POLS/CSSS 510:
Maximum Likelihood Methods for the Social Sciences

Christopher Adolph · Professor · Political Science and CSSS

University of Washington
Fall Quarter 2024

Class Meets	Office
MW 4:30–5:50 PM	Gowen Hall 145
Smith Hall 205	cadolph@uw.edu
Section Meets	**Teaching Assistant**
F 3:30–5:20 PM	Ramses Llobet
Taught via Zoom	rllobet@uw.edu

Overview and Class Goals. Most social science data do not neatly fit the assumptions of the linear regression model taught in introductory statistics courses. For example, social science data often consist of discrete categorizations or counts of events, rather than continuous outcomes. Observations may be correlated across periods, as in time series, or clustered into correlated groups, violating the linear regression assumption of independence. In this course, students will learn how use the method of maximum likelihood to derive statistical models that suit the particular behavior of their social science data and how to clearly communicate the substantive import of their findings to a broad audience, with special attention to visualization of both results and model fit. In the process, students will gain familiarity with basic statistical programming in R, a free and popular language. Topics of special interest to students will be covered as time permits.

Prerequisites. No specific courses are required; however, students should have a solid grounding in linear regression, as provided, for example, by courses in Political Science (POLS 501 and POLS 503), Sociology (SOC 504 and SOC 505), or Statistics (STAT/CSSS 504). Familiarity with or a willingness to quickly learn basic probability and matrix algebra is also required.

Course Requirements. Course evaluation will be based on (1) problem sets: five over the term, due by the start of class on specified dates, (2) a student poster presentation in class, also to be emailed to your instructor as a PDF, and (3) a research paper, due Tuesday 10 December 2024 at 3 PM by email in PDF format.

Students *may* work in small groups on problem sets, but each student must complete and submit their own write-up of the assigment through Canvas by the start of class on the day the assignment is due.

Students are encouraged (but not required) to collaborate in groups of two or threeon the poster presentation and paper, which should apply methods studied in the course (or with instructor approval, related methods of similar sophistication) to a student-chosen topic. Papers developing new methodological tools are also acceptable. Each paper writer or paper-writing group must propose a paper topic as part of the first assigned homework, so students should plan to form research partnerships quickly. Additional paper requirements and guidelines can be found on the course website.

Office Hours. Chris Adolph: By appointment via Zoom. Ramses Llobet: by appointment via Zoom.

Course Website. Consult `http://faculty.washington.edu/cadolph/mle` for problem sets, notes, and announcements.

Use of Generative Text and Images Prohibited. Students are prohibited from using generative text or generative images – so called-artificial intelligence tools such as ChatGPT or MidJourney – to assist in completing any course assigments. Students should not use chatbots based on large language models to complete class assignments because of the fundamental challenges these tools have in generating accurate statements ("hallucination") and the intrinsic inability of these tools to properly attribute sources of information. Moreover, a strong ethical case can be made against the use of either generative text or images in academic work due to the unauthorized use of copyrighted materials to train the models underlying these tools. Regardless of the merits

of these tools, reliance on them in an instructional environment deprives students of the opportunity to hone the research, writing, and coding skills required to evaluate or refine their outputs. Even if there is a case for using chatbots or AI art in some contexts, doing so in this course contradicts its core pedagogical aims.

Students may not use chatbots or AI art to produce, in whole or in part, either rough or final drafts of computer code, figures, assignment write-ups, presentations, or papers: use of chatbots to assist in any of these tasks will be considered cheating and/or academic fraud. With prior instructor approval, limited exceptions may be made *only* for the use of AI to process text or image data into usable machine-readable formats; in such cases, students should be mindful of ethical considerations in using such tools, practical considerations regarding the reliability of data processed using generative AI, and appropriate techniques for mitigating bias and hallucination. *If you are uncertain whether use of a specific resource violates these guidelines, ask your instructor before using it.*

Penalty for Cheating or Academic Fraud. Any student caught cheating or plagiarizing by the instructor on any assignment will receive a grade of **X** for the course and will be reported to the Dean's office in the College of Arts and Sciences.

Notice Required by State Law. *Washington state law requires that UW develop a policy for accommodation of student absences or significant hardship due to reasons of faith or conscience, or for organized religious activities. The UW's policy, including more information about how to request an accommodation, is available at Religious Accommodations Policy* (`https://registrar.washington.edu/staffandfaculty/religious-accommodations-policy`). *Accommodations must be requested within the first two weeks of this course using the Religious Accommodations Request form* (`https://registrar.washington.edu/students/religious-accommodations-request`).

Other relevant university policies. See this website:
`https://registrar.washington.edu/staffandfaculty/syllabi-guidelines`

Course textbooks

Books recommended for purchase marked with a ⋆.

⋆Andrew Gelman, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories*. Cambridge University Press. ISBN: 978-1107676510. Amazon: $41.47 (paperback)	*Excellent introduction to statistical inference, post-estimation simulation, and the regression models covered in our course, with great advice for best practices.*
Gary King. 1989. *Unifying political methodology*. University of Michigan Press. ISBN: 978-0472085545. Amazon: $29.95 (paperback)	*Classic introduction to maximum likelihood for social scientists. Emphasis on intuition and basic deriviations.*
J. Scott Long. 1997. *Regression models for categorical and limited dependent variables*. Sage Publications. ISBN: 978-0803973749.	*Clearest step-by-step introduction to the logic of logistic regression and related models, but expensive and scarce.*
⋆Michael D. Ward and John S. Ahlquist. 2018. *Maximum Likelihood for Social Science: Strategies for Analysis*. Cambridge University Press. ISBN: 978-1316636824. Amazon: $34.94 (paperback)	*Main text for the course; written by a former UW professor & a POLS/CSSS alumnus.*

Optional books for further study

A. Colin Cameron and Pravin K. Trivedi. 2013. *Regression analysis of count data*. 2nd. Ed. Cambridge University Press.

Andrew Gelman and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Norman Matloff. 2011. *The Art of R Programming: A Tour of Statistical Software Design*. No Starch Press.

Richard McElreath. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2nd. Ed. Chapman & Hall/CRC Press.

Will H. Moore and David A. Siegel. 2013. *A Mathematics Course for Political & Social Research*. Princeton University Press.

Required and optional articles

R. Michael Alvarez and Jonathan Nagler. 1998. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science*. 42(1) 55–96.

William D. Berry, Jacqueline H. R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science*. 54(1) 243–266.

Kosuke Imai and David A. van Dyk. 2005. "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation." *Journal of Econometrics*. 124: 311–334.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Interpretation and Presentation" *American Journal of Political Science* 44(2): 341–355.

Raftery, Adrian, 1995. "Bayesian Model Selection in Social Research (with Discussion)." *Sociological Methodology*. 111-196.

Brian Greenhill, Michael D. Ward, and Audrey Sacks. 2011. "The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models." *American Journal of Political Science*. 55(4): 990–1002.

Course outline

This outline of topics is a guideline and may be altered to meet course needs. In particular, the pace of the course may vary to make sure we are moving as fast as possible conditional on everyone understanding the material. Students should come to class having read the material for the next topic to be covered.

Readings from Ward & Ahlquist, King, and Long may at times overlap somewhat; feel free to focus on the sources that best fit your learning style, while bearing in mind that Ward & Ahlquist is the most up-to-date treatment.

Numbers listed after texts indicate chapters, unless otherwise noted.

Part I: Fundamentals

25 September – 2 October: Introduction to the course, probability, and R

Required: Handout on probability
Handout on matrix algebra
Gelman, Hill & Vehtari, 1.1–1.5, 4
Ward & Ahlquist, p. *xxv-xxvii*

Optional: Gelman, Hill & Vehtari, Appendix A
King 3

7–14 October: Overview of maximum likelihood estimation

Required: Ward & Ahlquist 1, 2, 4

Optional: Long 1, 2
King 1, 2, 4

PROBLEM SET 1 + PROJECT PROPOSAL DUE WED 9 OCTOBER BY START OF CLASS

Part II: Modeling Choice

16–21 October: Estimating & Interpreting models of binary data

Required: Ward & Ahlquist 3 and 6
King 5.1–5.3
Gelman, Hill & Vehtari 6.1–6.3, 7, 9.1–9.2
King, Tomz, and Wittenberg

Optional: Gelman, Hill & Vehtari 5
Long 3.1–3.9
Berry, DeMeritt, and Esarey

PROBLEM SET 2 DUE MONDAY 21 OCTOBER BY START OF CLASS

23–28 October: Fitting & selecting models of binary data

Required: Ward & Ahlquist 5
Long 4
Gelman, Hill & Vehtari 13, 14

Optional: Greenhill, Ward, and Sacks
Raftery

30 October–4 November: Modeling ordered categorical data

Required: Ward & Ahlquist 8
Long 5.1–5.7

Optional: King 5.4

PROBLEM SET 3 DUE WEDNESDAY 6 NOVEMBER BY START OF CLASS

6 November: Non-ordered data: Multinomial logit & Multinomal probit

Required: Ward & Ahlquist 9

Optional: Long 6
Gelman, Hill & Vehtari 15.4–15.5
Alvarez and Nagler
Imai and van Dyk

NO CLASS MONDAY 11 NOVEMBER — VETERANS DAY

Part III: Modeling Counts

13–20 November: Poisson models of counts / Overdispersed & zero-inflated counts

Required: Ward & Ahlquist 7, 10
King 5.5–5.9
Gelman, Hill & Vehtari 15.1–15.3, 15.8

Optional: Long 8.1–8.7

PROBLEM SET 4 DUE WEDNESDAY 20 NOVEMBER BY START OF CLASS

Part IV: Advanced Topics

25–27 November: Missing data and multiple imputation

Required: Ward & Ahlquist 12
Gelman, Hill & Vehtari 17.3–17.6

PROBLEM SET 5 DUE MONDAY 2 DECEMBER BY START OF CLASS

Part V: Student Poster Presentations

2–4 December: Student poster presentations

Students will prepare and present a poster on their research projects in progress; this yields valuable feedback prior to final write-ups. Requirements and suggestions for poster construction will be discussed in class. Presentation dates will be assigned to minimize discrepancy with student preferences. Early presentations are expected to be less complete and evaluated accordingly. Students unable to present on certain days due to schedule conflicts should inform the instructor.

FINAL PAPER DUE TUESDAY 10 DECEMBER AT 3 PM BY EMAIL