains a foundational tool, its assumptions—such as homoscedasticity, linearity, and normally distributed errors—are frequently violated when studying real-world political phenomena. Dependent variables of substantive interest are often categorical, ordinal, count-based, or censored, and data may be subject to selection bias, endogeneity, or missingness To address these challenges, this course introduces advanced statistical methods rooted in Maximum Likelihood Estimation (MLE), with an emphasis on Generalized Linear Models (GLMs), Survival Models, and Causal Inference Models. These frameworks extend the classical approach and are essential for modeling limited dependent variables, time-to-event data, and causal inference. The focus is both theoretical and mostly applied, equipping students to analyze data across American politics, comparative politics, and international relations.

The course is structured into three thematic parts:

1. **Generalized Linear Models (GLMs**): We begin by covering a suite of models for non-continuous outcomes, including logit, probit, multinomial logit, ordinal logit/probit, and count models such as Poisson and negative binomial regressions, selection bias models. Emphasis is placed on model specification, diagnostics, and interpretation in applied research.

2. **Survival and Event History Analysis**: We then explore models where the timing of events is the key variable of interest. This includes parametric and semi-parametric models (e.g., Cox proportional hazards, exponential, Weibull), with applications in political/social science processes.

3. **Causal Inference**: The final module addresses causal inference, a central concern in empirical social science. While MLE models estimate associations, understanding causal mechanisms requires more than prediction. This part introduces frameworks and tools for causal inference, including Potential Outcomes Framework (Rubin Causal Model), Counterfactual reasoning and assumptions, Matching methods (propensity score matching, coarsened exact matching), Difference-in-Differences (DiD) and fixed effects models, Instrumental Variables (IV), Regression Discontinuity Designs (RDD), and Sensitivity analyses and robustness checks.

Throughout the semester, students will gain hands-on experience applying these models to real political/social science datasets, using the python computer language. By the end of the course, students will be equipped not only to choose appropriate statistical models for their data, but also to critically evaluate causal claims and contribute rigorously to empirical debates in political science.

Computing Language:

Although social scientists are usually trained in using software packages such as Stata, SPSS, Mplus, SAS, and R, data sciences practitioners find those quite limiting and instead use languages such as Python, C++, Julia, and a few others.

This course introduces students to learning and using from scratch Python. Why? Not only is Python (like R) a free software, but it does also surpass R by far in its practicality and in the availability of very large numbers of powerful packages that make data analysis much richer and more versatile in every field of knowledge. Python is quite flexible in its semantics and comes very close to human natural language in many respects. Moreover, by and large, the fields of AI, machine learning, and more generally data sciences use Python, and hence this course will equip social scientists to join ranks with data scientists in other fields of knowledge in deploying powerful methodologies of AI to produce theoretical and practical knowledge. In short, doing statistical analysis using Python packages is quite straightforward and surpasses the language Stan (used mostly by statisticians) which R practitioners draw on in developing various R statistical packages.

Because the great majority of social scientists have never been introduced to Python, the course will start from the very beginners' level of Python. No prior knowledge or experience with Python is expected. We will spend together hours learning and applying Python throughout the whole semester. Every session of the course will contain a conceptual as well as a hands-on deployment of Python to analyze datasets pertaining to the session at hand. Students who are familiar with R will quickly realize that the similarities are quite large between the two languages and will have no difficulties moving between the two languages, as well as quickly understanding the versatility and practicality of Python. The class will be 'walked' into installing and deploying Python and various packages needed for the course on their personal computers during the first week of the semester. In learning how to do statistical analysis in any computer language, students must invest in learning 'software-ways' of how to do it, and Python is not unique in this respect.

Course Objectives/Learning Outcomes:

1. **Explain and Interpret Generalized Linear, Survival Analysis Models, and Causal Inference Models in Political/Social Science Research**: Students will articulate the concepts and assumptions of generalized linear models, survival analysis, and causal inference and interpret their applications in American politics, comparative politics, and international relations, and, more generally, social sciences.
2. **Apply Statistical Knowledge to Model Development in Political/Social Science**: Students will demonstrate the ability to apply their statistical knowledge to develop, estimate, and validate both linear and non-linear models using appropriate statistical software, focusing on practical applications in political/social science research.
3. **Analyze Political/Social Science Data with Advanced Techniques**: Students will employ their statistical knowledge to conduct rigorous data analyses pertinent to their research fields within political/social science, critically assessing the results and implications for American politics, comparative politics, and international relations, and, more generally, social sciences.
4. **Synthesize Learning into Applied Political/Social Research**: By the end of the semester, students will synthesize their learning to complete a research paper that employs the methodologies covered in the course. This paper should demonstrate a deep understanding of the statistical models and their practical applications in a specific political/social science research context, aiming to achieve a high academic standard.

SOME RECOMMENDED TEXTS

Generalized Linear Models, MLE, and more …

1. Michael Smithson and Edgar C. Merkle. 2014. Generalized Linear Models for Categorical and Continuous Limited Dependent Variables. CRC Press.
2. Alan Agresti. 2015. Foundations of Linear and Generalized Linear Models. Wiley Press.
3. Gerhard Tutz. 2012. Regression for Categorical Data. Cambridge University Press.

Survival Analysis

1. Nag, Avishek. Survival Analysis with Python. 2022. CRC.

Causal Inference

1. Hernán MA, Robins JM (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. https://miguelhernan.org/whatifbook

2. Keele L. (2015) The Statistics of Causal Inference: A View from Political Methodology. *Political Analysis*. 23(3): 313-335.
3. Imbens, Guido W. (2024) Causal Inference in the Social Sciences. *Annual Review of Statistics and Its Application*. (11): 123-152. https://doi.org/10.1146/annurev-statistics-033121-114601
4. Jiao L, Wang Y, Liu X, Li L, Liu F, Ma W, Guo Y, Chen P, Yang S, Hou B. (2024). Causal Inference Meets Deep Learning: A Comprehensive Survey. *Research* 7: 0467. https://doi.org/10.34133/research.0467
5. Morgan, Stephen L. and Christopher Winship. (2014). Counterfactuals and Causal Inference: Methods and Principles f or Social Research. Cambridge University Press. 2nd Edition. https://doi.org/10.1017/CBO9781107587991

Python Introductory Books

1. Fabio Nelli. 2018. Python Data Analytics with Pandas, NumPy, and Matplotlib. Second Edition. APress.
2. Claus Führer, Olivier Verdier, and Jan Erik Solem. 2021. Scientific Computing with Python: High-performance scientific computing with NumPy, SciPy, and Pandas. Second Edition. Packt.
3. Robert Johansson. 2019. Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib. Second Edition. APress.
4. Ashwin Pajankar. 2022. Hands-on Matplotlib Learn Plotting and Visualizations with Python 3. APress.

ADDITIONAL MATERIAL ON CANVAS

Additional readings and materials (Jupyter Notebooks) will be posted on canvas site for the course at appropriate times during the semester.

REQUIREMENTS AND ASSESSMENT

The requirement for this course is simple (as always): work diligently and persistently. This includes attending classes and working regularly on the computer applications, the homeworks, and the research paper. Each student should expect to be spending many hours learning how to excel in using the python packages used to estimate the models discussed in the class.

There will be several homework assignments that students must complete and turn in. The homework assignments are due on the specified dates; no late submission is acceptable. All materials related to a specific homework should be collected in a Notebook and uploaded to canvas. There will also be a final take-home exam, the specifics of which will be discussed in class in time. Roughly speaking, it will consist in answering several questions by analyzing a dataset that will be provided to you with the questions.

A major component of the course evaluation will be a term research paper. Each student will produce a manuscript of high quality using an appropriate modelling strategy (specifics of the paper are discussed down below).

DISTRIBUTION OF GRADES

10%: Weekly homework exercises.
All assignments are to be uploaded to canvas before the beginning of class on their respective due dates. No late submission will be accepted for any reason (except when justified with university sanctioned documentation). The problem sets will be assigned at the end of the lectures depending on what we cover in the lecture sessions. Students are expected their individual responses to the homeworks. Please beware of plagiarism.

10%: Each student will be assigned "presentations" for exercises practice session of the course which will consist in presenting the weekly assigned homework (this will be fully explained on the first day of class).

30%: Take-Home Final Examination

The final exam is a take-home and open-book, open-computer, open-anything-but-another-human-being (physical or virtual).

40%: A Research Paper
Each student is required to choose in consultation with the instructor a research topic. The student is required to find a dataset suitable for the topic and construct a set of research questions. The goal is to produce a high-quality, potentially publishable research manuscript, using a model (or models) discussed in the course, estimated using python packages.

10%: Paper Presentation.
Each student will present his/her paper at the end of the semester (Data/time: TBD). The presentation will consist of a ppt presentation for about 15 minutes followed by 5 minutes of Q & A.

Your final cumulative score will be translated into a letter grade according to the following schedule: 93 points or higher = A; 90–92.9 = A-; 87–89.9 = B+; 83–86.9 = B; 80–82.9 = B-; 77–79.9 = C+; 73–76.9 = C; 70–72.9 = C-; 67–69.9 = D+; 63–66.9 = D; 60–62.9 = D-; <60 = E.
Information on current UF grading policies for assigning grade points. This may be achieved by including a link to the web page:
https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/

RESEARCH PAPER GUIDELINES

The primary assignment for this course is a substantive, methods-driven research paper that applies one or more of the advanced statistical techniques discussed during the semester. To ensure thoughtful progress and allow for instructor feedback at critical stages, students will submit a series of short intermediate assignments. These will culminate in a final paper that emphasizes methodological rigor while engaging a substantive question of interest.

While the substantive topic is your own choice, remember that this is a methods course: your grade will be weighted more heavily toward your use, justification, and understanding of statistical modeling than on the novelty of your political science question. That said, the strongest papers integrate theory, data, and methods in a coherent and compelling way.

Development Timeline and Deliverables

1. Topic and Dataset Proposal
Due: October 13

- Choose a political/social science topic that interests you.
- Identify a suitable dataset that enables meaningful application of advanced statistical methods (GLMs, survival models, causal inference techniques, etc.).
- Discuss it with the instructor during office hours
- Submit a short (1–2 page) proposal that:
    - States your research question
    - Describes the dataset
    - Explains how the topic aligns with course methodology

Note: Emphasis should be placed on the methodological challenge and not solely on the substantive research question.

---

2. Data Exploration Report
Due: October 27

- Submit a brief report (2–3 pages) that describes:
    - Key variables and their measurement levels
    - Summary statistics and preliminary visualizations

  - Missing data or other data quality concerns
  - Initial reflections on potential modeling strategies

---

3. Research Design and Model Selection

Due: November 17

- Outline your emerging research design, including:
  - Main hypotheses
  - Identification strategy (if applicable)
  - Justification for your choice of model(s)
- Explicitly connect your methodological choice to course material.
- Avoid over-reliance on simple logit/probit models — these do not provide enough methodological complexity for this course. Instead, consider:
  - Multinomial or ordinal logit/probit
  - Count models (Poisson, negative binomial, zero-inflated)
  - Duration/survival models
  - Causal inference designs (Difference-in-Differences, Regression Discontinuity Designs, Instrumental Variables, Matching)

Submit a concise 2–3 page memo.

---

4. Final Paper Submission

Due: December – last class

Your final research paper should be 15–20 double-spaced pages, including references, and must include the following components:

- Preprocessing and Data Cleaning:
  - Discuss data preparation, variable construction, and any transformations.
- Model Construction and Justification:
  - Describe your chosen statistical model and explain why it is appropriate for your research design.
  - Clearly state model assumptions and assess whether they are met.
- Estimation and Results:
  - Present estimation output with interpretation.
  - Include robustness checks or alternative specifications as needed.
- Substantive and Methodological Conclusions:
  - Reflect on the validity of your model and analysis.
  - Discuss limitations and potential improvements for future research.

Final Submission Package Requirements

Submit a .zip folder to Canvas containing:

1. Your final research paper (as a .pdf, .docx, or .tex file).
2. A fully annotated Python Jupyter Notebook:
   - Includes all code for data preprocessing, model estimation, and result visualization.
   - Should allow full replication of the paper's analysis from start to finish.
3. Final dataset used in the paper (in .csv or .dta format).
4. Any relevant supplementary material (e.g., codebooks, appendices, graphs).

A Note on Research Goals

The instructor is committed to mentoring students through the paper-writing process and helping each student develop their work into a potentially publishable piece. Use office hours and draft feedback as opportunities to push your work beyond course requirements.

IMPORTANT DATES

Class Begins	**Monday, August 25**
Holidays: No classes	**September 1: Labor Day** **October 17-18: Homecoming** **November 11: Veteran's Day** **November 24 - 29: Thanksgiving Break**
Class Ends	**Monday, December 1**

COURSE OUTLINE:

Introduction and Overview: MLE in Social Science Research

Generalized Linear Models:
- Binary Models: Logit and Probit Analysis
- Ordinal Outcomes: Ordered Logit and Ordered Probit Analysis
- Limited Outcomes: Tobit Model
- Heckman Model and Other Sample Selection Models
- Regression Models for Count Dependent Variables

Survival Analysis
- The Logic of Survival Analysis and Duration
- The Cox Proportional Hazards Model
- Time-Varying Covariates and Unobserved Heterogeneities

Causal Inference
- Models for Causal Inference
- Potential outcomes and counterfactual reasoning
- Randomized experiments and designs, Classical causal analysis for randomized experiments
- Propensity score, Estimation of propensity score, Matching and trimming
- Instrumental variable estimation

IMPORTANT NOTES:

This course complies with all UF academic policies. For information on those polices and for resources for students, please see this link.