GSERM 2017
Regression III - Advanced Methods
Professor Christopher Zorn
Pennsylvania State University
`zorn@psu.edu`

Course Content

This class builds directly upon the foundations laid in Regression II, but expands it to consider a range of issues with and extensions of conventional ordinary and generalized linear regression models. After a brief review of the linear regression model, the course first considers issues of model specification, including methods for fitting nonlinearities and varying coefficients in regression model covariates. It then goes on to address nonlinear regression models more generally, both through data transformations and via models that explicitly relax the assumption of linearity. From there, the course moves toward advanced diagnostic methods, and to methods for conducting statistical inference that are robust to violations of the assumptions of the conventional linear model. The course concludes with an introduction to two classes of more advanced regression models: those with varying coefficients (sometimes referred to as "hierarchical" or "multilevel" models) and generalized linear models estimated via maximum likelihood.

Prerequisites (knowledge of topic)

Students should be familiar with ordinary and generalized least squares, at the level presented in Regression II. In addition, they should have a working familiarity with univariate differential and integral calculus, linear (matrix) algebra, and commonly-used univariate statistical distributions. There will be a brief overview of the necessary background knowledge at the beginning of the first day of the course.

Hardware

Course exercises will be completed on the students' own laptop computers. For purposes of this class, a laptop running any widely-used operating system (Windows, OS-X, Linux) will be acceptable.

Software

All lecture materials, slides, and in-class examples will be conducted using the R statistical language. Students are encouraged to come to class with current versions of both R (via `https://www.r-project.org`) and RStudio (`https://www.rstudio.com`) on their laptops. The instructor can also provide limited support for students wishing to use Stata (`http://www.stata.com`). Students electing to use other statistical software (e.g., SAS, PSPP, etc.) will be at a substantial disadvantage.

Structure

Day One:

- Brief Review of Linear (and other) Regression Models

- Linear Model Specification: Interactions, Polynomials, and More

Day Two:

- Non-Linearity: Data Transformations

- Non-Linearity: Nonlinear Models

Day Three:

- Outliers: Detection and Influence

- Robust Inference: Bootstrapping and Other Resampling Methods

Day Four:

- Varying Coefficients and Mixture Models

Day Five:

- Introduction to Maximum Likelihood and Generalized Linear Models

Literature

Mandatory

- Weisberg, Sanford. 2013. *Applied Linear Regression*, 4th Ed. New York: Wiley.

- Other readings in the available packet.

Supplementary / Voluntary

- Berk, Richard. 2010. "What You Can and Can?t Properly Do with Regression." *Journal of Quantitative Criminology* 26(4):481-487.

- Brambor, Thomas, William R. Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14:63-82.

- Freedman, D. A. 2006. "On the So-Called 'Huber Sandwich Estimator' and 'Robust' Standard Errors." *The American Statistician* 60:299-302.

- King, Gary, and Margaret E. Roberts. 2014. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What To Do About It." *Political Analysis* 22:1-21.

Mandatory Readings Before Course Start

- Weisberg, Sanford. 2014. ?Computing Primer for Applied Linear Regression, 4th Edition Using R.?

Examination

Students will be evaluated on two written homework assignments that will be completed during the course (20 percent each) and an in-class final examination given in the afternoon of the last day (60 percent). Homework assignments will typically involve a combination of simulation-based exercises and ?real data? analyses, and will be completed during the evenings while the class is in session.

Supplementary aids

The Examination-Aid Rules are "Practical Examination" and "Open Book." In addition to the course materials, useful documents include:

- Fox, John, and Sanford Weisberg. 2011. *An R and S-Plus Companion to Applied Regression*, Second Edition. Thousand Oaks, CA: Sage Publications.

- Nagler, Jonathan. 1996. "Coding Style and Good Computing Practices." *The Political Methodologist* 6(2):2-8.

Examination Content

The final examination will involve the application of the techniques taught in the class to one or more "live" data example(s). These will typically take the form of either (a) a replication and extension of an existing published work, or (b) an original analysis of observational data using regression. Students will be required to specify, estimate, and interpret various forms of regression models, to conduct and present diagnostics and robustness checks, and to give detailed justifications for their choices

Literature

- Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*, Second Edition. Thousand Oaks, CA: Sage Publications.

- Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Workload

At least 24 units, 45 minutes each, on 5 consecutive days.