

POLI 7964:
Advanced Topics in Research Methods:
Categorical Data Analysis

Fall 2014

James C. Garand
Emogine Pliner Distinguished Professor

Department of Political Science
Office: Stubbs 205
Phone: 578-2548
Email: pogara@lsu.edu

Introduction

The regression model is at the core of many statistical analyses of continuous dependent variables in the social sciences. But what happens when researchers estimate a model in which the dependent variable is not continuous but is, rather, categorical or discrete?

This is where categorical data analysis comes into the picture. In this seminar we will focus our attention on regression models that have *categorical* or *discrete* dependent variables. The major problem with models that have categorical or discrete dependent variables is that they often require nonlinear specifications or seemingly-complicated transformations to avoid violating the assumptions associated with ordinary least squares (OLS) regression. Non-linearity complicates statistical analyses, but fortunately the nonlinear procedures used to estimate models with categorical dependent variables can be easily handled by available statistical software. This seminar is intended as an overview of the major statistical techniques associated with categorical or discrete dependent variables.

The seminar will proceed as follows. First, we begin with a brief refresher on regression analysis and other statistical issues that are relevant to this course. Second, we will discuss the simplest kind of analysis of categorical dependent variables—i.e., cross-tabulation analysis. Third, we will discuss the various statistical techniques in the logit and probit families for dealing with binary, ordinal, and nominal (i.e., nonorderable discrete) dependent variables. In this section we will introduce maximum likelihood estimation (MLE) as an alternative to the least-squares approach for model estimation. Fourth, we will explore estimation procedures for count models—i.e., those with a dependent variable comprised of count data, which tells the researcher how many times something has happened. Here we will discuss Poisson and negative binomial regression models, both of which are appropriate for count data. Finally, we will consider estimation strategies for dependent variables that are censored (i.e., they take on a value only up to or down to a specific threshold).

Two major points should be made about this course. First, the focus of this course will be on the application of categorical data analysis to answer questions of interest to social scientists. To be sure, there will be some discussion of the mathematical underpinnings of the various techniques for categorical data analysis, but for the most part the focus will be on how social scientists (1) determine which statistical technique is appropriate for a given data set and (2) use these statistical techniques to model social

phenomena measured using categorical or discrete variables. Second, because most of the statistical techniques used for categorical or discrete dependent variables generate output that is often not interpreted intuitively, we will spend quite a bit of time on the presentation of results in a way that facilitates interpretation. This includes the use of well-designed tables and figures to demonstrate how specific independent variables affect a given categorical or discrete dependent variable.

Prerequisites

All students enrolled in this course should have completed an introductory statistics class such as POLI 7962 (or its equivalent), and all students should have completed (or at least be in the process of taking) a seminar in regression analysis such as POLI 7963 (or its equivalent). We will conduct a *brief* review at the beginning of the course on regression analysis, but students should be generally familiar with such ideas as descriptive statistics, sampling distributions, statistical inference, hypothesis testing, and basic regression analysis.

Course Requirements and Evaluation

Each student will be evaluated on the basis of the following:

Midterm Examination	30%
Final Examination	30%
Research Paper	40%

1. **Midterm Examination.** A comprehensive midterm examination covering all material from the first half of the course will be required for all students. Information pertaining to the format of the midterm examination will be provided to all students prior to the exam.
2. **Final Examination.** A comprehensive final examination covering all material from the second half of the course will be required for all students. Information pertaining to the format of the final examination will be provided to all students prior to the exam.
3. **Research Paper.** In addition to the midterm and final exams, each student will be required to write a research paper that utilizes one of the statistical techniques discussed during the semester. More detailed information about the paper will be provided later in the semester.

Required Reading

There is one required text for this course:

J. Scott Long and Jeremy Freese

Regression Models for Categorical Dependent Variables Using Stata (2th edition)

Stata Press

ISBN: 1-59718-011-4

In addition, there will be some additional required reading during the course. The required reading assignments will be made available in electronic format through Moodle or via email.

Computer Statistics Package

There are numerous statistics programs that can be used to conduct statistical analysis with a mainframe or personal computer. In this course we will use Stata 13.0, a commonly-used (and easy-to-use) statistics program. I should also note that if you have a copy of Stata 11.0 or Stata 12.0 that will work for most of what we are going to do.

If you are serious about doing quantitative research, I would encourage you to purchase a copy of Stata. Other programs (such as SPSS or SAS) are fine, but Stata is easy to learn and use, and I find that it is much more powerful and flexible for most applications than other programs. This is particularly the case for categorical data analysis, which will be the focus of this class. Stata will be the only program supported during this class, and computer assignments will be conducted using Stata.

Stata is expensive, but fortunately the Stata Corporation provides educational discounts. Stata can be purchased at the following web site:

<http://www.stata.com/order/new/edu/gradplans/campus-gradplan/>

Notice that there are multiple options. First, at the very least you should purchase Intercooled Stata 13.0 with either the six-month license (\$69) or a one-year license (\$98); this will give you access to Stata on your personal computer on a temporary basis. Second, a mid-range (and recommended) option is Intercooled Stata 13.0 with a perpetual license (\$189). Finally, if you (1) will be doing a lot of statistical analysis in your research in the future, particularly with larger data sets, and (2) can at all possibly afford it, I encourage you to purchase Stata / SE 13.0, which is the most powerful version of Stata but is somewhat pricey (\$395). Do not purchase the Small Stata 13.0, which is really for very small data sets and will not accommodate some of the data sets that we will use this semester.

Clarify and SPOST

There are two add-ons to Stata that we will use during the course of the semester. Both of these programs are free and can be downloaded and easily integrated into Stata. You should download and install these programs as soon as you have installed Stata on your computer.

Clarify. First, Gary King and collaborators have developed a program called Clarify, which is used by researchers to estimate a regression model and then generate predicted values on the dependent

variable, all-the-while manipulating the effects of some independent variables and holding the effects of others constant. Clarify is relatively easy to use and can generate results that can be presented in tables or figures. You can obtain a copy of clarify in one of two ways.

The easiest way is to type in the following command in Stata 12.0:

```
findit clarify
```

You will be directed to another page, where you will see the following link:

clarify from <http://gking.harvard.edu/clarify>

If you click on this link it will take you to a page where you can install Clarify directly in Stata. Click on:

(click here to install)

and follow the prompts. This will install Clarify in Stata.

Alternatively, you can download Clarify from the following link:

<http://gking.harvard.edu/clarify>

Once you have downloaded this program as a compressed file, you should place it temporarily in a folder on your hard drive; I would suggest that you create a new folder in your c:\ directory, such as c:\clarify. Once the zipped file is in this directory, you can extract the files from the compressed file, and then these files can be installed in Stata using the following commands:

```
net from c:\clarify (or whatever directory you placed the uncompressed files)
```

```
net install clarify
```

SPOST. Second, one of the authors of your text, J. Scott Long, is a foremost authority on the use of categorical data analysis in the social sciences. Long has developed an add-on program for Stata, called SPOST, that does many of the same things that Clarify does, as well as some other things that Clarify does not do. We will use SPOST during the course of the semester. In order to download SPOST, type the following command in Stata:

```
findit spost
```

This will open a viewer screen in Stata with a variety of links. Scroll down and click on the following link:

spost9_ado from <http://www.indiana.edu/~jslsoc/stata>

Note that this link has “ado” in the address. Stata will prompt you to install the files. Click on “Click here to install” and follow the directions. You will now have access to SPOST commands.

Stata Resources

Because many of you have not had experience with Stata, there are several Stata resources that I recommend. First, I will make available to you a .pdf introduction to Stata:

Tobias Pfaff, *A Brief Introduction to Stata with 50+ Basic Commands*

Even though this manual is designed for Stata 10.0 rather than Stata 12.0, the differences in the two versions are sufficiently small that Pfaff's manual will be very helpful to you.

There is also a lengthy (but informative) introduction to Stata that is a Powerpoint presentation converted to a .pdf format:

Christopher Baum, *Introduction to Stata*

In addition, I would like to direct you to the following web sites that are designed to assist Stata users. These are very useful web sites that provide detailed information about Stata commands. You should bookmark these web sites and refer to them often.

The first is a broad-based Stata web site housed at UCLA. One can find a wide range of information about Stata on this web site. I would encourage you to take a look at the links on this web site and familiarize yourself with what this site has to offer:

<http://www.ats.ucla.edu/stat/stata/>

This is the Stata "starter kit" for new users, also available on the UCLA web site:

<http://www.ats.ucla.edu/stat/stata/sk/>

Here is another UCLA site that has basic "learning models" for Stata procedures.

<http://www.ats.ucla.edu/stat/stata/modules/default.htm>

Another introductory web site is found on the University of North Carolina web site:

http://www.cpc.unc.edu/research/tools/data_analysis/statatutorial

Here is a web page at Princeton that provides a basic introduction to Stata:

<http://data.princeton.edu/stata/default.html>

This UCLA site has information on estimating models with categorical dependent variables in Stata. This is a very good resource for information about estimating such models:

http://www.ats.ucla.edu/stat/stata/topics/logistic_regression.htm

Finally, this is a Princeton web site that has a series of downloadable chapters about how to estimate a variety of models in Stata, including models with categorical dependent variables.

<http://data.princeton.edu/wws509/stata/>

Calculator

In addition, many of the statistical problems to be worked out during the semester require the use of a calculator with a square root function. Each student is strongly encouraged either to purchase such a calculator or otherwise to have one available. A simple calculator with a square root function can be purchased for well under \$10.

Graduate Assistant

There is no graduate assistant for this course.

Office Hours

Friday, 9:00 A.M. – 11:00 A.M.
Other hours by appointment

Academic Misconduct Statement

Academic misconduct is defined by the Code of Student Conduct. You are encouraged to familiarize yourself with the LSU policy on academic misconduct, particularly regarding plagiarism. The LSU Code of Student Conduct can be found on the web site for the LSU Dean of Students:

http://saa.lsu.edu/sites/saa.lsu.edu/files/attachments/Code%20of%20Student%20Conduct%20August%2009_0.pdf

Academic misconduct is a serious violation of university policy, but more importantly it is a significant scholarly violation for political scientists. Plagiarism and other forms of academic misconduct will not be tolerated in this course. Charges of academic misconduct will be turned over to the Dean of Students for appropriate disciplinary action.

I am serious about this.

Suggestions for Further Reading

In addition, the following are useful reference or supplemental readings. These works should be available in the Middleton Library, through J-STOR, or for purchase from Amazon.com or some other online book seller.

Logit and Probit Models

J. Scott Long, *Regression Models for Categorical and Limited Dependent Variables*.

John Aldrich and Forrest Nelson, *Linear Probability, Logit, and Probit Models*.

Scott Menard, *Applied Logistic Regression Analysis*.

Tim Liao, *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*.

Arthur DeMaris, *Logit Modelling: Practical Applications*.

Alan Agresti, *An Introduction to Categorical Data Analysis* (2nd edition).

Class Outline:

The following is a tentative outline of topics for the semester. The instructor reserves the right to make adjustments in the schedule of topics as necessary and with reasonable advance notice.

August	25	Introduction
September	1	Labor Day Holiday (no class meeting)
September	8	Statistics review J. Scott Long and Jeremy Freese, <i>Regression Models for Categorical Dependent Variables Using Stata</i> (2 nd edition), Chapters 1-2.
September	15	Analysis of cross-tabulations Required: David Rohde, "Risk-Bearing and Progressive Ambition: The Case for Members of the United States House of Representatives," <i>American Journal of Political Science</i> (February 1979: 1-26). M. Kent Jennings and Gregory Markus, "Partisan Orientations over the Long Haul: Results from the Three-Wave Political Socialization Panel Study," <i>American Political Science Review</i> (December 1984: 1000-1018). John Sides, "Are Racists in Only One Political Party?" <i>The Monkey Cage</i> , August 19, 2012. Downloaded from: http://themonkeycage.org/blog/2012/08/19/are-racists-only-in-one-political-party/
September	22	Analysis of cross-tabulations (continued)
September	29	Dichotomous dependent variables: Binary logit Required: J. Scott Long and Jeremy Freese, <i>Regression Models for Categorical Dependent Variables Using Stata</i> (2 nd edition), Chapters 3-4. Elizabeth Corey and James C. Garand, "Are Government Employees More Likely to Vote? An Analysis of Turnout in the 1996 U.S. National Election," <i>Public Choice</i> (April 2002: 259-83). James C. Garand and T. Wayne Parent, "Representation, Swing, and Bias in U.S. Presidential Elections, 1872-1988," <i>American Journal of Political Science</i> (November 1991: 1009-29). Bruno Frey, David Savage, and Benno Torgler, "Behavior Under Extreme Conditions: The Titanic Disaster," <i>Journal of Economic Perspectives</i> (Winter 2011: 209-22). Garand and Evans, "Votar o No Votar? Understanding Voter Turnout in Latin American Countries," unpublished paper, 2013.

Class Outline (continued):

- October 6 Dichotomous dependent variables: Binary logit (continued)
- October 13 Dichotomous dependent variables: Binary logit (continued)
- October 20 Ordinal dependent variables: Ordered logit

Required:

J. Scott Long and Jeremy Freese, *Regression Models for Categorical Dependent Variables Using Stata* (2nd edition), Chapters 5.

N. Susan Gaines and James C. Garand, "Morality, Locality, or Equality: Analyzing Determinants in Support for Same-Sex Marriage," *Political Research Quarterly* (September 2010: 553-67).

Shao, Garand, Keim, and Hamilton, "Science, Scientists, and Local Weather: Understanding Mass Perceptions of Global Warming," unpublished paper, 2013.

- October 27 Ordinal dependent variables: Ordered logit (continued)

- November 3 Nominal dependent variables: Multinomial logit

Required:

J. Scott Long and Jeremy Freese, *Regression Models for Categorical Dependent Variables Using Stata* (2nd edition), Chapters 6-7.

N. Kim Nguyen and James C. Garand, "Partisan Strength and Nonpartisanship Among Asian Americans," *American Politics Research* (May 2009: 375-408).

Mondak, "Reconsidering the Measurement of Political Knowledge," *Political Analysis* (2000: 57-82).

- November 10 Nominal dependent variables: Multinomial logit (continued)

- November 17 Count models: Poisson and negative binomial regression

Required:

J. Scott Long and Jeremy Freese, *Regression Models for Categorical Dependent Variables Using Stata* (2nd edition), Chapters 8.

James C. Garand and Kelly M. Burke, "Legislative Activity and the 1994 Republican Takeover: Exploring Changing Patterns of Sponsorship and Cosponsorship in the U.S. House," *American Politics Research* (March 2006: 159-88).

- November 24 Count models: Poisson and negative binomial regression (continued)

- December 1 Censored dependent variables: Tobit analysis

About the Instructor

James C. Garand (Ph.D., University of Kentucky, 1984) is the Emogene Pliner Distinguished Professor of Political Science at Louisiana State University.

Professor Garand has teaching and research interests in the fields of legislative politics, electoral politics, public opinion, public policy, state politics, racial and ethnic politics, domestic political economy, and research methodology and statistics. His research on a wide range of topics in American politics has been published in numerous journals, including the *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, *British Journal of Political Science*, *Political Research Quarterly*, *Western Political Quarterly*, *Comparative Political Studies*, *Legislative Studies Quarterly*, *PS: Political Science and Politics*, *American Politics Research*, *American Politics Quarterly*, *Public Choice*, *Social Science Quarterly*, and *Electoral Studies*, among others. His coedited book, *Before the Vote: Forecasting American National Elections*, was published by Sage Publications in 2000. His current research agenda includes numerous projects relating to the study of American politics.

Professor Garand received the 2006 LSU Distinguished Research Master Award in recognition of outstanding faculty accomplishments in research and scholarship. In 2009 he was recognized as an LSU "Rainmaker," an award given by the LSU Office of Research and Economic Development (ORED) for national and international recognition "for innovative research and creative scholarship." He served as President of the Southern Political Science Association in 2004, and he is also former president of the State Politics Section of the American Political Science Association. He served as Vice-President and Program Chair in 2001 for the Southern Political Science Association. Professor Garand is former editor of the *American Politics Quarterly*, one of the leading subfield journals in American politics. He currently serves on the editorial boards of the *American Politics Research*, *Journal of Political Marketing*, and *Ralph Bunche Journal of Public Affairs*, and he is a former member of the editorial boards of the *American Journal of Political Science*, *Journal of Politics*, *State Politics and Policy*, and *Legislative Studies Quarterly*. He also serves as a member of the International Advisory Board of the *Online Portal for Social Science Education in Methodology* (OPOSSEM).

Professor Garand has received numerous faculty awards. In 1997 Professor Garand received the LSU Alumni Association Distinguished Faculty Award in recognition of sustained excellence in teaching, research, and service. In 2012 he received the Tiger Athletic Foundation Undergraduate Teaching Award recognizing excellence in teaching in the LSU Honors College. In 2001 he received the LSU Foundation Distinguished Faculty Award in recognition of his excellence in graduate teaching, and he is the 1990 recipient of the university-wide Student Government Association Teaching Excellence Award for undergraduate teaching. He is also a recipient of the Alpha Lambda Delta Freshman Honor Society certificate of recognition for superior instruction of freshman students during the Fall 2000 semester.