POS 6xxx: Maximum Likelihood Estimation – Fall 2024

Department of Political Science,
University of Florida

Monday: Periods 5-7; MAT 0251

Instructor: Prof. Badredine Arfi
Office: 221 Anderson Hall
Office Hours: W 1-3pm
Or: by appointment

Phone: (352) 273 2357
Email: barfi@ufl.edu

COURSE DESCRIPTION AND OBJECTIVES

The Classical Linear Regression Model (CLRM) is a useful but limited tool for statistically studying most political phenomena. The dependent variables of interest to political scientists often do not fit the assumptions of the CLRM. To address these limitations, methodologists and statisticians have developed sophisticated methods for analyzing data with qualitative and ordinal dependent variables, count models, and situations involving selection bias. Maximum Likelihood Estimation (MLE), as applied in Generalized Linear Models (GLMs) and Survival Models, provides a general framework for estimating models that address these complexities.

This course focuses on applied research in political science, using MLE-based methodologies to analyze data within American politics, comparative politics, and international relations. Students will learn the theory and practical application of GLMs and survival models, using these powerful statistical tools to investigate problems of interest in political science. Students are expected to use these powerful statistical models to research problems that interest them.

The course consists of two parts. The first part deals with several models suited for analyzing problems with categorical and limited dependent variables, commonly known as Generalized Linear Models (GLMs). The second part introduces the framework of event history modeling, also known as survival analysis, where the dependent variable is duration or time-to-event.

The course starts with the basic concepts of statistical analysis in GLMs and incrementally covers more advanced computational methods. By the end of the semester, students will be equipped to conduct applied political research using advanced statistical methodologies.

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.

Yet, 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 and Survival Analysis Models in Political Science Research**: Students will articulate the concepts and assumptions of generalized linear models and survival analysis and interpret their applications in American politics, comparative politics, and international relations.
2. **Apply Statistical Knowledge to Model Development in Political 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 science research.
3. **Analyze Political Science Data with Advanced Techniques**: Students will employ their statistical knowledge to conduct rigorous data analyses pertinent to their research fields within political science, critically assessing the results and implications for American politics, comparative politics, and international relations.
4. **Synthesize Learning into Applied Political 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 science research context, aiming to achieve a high academic standard.

SOME RECOMMENDED TEXTS

Generalized Linear Models, MLE, and more …

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

Survival Analysis

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

Python Introductory Books

- Fabio Nelli. 2018. Python Data Analytics with Pandas, NumPy, and Matplotlib. Second Edition. APress.
- 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.
- Robert Johansson. 2019. Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib. Second Edition. APress.
- Ashwin Pajankar. 2022. Hands-on Matplotlib Learn Plotting and Visualizations with Python 3. APress.

ADDITIONAL MATERIAL ON CONVAS

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/

SPECIFICS ON THE RESEARCH PAPER

For the instructor to provide guidance in the preparation of the paper, students are required to turn in various brief intermediate papers throughout the semester as follows:

1. Find a topic that interests you and a suitable **research dataset** to analyze using a statistical method that falls within the scope of the material covered in this course. Important proviso: you should be guided by the fact that this is a paper for a methods course and hence the emphasis will be put more on the methods part of the paper, and not on the substantive research question that the paper is pursuing (of course, the two aspects are not mutually exclusive). Submit a report summarizing this step of the paper.
   Due Date: September 16

2. Report on the data and various aspects of it. **Due Date: October 7**

3. Begin developing the research design and hypotheses as well as choosing the right statistical model for that purpose. Students are strongly discouraged from using what is commonly called as *logit* or *probit*. These are too simple to provide enough learning challenges, and hence this would defeat the learning goals of the course through a research paper. Submit a short report on this.
   Due Date: October 28

4. Finalize the research paper focusing mostly on the **methodological aspect** of it without of course neglecting the substantive questions:
   - Show and discuss how you preprocess the data.

- Construct an appropriate model for the estimation and choose the corresponding python package for that purpose.
- Carry out a full analysis considering the assumptions and limitations of the model and the data.
- Draw conclusions on the validity of the model and suggest potential ways to improve your own analysis.

The final paper should be about 15-20 pages long, including the bibliography. **Due Date: December 2.**

5. **Note on the Final Submission of the Paper**:

Students are required to submit to canvas a **zip** folder that contains the paper (written in a professional format suitable for an academic journal as word, pdf, or latex file), an annotated python Notebook file displaying their complete code and analysis that one would need to replicate the analysis of the paper from beginning to end, and the final dataset used for the paper and, if need be, supplementary materials that are deemed important to understand the paper and its analysis.

The instructor is committed to 'walk the walk' with each student in making his/her research paper a potentially publishable piece.

IMPORTANT DATES

Class Begins	**Monday, August 26**
Holidays: No classes	**September 2: Labor Day** **October 18-19: Homecoming** **November 11: Veteran's Day** **November 25 - 30: Thanksgiving Break**
Class Ends	**Monday, December 2**

COURSE OUTLINE:

Generalized Linear Models in Political Science:

1. Introduction and Overview: GLMs & MLE in Political Science Research
    - Introduction to Generalized Linear Models (GLMs) and Maximum Likelihood Estimation (MLE) with examples from political science research.
    - Applications in American politics (e.g., voter turnout), comparative politics (e.g., regime type classification), and international relations (e.g., alliance formation).

2. Deeper into MLE & Model Diagnostics in Political Science
    - In-depth exploration of MLE and diagnostic techniques for model validation.
    - Case studies from political science literature, including election forecasting (American politics), corruption measurement (comparative politics), and conflict prediction (international relations).

3. Binary Models: Logit and Probit in Political Behavior Analysis
    - Analysis of binary outcomes using Logit and Probit models.
    - Practical applications in voter behavior (American politics), democratization likelihood (comparative politics), and treaty ratification (international relations).

4. Ordinal Outcomes: Ordered Logit and Ordered Probit Analysis in Political Surveys
    - Techniques for analyzing ordinal data with Ordered Logit and Ordered Probit models.
    - Examples from public opinion on policy issues (American politics), satisfaction with democracy (comparative politics), and conflict intensity (international relations).

5. Nominal Outcomes: Multinomial Logit and Related Models in Party Choice and Issue Voting
    - Analysis of nominal outcomes using Multinomial Logit and related models.
    - Applications in studying party choice (American politics), coalition formation (comparative politics), and UN voting behavior (international relations).

6. Limited Outcomes: Tobit Model in Political Participation and Campaign Contributions
    - Use of the Tobit model for analyzing limited dependent variables.
    - Practical applications in campaign finance (American politics), political participation (comparative politics), and foreign aid distribution (international relations).

7. Heckman Model and Other Sample Selection Models in Political Science
    - Addressing sample selection bias with the Heckman model and other techniques.
    - Case studies in legislative studies (American politics), election fraud detection (comparative politics), and international trade agreements (international relations).

8. Regression Models for Count Dependent Variables in Legislative Behavior and Conflict Studies
    - Application of regression models for count data.
    - Examples from legislative bill sponsorship (American politics), protest events (comparative politics), and conflict incidence (international relations).

Survival Analysis in Political Science

9. The Logic of Survival Analysis in Political Tenure and Conflict Duration
    - Introduction to survival analysis and its relevance in political science.

- Applications in congressional career longevity (American politics), regime survival (comparative politics), and peace duration (international relations).

10. Parametric Models for Single-Spell Duration Data in Political Events
    - Exploration of parametric models for analyzing single-spell duration data.
    - Practical examples from gubernatorial terms (American politics), cabinet stability (comparative politics), and alliance longevity (international relations).

11. The Cox Proportional Hazards Model in Political Science
    - Application of the Cox Proportional Hazards Model.
    - Case studies in judicial career length (American politics), democratic transitions (comparative politics), and conflict termination (international relations).

12. Diagnostic Methods for Survival Models in Political Research
    - Diagnostic techniques for validating survival models.
    - Practical applications in political science research, including election cycle effects (American politics), regime change (comparative politics), and arms control agreements (international relations).

13. Inclusion of Time-Varying Covariates in Political Survival Analysis
    - Incorporating time-varying covariates into survival analysis.
    - Examples from policy impact over time (American politics), economic performance and government survival (comparative politics), and international sanctions (international relations).

14. Unobserved Heterogeneities in Survival Analysis in Political Contexts
    - Addressing unobserved heterogeneities in survival models.
    - Case studies in mayoral tenure (American politics), autocratic regime durability (comparative politics), and international conflict resolution (international relations).

15. Models for Multiple Events in Political Science
    - Techniques for analyzing multiple event data.
    - Practical applications in repeated electoral contests (American politics), coups and revolutions (comparative politics), and recurring diplomatic negotiations (international relations).

IMPORTANT NOTES:

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at https://gatorevals.aa.ufl.edu/students/. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via https://ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students at https://gatorevals.aa.ufl.edu/public-results/.

UF students are bound by The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students at the University of Florida, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Conduct Code specifies a number of behaviors that are in violation of this code and the possible sanctions. Click here to read the Conduct Code. If you have any questions or concerns, please consult with the instructor or TAs in this class.

Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor.

A "class lecture" is an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a University of Florida course. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.

Publication without permission of the instructor is prohibited. To "publish" means to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 StudentHonor Code and Student Conduct Code.

Campus Resources: Health and Wellness

U Matter, We Care: If you or someone you know is in distress, please contact umatter@ufl.edu, 352-392-1575, or visit U Matter, We Care website to refer or report a concern and a team member will reach out to the student in distress.

Counseling and Wellness Center: Visit the Counseling and Wellness Center website or call 352-392-1575 for information on crisis services as well as non-crisis services.

Student Health Care Center: Call 352-392-1161 for 24/7 information to help you find the care you need, or visit the Student Health Care Center website.

University Police Department: Visit UF Police Department website or call 352-392-1111 (or 9-1-1 for emergencies).

UF Health Shands Emergency Room / Trauma Center: For immediate medical care call 352-733-0111 or go to the emergency room at 1515 SW Archer Road, Gainesville, FL 32608; Visit the UF Health Emergency Room and Trauma Center website.

GatorWell Health Promotion Services: For prevention services focused on optimal wellbeing, including Wellness Coaching for Academic Success, visit the GatorWell website or call 352-273-4450.

In response to COVID-19, the following recommendations are in place to maintain your learning environment, to enhance the safety of our in-classroom interactions, and to further the health and safety of ourselves, our neighbors, and our loved ones.

- If you are not vaccinated, get vaccinated. Vaccines are readily available and have been demonstrated to be safe and effective against the COVID-19 virus. Visit one.uf for screening / testing and vaccination opportunities.
- If you are sick, stay home. Please call your primary care provider if you are ill and need immediate care or the UF Student Health Care Center at 352-392-1161 to be evaluated.
- Course materials will be provided to you with an excused absence, and you will be given a reasonable amount of time to make up work.

Academic Resources

E-learning technical support: Contact the UF Computing Help Desk at 352-392-4357 or via e-mail at helpdesk@ufl.edu.

Career Connections Center: Reitz Union Suite 1300, 352-392-1601. Career assistance and counseling services.

Library Support: Various ways to receive assistance with respect to using the libraries or finding resources.

Teaching Center: Broward Hall, 352-392-2010 or to make an appointment 352- 392-6420. General study skills and tutoring.

Writing Studio: 2215 Turlington Hall, 352-846-1138. Help brainstorming, formatting, and writing papers.

Student Complaints On-Campus: Visit the Student Honor Code and Student Conduct Code webpage for more information.

On-Line Students Complaints: View the Distance Learning Student Complaint Process.