Advanced Regression

University of Kentucky BST 760, Spring 2013 Credit: 3.0

Lecture: 9:30 p.m - 10:45 p.m., Tuesdays and Thursdays Room 221, Multidisciplinary Science Building (MDS 221)

Instructor: Patrick Breheny, Ph.D.
Office: Room 205D, Multidisciplinary Science 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 applied as well as methodological statistical research. Particularly in the health and social sciences, we are often interested in outcomes that do not follow a normal distribution, such as binary outcomes (survived/died, successful/unsuccessful therapy) and counts (number of infections/cases of cancer/complications at a hospital or in a county). 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.

Course objectives:

- To develop the mathematical tools and likelihood theory necessary to describe generalized linear models and carry out inference concerning model parameters
- To provide familiarity with the computational tools and software needed to fit logistic and Poisson regression models and extract useful information from the fitted models
- To develop and reinforce techniques and experience for successful modeling in practical research, including model checking and diagnostics, model selection, and the visualization and written 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/S13/index.html

Grading: Your grade will be based on a weighted average of homework (16%), two exams (16% each), and projects (16% each for the two mid-term projects, 20% for the final project). The grading scale is below:

A: 85-100 B: 70-85 C: 55-70 F: 0-55

Exams: The two midterm exams are of equal weight, and will be closed-book tests administered during class time on the following dates:

Exam I Tuesday, February 12 Exam II Thursday, March 7

If you cannot take the exams on these dates, let me know as soon as possible so that we may work out a makeup date/time. Note that there is no final exam.

Homework: Assignments will be given approximately weekly, although no homework will be assigned on weeks in which there is an exam or in which a project is due. Thus, there will be more frequent homework assignments at the beginning of this course, but fewer later in the course as we move into more comprehensive exams and projects. Solutions will be posted the morning after the due date; beyond this point, late homework cannot be accepted. You may work with other students on the homework assignments, but all students must write and turn in their own solutions.

Assignments may be turned in either electronically or as a physical copy (preferred). If they are turned it electronically, they must be turned in as a .pdf file (*i.e.*, no Word documents). Each assignment will also involve writing code (to analyze data, make plots, etc.). Please turn in this code **separately and electronically**.

All electronic submissions should be made via DROPitTOme at http://www.dropitto.me/ pbreheny and should follow the following naming convention: last name, assignment number, proper extension. So, for example, if Adam Smith was taking this course and turning in assignment 1, he would name the file Smith1.pdf. The associated code would be Smith1.R or Smith1.sas.

Projects: There will be three projects in this course. For projects 1 and 2 (worth 16% each), I will supply the data set and define the questions of interest. There will also be a final project (worth 20%), for which you will acquire the data and define the questions of interest as well as carry out the analysis. For all three projects, you will write a report of your findings. You may discuss your analysis with other students, but each student's report must be written independently.

For the final project, you will also present your work to the class. We will discuss this in more detail later in the semester, but it is never too early to start thinking about your final project and what data you might wish to analyze.

Proofreading: If you see any typos in my notes (no matter how small), please tell me about them! Doing so will not only benefit you, but also myself, your classmates and any future students of this course.

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.

Academic honesty: Academic honesty is highly valued at the University of Kentucky. You must always submit work that represents your original words or ideas. If any words or ideas used in an assignment or project do not represent your original words or ideas, you must cite all relevant sources and make clear the extent to which such sources were used.

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 you have a documented disability that requires academic accommodations, please see me as soon as possible during scheduled office hours. If you have not already done so, please register with the Disability Resource Center for coordination of campus disability services available to students with disabilities.

Religious observances: If a religious observance prevents you from taking an exam or finishing an assignment or project, please let me know in advance so that we can make arrangements for you to make up the work.

Inclement weather: The University of Kentucky has a detailed policy for decisions to close in inclement weather. The snow policy is described in detail at http://www.uky.edu/MicroLabs/documents/temp/policies-weather.htm or you can call (859) 257-5684.

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

Course schedule: (subject to change)

- 1-10 Introduction; causality and confounding
- 1-15 Overview of R
- 1-17 Transformations
- 1-22 Generalized linear model framework
- 1-24 Exponential families; maximum likelihood estimation
- 1-29 Maximum likelihood estimation (cont'd)
- 1-31 Asymptotic multivariate maximum likelihood theory
- 2-5 Asymptotic multivariate maximum likelihood theory (cont'd)
- 2-7 Weighted least squares
- 2-12 Exam I
- 2-14 GLM model fitting & the iteratively reweighted least squares algorithm
- 2-19 GLM model fitting & the iteratively reweighted least squares algorithm (cont'd)
- 2-21 Logit functions and the logistic regression model
- 2-26 Logistic regression: Estimation and inference for probabilities and odds ratios
- 2-28 Logistic regression and case-control studies
- 3-5 Wald vs. likelihood ratio approaches to inference

3-7 Exam II

- 3-12 No class (spring break)
- 3-14 No class (spring break)
- 3-19 Logistic regression: Residuals and diagnostics
- 3-21 Logistic regression model selection
- 3-26 Logistic regression: Case study
- 3-28 Logistic regression: Case study (cont'd)
- 4-2 The Poisson regression model, estimation, and inference **Project 1 due**
- 4-4 Poisson regression: Model building, diagnostics, offsets
- 4-9 Overdispersion: Quasi-likelihood and negative binomial models
- 4-11 Poisson regression: Case study
- 4-16 Multinomial regression
- 4-18 Multinomial regression: Case study
- 4-23 The proportional odds model **Project 2 due**
- 4-25 Proportional odds modeling: Case study
- 5-2 Final projects due

PhD program competency attainment for BST 760 – Advanced Regression

Key:

Competency level	Number	Description
Unaware	0	No information or skill in this area
Aware	1	Able to identify the concept or skill but with limited
		ability to perform or apply it independently
Knowledgeable	2	Able to apply and describe the concept or skill
Proficient	3	Able to synthesize, critique, or teach the concept or skill

Biostatistics competencies attained:

	Level	
Competency	attained	
Understand the interface between biostatistics and epidemiology	1	
Demonstrate advanced proficiency to apply concepts and methods from these		
disciplines jointly		
Demonstrate the ability to review and critically evaluate the literature in a		
substantive area of research, be able to identify gaps in knowledge and be able		
to formulate original research hypotheses or statements		
Evaluate the strengths and limitations of epidemiologic reports	1	
Draw appropriate inferences from data	2	
Communicate research results orally and in writing to lay and professional audiences	2	
Demonstrate an understanding of concepts of probability and statistical inference		
as they apply to problems in public health		
Demonstrate proficiency in using computing tools commonly encountered in		
epidemiology and biostatistics		
Understand the principles of epidemiologic study design and be able to calculate		
the appropriate epidemiologic measures for most typical designs		
Become proficient at and be able to evaluate the strengths and limitations of advanced	2	
designs including multivariate linear models, generalized linear models, longitudinal		
models, mixed effects models, and survival models both parametric and		
nonparametric		
Understand the principles of chronic and infectious disease epidemiology	0	
Demonstrate an understanding of research methods used in epidemiology and		
biostatistics		
Demonstrate knowledge of the public health system in the commonwealth and the	0	
country		