POL SCI 733
ADVANCED REGRESSION
Spring 2018

Instructor: Daniel Stegmueller (ds381@duke.edu)

Class Time and Location: Lectures: Mon & Wed 10:05am to 11:20am, LSRC A156. Labs: Fridays 1:25pm to 2:40pm, Gross Hall 276.

Office Hours: Thursdays, 4 to 6pm, Gross Hall 294K.

Teaching Assistants:

- Ahn Le (anh.le@duke.edu) Office Hours: Fri, 2:45-3:45pm in Bunche Lab
- Haohan Chen (haohan.chen@duke.edu), Office Hours: Fri, 2:45-3:45pm in Bunche Lab

Books with relevant readings

- Faraway, Julian. 2005. Extending the Linear Model with R. Boca Raton: Chapman & Hall.
- Gelman, Andrew, and Jennifer Hill. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press.

Helpful background reading

- Eliason, Scott R. 1993. Maximum Likelihood Estimation: Logic and Practice. Thousand Oaks: Sage.

Grades

Grades will be based on the following items:

- **30%** — Five homework assignments
- **30%** — Two short papers
- **40%** — Final research paper

Homework (30%)

Each homework assignment will deal with the concepts and methods discussed in the previous weeks lectures. Homework assignments will be posted after Fridays lab and should be turned in via the course website by the beginning of lab the following week. For each homework, you are required to cross-grade another student's homework. The deadline for the submission of the cross-graded homework is the beginning of the Wednesday lecture. For the cross-grading procedure, we will use the following website (some of you will be familiar with this system from POLSCI 630): `http://ps733-s18.herokuapp.com`. You are required to submit your homework as a PDF document created by knitr or sweave as well as its source code. The submitted PDF file is expected to contain (1) written answers, (2) R code, and (3) R output.[1] Do not make any reference to your identity in the submitted documents, to ensure that

[1] As an aside, learn to use LaTeXwhile you can. A short introduction can be found here.

cross-grading takes place anonymously. A maximum of 20 points can be earned for each homework. For your TAs' sake, please try to keep your code and output concise and organized!
As cross-grader, you have an important responsibility in this class. You are expected to take this responsibility seriously and carefully evaluate the homework you have received from another student. If the TAs recognize that a student does not grade another person's homework carefully and in accordance with the answer key, they may reduce the grader's homework score.
You can discuss you answer strategies but are expected to write your answers and code individually, without copying from other students or external sources (such as websites or online lectures). If we recognize that answers are copied from others or external sources, the grade of the involved students will be reduced, and further action may be taken in line with the University's policies on academic dishonesty. These assignments are not busywork; they are designed to help you better understand the course material.
If the homework is late, 3 points will be subtracted from the grade of the homework for each 24-hour period it is late. If the submission of the cross-grading is late, 1.5 points will be subtracted from the grade of the grader's homework for each 24-hour period it is late.

Short papers (30%)

Treat each paper assignment as practice for writing (only!) the methods and results sections of a research article to be submitted to a peer-reviewed journal.

Paper 1: Limited Dependent Variables

Choose two out of the three options below (pick what is most helpful for your substantive research interests). (1) Find a dataset with an binary or ordered dependent variable and several relevant independent variables and estimate an ordered probit or logit model; report all regression estimates, and provide an appropriate graphical representation of the relationship of at least three of your model variables with the dependent variable, including representations of uncertainty. (2) Find a dataset with a nominal dependent variable and several relevant independent variables, and estimate a multinomial logit model; provide an appropriate graphical representation of the relationship of at least three of your model variables with the dependent variable, including representations of uncertainty. (3) Find a dataset with a count dependent variable and several relevant independent variables. Estimate a poisson model, report all regression estimates, and graphically report and interpret your results for at least three model variables including representations of uncertainty. Then, re-estimate using a quasi-poisson or a negative binomial model and compare to the results from the poisson model. You may divide this paper into two separate "mini-papers," or you may write combined methods and results sections. In all cases, discuss the strengths and weaknesses of your chosen estimation strategy. Which shortcomings are merely "technical", which ones are threats to credible causal inferences?

Paper 2: Hierarchical data

Find a dataset with a "multilevel" or panel data structure. First, propose and test two hypotheses using unit-specific constants. The first hypothesis should entail a relationship between the DV and a "first-level" variable, and the second should entail a relationship with a "second-level" variable. Second, propose and test a hypothesis that entails some form of effect heterogeneity (i.e., where the effect of a covariate varies over units). The dependent variable can be anything you wish. Describe and justify all data and modeling choices, and pay close attention not just to statistical details but also to issues of inference.

Final paper (40%)

You have two options for your final paper. Begin looking into this early on in the semester. You must clear your project with me by March 21st at the latest. The final paper is due on May 4th by 1pm!

Option 1:

Go to one of the major journal websites and find a paper on a topic that interests you. Using the provided replication data, or other relevant publicly available data, you should improve upon or extend the analysis

reported by the authors of this paper using the methods learned in this course. In other words, your job is to find an existing analysis and make it better. That might mean using a more appropriate methodology than the authors of the paper to check the robustness of the results, it might mean using a more appropriate model specification, or it might mean extending the results of the paper by proposing and testing a new hypothesis that the authors did not check. There are a large number of options here. The only requirements are that you use one or more of the methods learned in this class, and that your contribution is meaningful enough to warrant writing the paper! With respect to the latter use the following rule of thumb: would my paper be desk-rejected at a peer-reviewed journal? If the answer to this is "probably yes," then find a different topic.
You will write-up your new results in a research paper. This means: (1) introduce the topic and the previous research on this topic (obviously including the paper that inspired your own paper), (2) explain the purpose of your own paper, (3) describe the data and variable operationalizations you will use, (4) describe your modeling approach, (5) present and interpret your findings, and (6) conclude. Again, the paper should (with only relatively minor edits) be appropriate as a submission to an actual journal. We will grade your paper as-if we were reviewing your article for a journal.

Option 2:

Write a full research article on a topic of your choosing that could (with minor edits) be submitted to a peer-reviewed journal. This means: (1) introduce the topic and the previous research on this topic, (2) explain the purpose of your own paper, (3) describe the data and variable operationalizations you will use, (4) describe your modeling approach, (5) present and interpret your findings, and (6) conclude. The only requirement (in the context of this class) is that you use one or more of the methods learned in this class. We will grade your paper as-if we were reviewing an article for a journal.
Whichever option you choose, use this paper as an opportunity to advance your research agenda.

Policy & other considerations

Don't do stupid things.
I will follow the University's policy in any event of plagiarism and academic dishonesty.

Course Outline:

Week 1 (Jan 10): Introduction and Syllabus

Week 2 (Jan 17): Inference, hypothesis testing, and uncertainty.

Week 3 (Jan 22 & 24): Introduction to MLE I

Week 4 (Jan 29 & Jan 31): Introduction to MLE II

Week 5 (Feb 5 & 7): Dichotomous Outcomes

Week 6 (Feb 12 & 14): Ordered Categorical Outcomes

Week 7 (Feb 19 & 21): Unordered Categorical Outcomes

Week 8 (Feb 26 &Feb 28): Count Outcomes and Contingency Tables

Week 9 (Mar 5 & 7): How to Handle Missing Data. EM and MI

Week 10 (Mar 12 & 14): Spring Break

Week 11 (Mar 19 & 21): Models for repeated cross-sections and panel data I

Week 12 (Mar 26 & 28): Models for repeated cross-sections and panel data II

Week 13 (Apr 2 & 4): Models for repeated cross-sections and panel data III

Week 14 (Apr 9 & 11): Nonparametric Regression and Additive Models

Week 15 (Apr 16 & 18): Catch-up session