Regression Methods for Population Health
Biostatistics & Medical Informatics/Population Health 552

Spring 2016

Course Instructor

Ronald Gangnon
Associate Professor, Biostatistics and Population Health
Room 603 WARF Office Building
ronald@biostat.wisc.edu
Office hours: 2:30-3:30pm Tuesdays and Thursdays or By appointment

Teaching Assistants

Patricia Jewett
Room 307 WARF Office Building
pjewett@wisc.edu
Office hours: 10:00am-noon Fridays

Wesley Chang
Room 660 WARF Office Building
wchang44@wisc.edu
Office hours: 1:00-3:00pm Mondays

Lectures (PM)

4:00-5:15pm Tuesdays and Thursdays
1220 HSLC

Lab Sessions (AM)

8:00-9:15 Tuesdays and Thursdays
2121 HSLC

Course Description and Learning Objectives

The course is an accessible introduction to the primary statistical tools used in epidemiology and health services research: multiple linear regression for continuous outcomes, logistic regression for binary outcomes and Cox proportional hazards regression for time-to-event outcomes, with emphasis on their proper use and interpretation. Analyses of a variety of real data sets will receive considerable attention. Non-essential mathematical complexities will be avoided. The development of the ability to interpret results and to evaluate critically the methods used is of paramount importance.

Upon successful completion of the course, students will be able to

1. State the assumptions underlying linear, logistic, survival and Poisson regression models, recognize and address violations of those assumptions, and estimate and

interpret regression models to answer epidemiologic and public health research questions.
2. Critique uses of linear, logistic, survival and Poisson regression models in the epidemiologic and public health literature.
3. Translate epidemiologic concepts into statistical modeling assumptions, and explain statistical modeling assumptions in epidemiologic terms.
4. Recognize applications that require methods beyond their expertise, and identify resources to learn about more advanced techniques.

"Statistics is, or should be, about scientific investigation and how to do it better, but many statisticians believe it is a branch of mathematics…. Now I agree that the physicist, the chemist, the engineer, and the statistician can never know too much mathematics, but their objectives should be better physics, better chemistry, better engineering, and in the case of statistics, better scientific investigation. Whether in any given study this implies more or less mathematics is incidental." –George Box

Course Requirements and Evaluation

1. Staying current with the assigned readings, completing the pre-class *minute papers* and participating in class.
2. Participating in the lab sessions and completing the lab exercises.
3. Four data analysis assignments, one for each module of the course.
4. One critical review of a scientific paper.
5. One in-class final exam.

Textbooks

Regression with Linear Predictors, 2010. Andersen PK and Skovgaard LT. Springer: New York. ISBN: 978-1-4614-2627-1.

Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers, 3rd Edition. 2011. Katz MH. Cambridge University Press: Cambridge. ISBN: 978-0-521-14107-9.

Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. 2nd Edition. 2015. Harrell FE. Springer: New York. ISBN: 978-3-3191-9424-0.

Assigned Readings and Minute Papers

Assigned readings should be read prior to lecture. Prior to the start of the lecture(s) on a given topic, you will complete a "minute paper" on the assigned readings (due at noon on the day of an afternoon lecture or 4pm on the prior day for a morning lecture). There will be a total of 23 minute papers. You are allowed to miss one minute paper without penalty. For each additional missed minute paper beyond the first, your final course grade will be lowered by 1%.

1/19	Course Introduction, Review of Basic Concepts Andersen & Skovgaard: Chapter 1-2

Single Predictor Linear Regression

1/21	Linear Regression with a Single Binary Covariate Andersen & Skovgaard: 3.1.1
1/26 1/28	Linear Regression with a Single Categorical Covariate Andersen & Skovgaard: 3.2.1 Katz: 4.2, 4.5, 7.1-7.2, 8.5 Encyclopedia of Biostatistics: ANOVA, Multiple Comparisons
2/2 AM 2/2 PM	Linear Regression with a Single Quantitative Covariate (Linear Effect) Andersen & Skovgaard: 4.1.1, 9.1-9.6 Encyclopedia of Biostatistics: Simple Linear Regression, Standardized Coefficients
2/9 2/11	Linear Regression with a Single Quantitative Covariate (Non-Linear Effect) Andersen & Skovgaard: 4.2 Katz: 4.3-4.4 Harrell: 2.4, 2.6 Encyclopedia of Biostatistics: Polynomial Regression, Spline Smoothing, Categorizing Continuous Variables

Multiple Predictor Linear Regression

2/16	Multiple Linear Regression Andersen & Skovgaard: 5.1, 5.3.1 Katz: 1.1, 3.2, 5.1-5.3, 8.3.A, 9.7 Harrell: 2.1-2.3.2 Encyclopedia of Biostatistics: Multiple Linear Regression
2/18	Confounding Andersen & Skovgaard: 6.1.1-6.1.3 Katz: 1.2-1.3, 2.2, 6.1-6.2 Encyclopedia of Biostatistics: Confounding
2/23	Interaction Andersen & Skovgaard: 5.2, 5.3.2, 6.1.7

	Katz: 1.4, 7.3, 8.4, 9.7 Harrell: 2.3.2 Encyclopedia of Biostatistics: Interaction, Effect Modifications
2/25	Mediation Katz: 6.3 Baron & Kenney (1986)
3/1	Predictor Selection Andersen & Skovgaard: 6.1.4-6.1.5, 6.2.1 Katz: 6.3-6.7, 7.8, 12.1 Harrell: 4.1-4.12

Binary, Multinomial and Poisson Regression

3/3 3/10 AM	Logistic Regression Andersen & Skovgaard: 3.1.2, 3.2.2, 4.1.2 Katz: 3.3, 7.9, 8.2.B, 8.3.B Harrell: 10.1-10.3, 10.5, 10.10 Encyclopedia of Biostatistics: Logistic Regression, Likelihood Ratio Tests
3/10 PM	Case-Control Studies Andersen & Skovgaard: 7.4.2 Encyclopedia of Biostatistics: Conditional Logistic Regression
3/15	Alternatives to the Logit Link Andersen & Skovgaard: 7.4.1 Encyclopedia of Biostatistics: Generalized Linear Model
3/17	Ordinal Multinomial Outcome Andersen & Skovgaard: 7.1.1 Katz: 3.4, 8.2.C, 8.3.C, 9.8 Harrell: 13.1-13.2, 13.3.1-13.3.3, 13.3.5, 13.3.7, 13.4.1-13.4.3, 13.4.5, 13.4.7 Encyclopedia of Biostatistics: Proportional Odds Model
3/29	Nominal Multinomial Outcome Andersen & Skovgaard: 7.1.2 Katz: 3.5, 8.2.D, 8.3.D Encyclopedia of Biostatistics: Polytomous Data
3/31 4/5	Poisson and Negative Binomial Regression Andersen & Skovgaard: 7.2 Katz: 3.10-3.11, 7.7, 8.2.F, 8.3.F Encyclopedia of Biostatistics: Poisson Regression, Poisson Regression in Epidemiology, Overdispersion, Negative Binomial Model

Survival Analysis

4/7	Survival Data, Censoring and Truncation Katz: 3.6-3.8 Harrell: 17.1-17.3 Encyclopedia of Biostatistics: Survival Analysis, Censored Data, Time Origin
4/12	Kaplan-Meier Estimator and Logrank Test Andersen & Skovgaard: 3.1.3, 3.2.3 Harrell: 17.5 Encyclopedia of Biostatistics: Kaplan-Meier Estimator, Logrank Test
4/14 4/19	Cox Proportional Hazards Regression Andersen & Skovgaard: 4.1.3 Katz: 3.6, 7.4-7.6, 8.2.E, 8.3.E Harrell: 20.1-20.2, 20.4-20.5, 20.6.1 Encyclopedia of Biostatistics: Cox Regression Model
4/21	Time Dependent Covariates Andersen & Skovgaard: 6.2.3 Katz: 13.1-13.2 Encyclopedia of Biostatistics: Time Dependent Covariate
4/26	Non-Proportionality Katz: 3.9, 9.9-9.10 Harrell: 20.6.2, 20.7
4/28	Multiplicative Hazard Models (Parametric) Andersen & Skovgaard: 7.5.1 Harrell: 18.2 Encyclopedia of Biostatistics: Parametric Models in Survival Analysis
5/3	Accelerated Failure Time Models Andersen & Skovgaard: 7.5.2 Harrell: 18.3 Encyclopedia of Biostatistics: Accelerated Failure Time Models
5/5	Competing Risks Andersen & Skovgaard: 8.1.3 Harrell: 17.6 Encyclopedia of Biostatistics: Competing Risks

Lab Sessions

Attendance and participation in the lab sessions, which start on 1/26, is mandatory. In a typical lab session, students will perform a series of data analyses related to the material covered in the prior lecture followed by a group discussion of the exercise. If you are unable to attend a lab session due to extenuating circumstances, please notify the course instructor in advance to make alternative arrangements to complete the lab exercises. For each unexcused absence, your final course grade will be lowered by 1%.

Data Analysis Assignments

There will be four data analysis assignments, one for each module of the course. For each assignment, you will perform a statistical analysis of a real dataset of your choice with an appropriate response variable. You will find the analysis more interesting, rewarding and educational when it is based on data that are of genuine interest to you. If you have questions about the appropriateness of a given dataset, please discuss with the course instructor.

You will analyze the data in a full and appropriate fashion using appropriate methods and techniques. You will prepare a complete written report of the analysis. A good report will include the following elements: (1) a statement of the problem and why it is interesting, (2) a description of the data and its source, (3) the research questions of interest, (4) any graphical or statistical methodology used, (5) discussion of the results obtained, including any adjustments to the data or corrective actions that might have been taken, (6) implications of your results for the real-world problem, and (6) a summary of what you've learned from the analysis. You should report and summarize the implications of your "best" model even if you ultimately decide that it is deficient for some reason (you should also discuss why you think it's deficient, and what you might be able to do to fix it, even if you don't have the tools yet to do so). Remember "all models are wrong; the practical question is how wrong do they have to be to not be useful."

You will turn in a hardcopy of your report at 4pm on 2/18 (single predictor linear regression), 8am on 3/10 (multiple predictor linear regression), 4pm on 4/12 (binary, multinomial and Poisson regression) and 4pm on 5/12 (survival analysis). Each data analysis assignment will be worth 17.5% of your final grade.

Critical Review of Scientific Paper

You will select a scientific paper of substantive interest to you from a scientific journal. The article must be an original research article; it cannot be a review article or editorial. It should have been published in 2015 or 2016. You will critically evaluate paper using the checklist for critiquing a research article from Kuyper (1991) Bringing Up Scientists in the Art of Critiquing Research *BioScience* 41(4): 248-250. If you have questions about the appropriateness of a given article, please discuss with the course instructor.

You will turn in hardcopies of the completed review paper and the article at 4pm on 4/26. The critical review assignment will be worth 10% of your final grade.

Final Exam

There will be a cumulative, in-class final examination. It will be similar in format to the Biostatistics section of the Population Health/Epidemiology PhD Qualifying Exam. For the exam, you will need a calculator, and you can bring two pages of notes. The exam will be held from 4-7pm on 5/12 in 1220 HSLC. The final exam will be worth 20% of your final grade.

Software

You are welcome to use any statistical software for the data analysis assignments and lab exercises. The instructors will provide support for SAS, and course instruction will use SAS. If you choose to use other software, instructors will provide as much assistance as they can, but support cannot be guaranteed.

There are several options for getting access to SAS software at no charge. SAS University Edition is the recommended option. It is available for all computer operating systems (Windows, Linux, Mac) from www.sas.com/en_us/software/university_edition.html. During the scheduled lab session on 1/21, the instructor and TAs will provide assistance with setting up SAS University Edition on your laptop.

The regular edition of SAS is available for Windows and Linux operating systems. It can be downloaded from the UW-Madison Campus Software Library (https://software.wisc.edu/cgi-bin/ssl/csl.cgi). (Note that some of these files are very large. They will take a long time to download even on a wired connection. It is strongly recommended that you not use a wireless connection to download these files.)

If you have not taken BMI/PHS 451 Introduction to SAS for Population Health, you should review the Introduction to SAS class notes from the UCLA Institute for Digital Research and Education (http://www.ats.ucla.edu/stat/sas/notes/default.htm).