9592: Maximum Likelihood Estimation (Methods 3)

Tuesday 9:30AM - 12:30PM
9/9/25-12/9/25, SSC 7200

Instructor: Dave Armstrong
Office: SSC 7310
Hours: 11AM-12PM Wednesday or by appointment
E-mail: dave.armstrong@uwo.ca

This course is designed to get you thinking beyond linear models to models for other kinds of dependent variables - counties, categories, durations and more. You will also be asked to apply these new skills to answer social scientific questions. An ancillary goal of the course is to enable you to be critical consumers of quantitative work. Often times, these two goals go hand-in-hand and as you get more comfortable with doing statistical analyses and you learn how to understand, implement and interpret more and more complicated methods, you almost necessarily become more savvy consumers of quantitative work.

The course will be taught from a more applied, rather than mathematical, point of view. While there will certainly be some math in the course, I will approach these concepts assuming no particular comfort with mathematics aside from basic arithmetic (and even then, we will go over, important mathematical results and formulae). If you are having trouble with the concepts, please ask questions in class. Others will likely be confused by some of the same issues. If you continue to have trouble in the course, please see me during office hours or make an appointment and we can clear up any problems as they arise. Statistics is a cumulative enterprise, so a lack of understanding early can compound itself as the course moves forward.

As you are all graduate students, I expect that you will attend class regularly, do the readings and ask questions when something is confusing. You are ultimately responsible for knowing the material. I will do my best to teach it in a way that is accessible and coherent, but if you do not understand something, you need to take responsibility for figuring it out by asking questions, either in or outside of class. If you miss class, you are responsible for learning the material you missed in a manner that proves least distracting for the other participants in the course. Also, the late work policy is that late papers are not accepted (rare exceptions may be allowed on a case-by-case basis when arrangements are made before the due date).

Computing

This class has both theoretical and applied aspects. Thus, you will not only learn the reasoning behind statistical analysis, you will also learn how to analyze data yourselves using **R**. R is open-source and as such, is freely available. We will try to put aside an hour of each class to work on data examples.

When you have work using the computer that needs to be turned in, it should be done in RMarkdown and the RMarkdown file should be submitted along with any data to which I do not otherwise have access. I will provide templates for homework in an effort to unify the visual aesthetic. For those of you who are unfamiliar with RMarkdown, there is a good tutorial here here

Grading

You final grade in the course will depend on the following:

Homework	50%
Final Paper	50%

Homework

You will get assignments on each topic of various lengths and types - they will all be applied and should be done with the computer. You should consider your colleagues a resource and I encourage you to discuss the problem sets with them. That said, each person must turn his/her/their own, original answers to the homework problems - the answers should be in your words.

Final Paper

The class culminates in a final paper that will be written in the form of published papers you have read. It will need to have all relevant parts of a paper - introduction, (short) literature review, theory/hypotheses, data/methods, results and conclusion. You should use one of the methods we discuss in class. There is no formal requirement for length, but let me suggest a couple of things. First, your literature review shouldn't be more than three or four pages. You should present your theory - the way you think the conceptual pieces fit together. This discussion should include a formal presentation of hypotheses. You should describe the data you're using - where you got it, what you did to it and how you think the variables you are using match the concepts they are meant to measure. You should talk about the procedure you use for testing the hypotheses along with the strengths and possible weaknesses of said procedure. You need to discuss the results and then conclude by putting the results back in context and highlighting the most important results.

Textbook

There is no single textbook used for this course. The required readings for this course will be posted on OWL. I detail these in the outline below.

AI Policy

Large Language Models, like the ones used by ChatGPT and Copilot are powerful tools, especially for coding. I use them from time to time to do tedious tasks or generate some template code that I can then reconfigure and extend. That said, there are two disadvantages to using LLMs to do all your coding.

1. The obvious one - you won't really learn how R works. You will say something like, "I will study the code and learn from it" but you won't, at least not for very long. You will eventually end up just cutting and pasting the code and letting ChatGPT do all the debugging for you. Without a good understanding of how R works, you won't be able to do much in the situation where ChatGPT does not give you something useful and you may not know the kinds of questions to ask to make it be helpful.

2. Maybe also obvious now - LLMs are not perfect. They make mistakes and sometimes the mistakes do not turn up as errors in the code that are easily debugged. They may also make choices about the statistical models you are running that are not optimal.

For these reasons, I am allowing you to use LLMs to help you with coding, but I am requiring that you disclose when you do so. If you use an LLM to help you with coding, please indicate so in your assignment and include a transcript of the conversation you had with the LLM. If you do not do this, I will assume that all the code is your own work and if it is not, I will consider it a violation of academic integrity. If you have any questions about this policy, please ask me.

Miscellaneous

Information on academic integrity and penatiles for violation thereof can be found here.

Outline

1. Maximum Likelihood Estimation I: Likelihood and Simple Models (9/9)
    - King "Unifying Political Methodology" (pp. 59-67)
    - Fox "Applied Regression" (Appendix, D.6.1-D.6.2)
2. Binary DV Models (9/16)

- Andersen and Armstrong (2021) *Presenting Statistical Results Effectively.* Chapters 10 and 11 (§11.1-11.3).
- Faraway "Extending the Linear Model with R" (Chapter 2)
- Application: Fred Cutler "Political Conditions for Electoral Accountability in Frederalism" (*CJPS*, 2017)

3. Understanding Interactions in Binary DV Models (9/23)

   - Rainey, Carlisle. (2016). "Compression and Conditional Effects: A Product Term Is Essential When Using Logistic Regression to Test for Interaction." *Political Science Research and Methods*, 4(3), 621-639. doi:10.1017/psrm.2015.59.
   - Application: Jaroslav Tir and Doug Stinnett "Weathering climate change: Can institutions mitigate international water conflict?" (*JPR*, 2012)

4. Class Cancelled - National Day for Truth and Reconciliation (9/30)

5. Model Fit and Diagnostics (10/7)

   - Robert Andersen and David Armstrong (2021) *Presenting Statistical Results Effectively* Chapter 11, §11.4-11.5
   - Scott Long (1997) *Regression Models for Categorical and Limited Dependent Variables* Thousand Oaks, CA: Sage. Chapter 4.
   - Kenneth Burnham and David Anderson (2004) "Understanding AIC and BIC in Model Selection" *Sociological Methods and Research* 33(2): 261-304.
   - James, Gareth et. al. (2021) *Introduction to Statistical Learning with R* Chapter 5 §5.1 (Cross Validation).

6. Ordinal Regression Models (10/14)

   - Robert Andersen and David Armstrong (2021) *Presenting Statistical Results Effectively* Chapter 11, §11.6.1-11.6.3
   - Faraway "Extending the Linear Model with R" Chapter 7 §7.4
   - William Clark and Matt Golder (2021) *Introduction to Interaction Models* Chapter 9.
   - Application: Jennifer Merolla, Laura Stephenson and Elizabeth Zechmeister "Can Canadians Take a Hint? The (In)Effectiveness of Party Labels as Information Shortcuts in Canada" (*CJPS*, 2008)

7. Multinomial Regression Models (10/21)

   - Robert Andersen and David Armstrong (2021) *Presenting Statistical Results Effectively* Chapter 11, §11.6.4-11.6.5

- Faraway "Extending the Linear Model with R" Chapter 7 §7.1
- William Clark and Matt Golder (2021) *Introduction to Interaction Models* Chapter 10.
- Application: Cameron Anderson and Laura Stephenson "Environmentalism and Party Support in Canada: Recent Trends outside Quebec" (*CJPS*, 2011)

8. MLE IV: Count Models (10/28)

   - Faraway "Extending the Linear Model with R" (Chapter 5)
   - Application: Paul Thomas, Peter Loewen and Michael MacKenzie, "Fair Isn't Always Equal: Constituency Population and the Quality of Representation in Canada" (*CJPS*, 2013)

9. Fall Reading Week (11/4)

10. Generalized Additive Models of Location, Scale and Shape (11/11)

    - James, Gareth, Daniela Witten, Trevor Hastie and Robert Tibshirani (2023) *An Introduction to Statistical Learning with Applications in R,* 2^{nd} *ed.*, Chapter 7
    - Stasinopoulos, D. M., and Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. Journal of Statistical Software, 23(7), 1–46. https://doi.org/10.18637/jss.v023.i07

11. Hiearchcal Models (11/18)

    - Andrew Gelman and Jennifer Hill (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models* Chapters 11-14
    - Bell and Jones "Explaining Fixed Effects: Random Effects Modeling of Time-Series Corss-Sectional and Panel Data" (*PSRM*, 2014)

12. Missing Data and Multiple Imputation (11/25)

    - Stef Van Buuren (2018) *Flexible Imputation of Missing Data* Chapters 1-5.

13. Measurement and Scaling (12/2)

    - Paul Spector (1992) *Summated Rating Scale Construction* Chapters 1-5.
    - David Bartholomew et. al. (2008) *Analysis of Multivariate Social Science Data* Chapters 7-8.

14. Wrap up and Research Presentations (12/9)