

**Categorical Data Analysis:
Models for Categorical, Ordinal, and Count Outcomes
ICPSR SUMMER PROGRAM
JULY 13 – JULY 31, 2026**

Instructor: Dr. Mark Nieman (mark.nieman@utoronto.ca)

Time and Location: 9am–12pm, TBA

Office Hours: by appointment

Teaching Assistant: TBA **TA Office Hours:** TBA

This class introduces a variety of statistical techniques for limited and categorical dependent variables relevant to political science research. The objective is for you to become familiar with these techniques to understand how, when, and why to use them. We therefore emphasize empirical applications, and a large portion of class time (approximately 40% of classes) is devoted to hands-on use and interpretation of these methods on computers. We recommend using the statistical programs R, since all sample code for assignments are written in this programming language. We assume some familiarity with the concepts of maximum likelihood, linear algebra, calculus, and probability theory, but we may review these topics as necessary.

The course covers a wide variety of estimators, including those for binary, ordered, polychotomous, and multivariate outcome variables. We also discuss issues relating to truncation, censoring, and non-random sample selection. Additional and related topics will be covered as necessary. The main tools through which you will familiarize yourself with these methods are Maximum Likelihood Estimation and Monte Carlo analysis, which will be presented in the first and second weeks. Many of your homework assignments and in-class work will involve these two techniques.

Requirements

Grades are based on three homework assignments.

The best way to learn the material is to use the estimators. We will assign homework each week. Some homework questions will specify a model and ask you to run a Monte Carlo analysis that involves generating data and then estimating parameters using a few different assumptions (both correct and incorrect). When you turn in the homework, we want you to upload an electronic copy of your html document (knit from an RMarkdown document) and a RMarkdown file (just in case), to Canvas.

Required Texts

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. USA: Sage Publications.

Recommended:

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor, MI: University of Michigan Press.

Course Outline

Week 1

- Monday—Welcome and introduction; Linear Probability Model
 - Long, Ch 1–2.
- Tuesday—Logistic Regression; Transformational and latent variable approach for binary outcomes; Maximum likelihood estimation; Odds ratios.
 - Long, Ch 3.
- Wednesday—Probit regression; Estimation of BRM; Interpretation.
- Thursday—Logit/Probit model fit; Compression; Separation.
 - Long, Ch 4.
 - Berry, William D., Jacqueline HR 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): 248–266.
 - Zorn, Christopher. 2005. A Solution to Separation in Binary Response Models. *Political Analysis* 13(2): 157–170.
- Friday—Ordinal variables; a latent variable model.
 - Long, Ch 5.

Week 2

- Monday—Estimation of ORM; Parallel Regression Assumption; Interpretation.
- Tuesday—Multinomial logit as a set of BLMs; IIA; Interpretation
 - Long, Ch 6.
 - Alvarez, R. Michael and Jonathan Nagler. 1998. When Politics and Models Collide: Estimating Models of Multiparty Elections. *American Journal of Political Science* 42(1): 55–96.

-
- Wednesday—Count dependent variable; Poisson regression; Negative binomial model; Estimation and Interpretation; Model fit.
 - Long, Ch 8, Sections 1–2 (Poisson), Sections 3–7 (negative binomial).
 - King, Gary, 1988. Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model. *American Journal of Political Science* 32: 838–863.
 - Thursday—Temporal Dependence; Duration analysis; Temporal splines and polynomials with binary outcomes.
 - Box-Steffensmeier, Janet M., and Bradford S. Jones. 1997. Time is of the Essence: Event History Models in Political Science. *American Journal of Political Science* 41(4): 1414–1461.
 - Carter, David B. and Curtis S. Signorino. 2010. Back to the Future: Modeling Time Dependence in Binary Data. *Political Analysis* 18(3): 271–292.
 - Williams, Laron K. 2018. Temporal Dependence and the Sensitivity of Quantities of Interest: A Solution for a Common Problem. *International Studies Quarterly* 62(4): 892–902.
 - Friday—Non-Constant Variance with Latent Variables.
 - Williams, Richard. 2009. Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups. *Sociological Methods & Research* 37(4): 531–559.
 - Arena, Philip and Glenn Palmer. 2009. Politics or the Economy? Domestic Correlates of Dispute Involvement in Developed Democracies. *International Studies Quarterly* 53(4): 955–975.

Week 3

- Monday—Bivariate and Multivariate Probit.
 - Greene, William. 1993. Bivariate and Multivariate Probit. *Econometric Analysis*. Upper Saddle River, NJ: Prentice-Hall, 2nd Ed. Chapter 21, Section 6.
- Tuesday—Extensions of the Bivariate Probit.
 - Kimball, Anessa L. 2006. Alliance Formation and Conflict Initiation: The Missing Link. *Journal of Peace Research* 43(4): 371–389.
 - Xiang, Jun. 2010. Relevance as a Latent Variable in Dyadic Analysis of Conflict. *Journal of Politics* 72(2): 484–498.
 - Reed, William. 2000. A Unified Statistical Model of Conflict Onset and Escalation. *American Journal of Political Science* 44(1): 84–93.

-
- Wednesday–Censoring/Truncation and Sample Selection
 - Long, Ch 7
 - Hug, Simon. 2003. Selection Bias in Comparative Research: The Case of Incomplete Data Sets. *Political Analysis* 11(3): 255–274.
 - Thursday–Strategic Models.
 - Signorino, Curtis S. 1999. Strategic Interaction and the Statistical Analysis of International Conflict. *American Political Science Review* 93(2): 279–297.
 - Signorino, Curtis S. 2003. Structure and Uncertainty in Discrete Choice Models. *Political Analysis* 11(4): 316–344.
 - Kenkel, Brandon and Curtis S. Signorino. 2014. Estimating Extensive Form Games. *R. Journal of Statistical Software* 56(8): 1–27.
 - Nieman, Mark David. 2018. Strategic Binary Choice Models with Partial Observability. *Statistical Sinica* 28(4): 2089–2105.