Advanced Regression

University of Kentucky
BST 760, Spring 2011
Credit: 3.0

Lecture: 3:30 a.m - 5:20 p.m., Thursdays
Room 222, Multidisciplinary Science Building (MDS 222)
Lab: 3:30 a.m - 5:20 p.m., Tuesdays
Room 207, College of Public Health Building (CPH 207)

Instructor: Patrick Breheny, Ph.D.
Office: Room 203C, College of Public Health Building
Phone: 218-2077
e-mail: patrick.breheny@uky.edu
Office hours: Whenever I'm in my office, or by appointment

Course description: Regression models are ubiquitous in the research that goes on in the health and social sciences, as well as in methodological statistical research. Understanding the fundamentals of these models is critical for anyone in epidemiology or biostatistics. It is the purpose of this course – a core course in the PhD program for Epidemiology and Biostatistics – to provide that understanding. We will begin by developing the necessary mathematical tools to describe these models, and then use those tools to study linear regression, ANOVA, logistic regression, and Poisson regression, scrutinizing their unique aspects but also bringing them under the collective framework of generalized linear models. Along the way, we will develop techniques necessary for successful modeling in practical research, including model checking and diagnostics, model selection, and the visualization and presentation of fitted models.

Text:

- Dobson A.J. and Barnett A.G. (2008) *An Introduction to Generalized Linear Models*. Third edition. CRC Press.

Prerequisite: BST 675.

Corequisite: BST 676.

Course website: The course notes, assignments, data sets, and other relevant materials will be made available on the course web site:
`http://web.as.uky.edu/statistics/users/pbreheny/760-S11/index.html`

Grading: Your grade will be based on a weighted average of homework (20%), two exams (30% each), and a final project (20%).

Homework: There will be weekly assignments in this course, to be distributed on Thursday and collected the following Thursday. Late homework will not be accepted; however, only assignments that benefit your grade will be included. Any assignment (such as a 0) that lowers your grade will be automatically dropped and its weight added to the other components (exams and project) of your final grade.

Example: Annie scores 85% on her tests and final project; 4% of her homework grades are below 85%. These homework grades are dropped. Annie's overall grade is then a weighted average based on 16% homework, 63% exams, and 21% final project. Note that Annie's homework grade is guaranteed to be above 85% if she gets to drop all assignments in which she scored below 85%; thus homework cannot possibly lower your grade in this course. If you have questions on this, ask me.

Final project: There will be a final project for this class, which will consist of analyzing a data set that you have gathered. You will write a report of your findings as well as present your project to the class. We will discuss this in more detail later in the semester, but I wanted to mention it now so that you can start thinking about what data you might wish to analyze.

Proofreading: This is the first time I have taught this course, and my notes are likely to be riddled with typographical errors. If you see them, please tell me about them! Doing so will convince me that you are paying close attention in class, and leave me with a favorable impression of your work ethic.

Electronic communication: I will occasionally send e-mails to the class (to the account listed for you in the campus directory), so please check that account regularly.

Cheating: The University of Kentucky takes cheating on examinations very seriously, and has in place a number of rather severe academic sanctions, a summary of which can be found at `http://www.uky.edu/Ombud/acadoffenses/index.htm`

Complaints: Students with suggestions or complaints should see me first, and if we cannot come to an agreement, I will direct you to the head of the department.

Disabilities: If anyone has a disability requiring special accommodations, please let me know as soon as possible, so that these arrangements can be made.

I look forward to getting to know you, and I hope that we have a great semester together!

Course outline:

1. Introduction
   (a) Experiments, observational studies, and models
   (b) Simple linear regression
2. Matrix algebra
3. Normal linear models
   (a) Basic terminology, notation, and concepts
   (b) Estimation and model fitting
   (c) Diagnostics and residuals
   (d) Inference: standard errors, confidence intervals, hypothesis testing
   (e) ANOVA
      i. Experimental design: blocking, balanced and incomplete blocks, nested and crossed factors, factorial designs
      ii. One-, two-, and multi-factor ANOVA
      iii. ANOVA tables
   (f) Model building (confounding, collinearity, interactions, variable selection)
4. Generalized linear models
   (a) Exponential families
   (b) Generalized linear models
   (c) Maximum likelihood estimation and model fitting
   (d) Likelihood theory and asymptotic inference
5. Logistic regression
   (a) Logit functions and the logistic regression model
   (b) Model diagnostics
   (c) Model building
   (d) Reporting and summarizing logistic regression models
6. Poisson regression
   (a) Poisson regression and log-linear models
   (b) Model building and diagnostics
   (c) Reporting and summarizing Poisson regression models
7. Further topics (as time permits), such as nonparametric regression, ordinal and polytomous logistic regression, ridge regression, random effects, ...