PUBHLTH 690NR (3 credits)
Biostatistics Methods II: Statistical Modeling and Data Visualization
Spring 2014 :: T/Th 11:15-12:30 :: Arnold 103

INSTRUCTOR
Professor Nicholas G Reich Ph.D.
425 Arnold House
(413) 545-4534
nick [at] umass.edu
on twitter: @reichlab

Office Hours: TBD
Teaching assistant: TBD
Course notes on GitHub

MATERIALS

Required Textbook

Kutner, M., Nachtsheim, C., Neter, J., and Li, W. (2004). *Applied Linear Statistical Models*, *4th Edition*.

Recommended Textbooks

Hosmer DW and Lemeshow S. (2000). *Applied Logistic Regression*. *2nd Edition*.

Diez, D., Barr, C. and Çetinkaya-Rundel, M. (2012). *OpenIntro Statistics*. *2nd Edition*. free online textbook.

Software

R :: r-project.org (or just Google "r")

RStudio :: rstudio.org

Prerequisites

Biostatistics Methods 1, or equivalent. Otherwise, permission of instructor required. Working knowledge of basic matrix methods and calculus will be helpful. Familiarity with the R statistical programming language is expected.

Course Goals

The aim of this course is to provide fundamental statistical concepts and tools relevant to the practice of summarizing, analyzing, and visualizing data. Continuing where Methods 1 (PUBHLTH 590AF) leaves off, this course will build your knowledge of the fundamental principles of biostatistical inference. The course will focus on linear regression and generalized linear regression models. We will use a variety of examples and exercises from medical and public health research.

Learning Goals *(By the end of the course students will be able to...)*

- perform advanced programming tasks in R, the language of modern statistical computing,
- independently formulate, fit, and interpret regression models to weigh evidence for/against hypotheses about associations between variables,
- interpret commonly used statistics in public health studies,
- diagnose the "goodness-of-fit" of a given regression model, both on its own and relative to other regression models,
- create powerful data visualizations (using R package ggplot2) that reveal features of data or fitted models,
- design and run simulation studies,
- write reproducible statistical analysis reports using knitr.

Expectations

Things you should expect from me:

- timely feedback on assignments and quizzes
- response to emailed questions in < 2 working days (often sooner)
- attention to your questions related to coursework during office hours
- instruction in how to write, research, and debug R code

Things you should not expect from me:

- time for frequent non-office hour drop-in questions
- comments on a research project that is unrelated to your coursework
- writing your code for you or *extensive* debugging of your code

Types of Assignments and Activities, with Grade Contributions

Problem sets and readings (25%): Throughout the semester, I will assign several chapters of reading in a textbook and some accompanying problems (usually from the textbook). There will be between 3 and 5 problem sets throughout the semester. You will be required to hand in reproducible solutions, i.e. a knitr file that reproduces your answers.

Quizzes (25%): There will be occasional quizzes, some announced, some unannounced. They will be short (about 10 minutes), in-class quizzes that will test your understanding of material covered in the course up to that time. The quizzes will not be designed to be difficult, as they are largely designed to evaluate participation, engagement with the material, and attendance. Quizzes will take many forms: short answer or multiple choice questions, an in-class survey of results of a take-home lab assignment, etc... I will drop your lowest quiz score when calculating your final grade.

Participation (10%) : You are expected to attend class every day, to participate actively in group work and discussions, and to provide helpful and insightful feedback to fellow students.

Labs: We will be doing some hands-on in-class work. Often, there will be some additional work required to finish the labs after class ends. Labs will not be graded in and of themselves, but your understanding of the material covered by the labs will be evaluated often by short presentations and quizzes.

Short presentations (15%): In almost all class periods, there will be several short assignments given by individual students during class. The presentations will each be about 5 minutes long and typically will present results from in-class labs or other brief assignments given the previous class period. Presenters will be chosen randomly from the class roster. These presentations will count towards your participation grade.

Independent Final Project (25%): A large component of the course will be an independent project which will be presented to and evaluated (in part) by your classmates. A separate handout will provide details.

Course Policies

Collaboration on homework is expected and encouraged, although you must write up your own assignment. No copying or cutting and pasting. Quizzes must be completed without assistance from your classmates. Your independent projects must be your own work. You may discuss your project with others and even solicit ideas and advice, but at the end of the day, you must complete all the analysis and write-up on your own. Any explicitly borrowed ideas must be cited appropriately.

Late assignments: Completing homework assignments on time will be vital to not falling behind in this course. It is expected that you hand in assignments on time. If an assignment is handed in late, you will receive zero credit for the number of problems that your homework assignment is late. Days late will be rounded rounded up: i.e. if your problem is less than 24 hours late you will receive a zero for one problem. Note that while there may be many problems assigned for a given problem set, it is possible that a few problems will be graded.

Make-up quizzes: Make-ups will not be allowed. I will drop the lowest quiz score when calculating this portion of your grade. Quizzes may be unannounced.

Attendance is required. Absences (excused or not) will impact your participation grade.

All mobile devices (smart phones, dumb phones, tablets, etc...) must be turned off at the start of class and may not be used during class time.

Formal CEPH Course Competencies

- Describe the role biostatistics serves in public health.
- Distinguish among the different measurement scales and the implications for selection of statistical methods to be used based on these distinctions.
- Describe conceptual frameworks (statistical literacy) in biostatistics
- Apply biostatistical methods to the design of studies in public health.
- Use computers to appropriately store, manage, manipulate and process data for a research study using modern software.
- Apply descriptive techniques commonly used to summarize public health data.
- Describe the basic concepts of probability, random variation and selected, commonly used, probability distributions.
- Select and perform the appropriate descriptive and inferential statistical methods in selected basic study design settings.
- Describe appropriate methodological alternatives to commonly used statistical methods when assumptions are violated.
- Integrate analysis strategies in biostatistics with principles and issues in epidemiology.
- Apply basic informatics techniques with vital statistics and public health records in the description of public health characteristics.
- Interpret results and critically evaluate basic statistical aspects of public health research and practice reported in the literature.
- Assist in the application of statistical theory to applied statistical problems.
- Develop a conceptual framework that integrates techniques and methods in biostatistics
- Critically evaluate statistical aspects of public health research reported in the literature
- Develop written and oral presentations based on statistical analyses for both public health professionals and educated lay audiences.
- Apply statistical methods to solve problems in the health sciences and carry out theoretical research in statistical methodology.

Academic Honesty Policy Statement Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst.

Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. The procedures outlined below are intended to provide an efficient and orderly process by which action may be taken if it appears that academic dishonesty has occurred and by which students may appeal such actions.

Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent. For more information about what constitutes academic dishonesty, please see the Dean of Students? website.

Disability Statement The University of Massachusetts Amherst is committed to making reasonable, effective and appropriate accommodations to meet the needs of students with disabilities and help create a barrier-free campus. If you are in need of accommodation for a documented disability, register with Disability Services to have an accommodation letter sent to your faculty. It is your responsibility to initiate these services and to communicate with faculty ahead of time to manage accommodations in a timely manner. For more information, consult the Disability Services website.