PUBHLTH 460 (3 credits)
Telling Stories with Data: Statistics, Modeling, and Data Visualization
Spring 2023 :: T/Th 10:00-11:15am :: ILC S311

INSTRUCTORS

Nicholas G Reich, nick [at] umass.edu, Reich Lab website, @reichlab
Lead teaching assistant: Weidong Wang
Course assistants: Sophia Fortier, Minh Le
Instructor and TA Office Hours: See Moodle

MATERIALS

Required Textbook

Baumer, Ben. 2021. *Modern Data Science with R*. (Free online textbook)

Recommended Textbook (freely available online)

Diez D, Barr C, and Çetinkaya-Rundel M. 2015. *OpenIntro Statistics, 3rd Ed.*.

Wickham H and Grolemund G. 2017. *R for Data Science*

Software (both free downloads)

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

RStudio :: rstudio.org

PREREQUISITES

One of any of the following introductory stats courses taught at UMass: BIOSTAT 223, STAT 111, STAT 240, STAT 501, ResEcon 212, PSYCH 240. If you have not taken an intro stats course at UMass but still want to enroll in this course, you are encouraged to petition the instructor for permission, especially if any of the following apply: (a) you have taken AP Stats in high school, (b) you have taken a college-level intro stats course just not one of the ones listed above, or (c) you are confident in your quantitative skills and your ability to succeed in a fast-paced, advanced introductory course. Additionally, prior programming experience with R or concurrent enrollment in PUBHLTH 497D (Introduction to Statistical Computing with R) is required.

COURSE DESCRIPTION

The aim of this course is to provide students with the skills necessary to tell interesting and useful stories in real-world encounters with data. Specifically, they will develop the statistical and programming expertise necessary to analyze datasets with complex relationships between variables. Students will gain hands-on experience summarizing, visualizing, modeling, and analyzing data. Students will learn how to build statistical models that can be used to describe and evaluate multidimensional relationships that exist in the real world. Specific methods covered will include linear and logistic regression. Students will work with the R statistical computing language and by the end of the course will require substantial independent programming. The course will not provide introductory training in R programming. To the extent possible, the course will draw on real datasets from biological and biomedical applications. This course is designed for students who are looking for a second course in applied statistics/biostatistics (e.g. beyond BIOSTATS 223 or STAT 240), or an accelerated introduction to statistics and modern statistical computing.

LEARNING GOALS *(By the end of the course students will be able to...)*

- design data-driven experiments and analyses to answer specific questions,
- use data to identify and distinguish patterns of randomness vs. non-randomness,
- create powerful data visualizations that reveal and highlight important features of data or models,
- understand and critique statistical model equations as representations of a given real-world setting,
- formulate, fit, and interpret statistical models to designed to answer specific scientific questions,
- weigh evidence for/against hypotheses about associations between variables,
- diagnose the appropriateness or "goodness-of-fit" of a given model,
- write concise, professional, and reproducible statistical analysis reports.

Expectations

This course will require you to work thoughtfully, carefully, and independently and will require substantial work outside of class time. Because we will be using a more project-driven approach in this course, with assignments that will build upon one another into a final product, it is vital that you do not fall behind. If you feel as though you are falling behind or starting to lose a handle on the content, I expect you to talk to me during office hours or during a separate appointment so that I can help as much as I can to set you back on track. Please do not wait to talk to me if you start to fall behind.

I also expect you to devote substantial outside-of-class time to your work for this course, typically involving 5-10 hours per week. I anticipate that this work will be divided among:

- finishing in-class activities
- reading assigned articles and chapters
- reviewing your notes or the recorded lectures
- working on assignments
- conducting project work
- preparing for exams

Things you should expect from me:

- timely feedback on assignments and quizzes
- response to questions via Discord 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 (35%): There will be approximately five lab assignments and five coding challenges that you will complete over the course of the semester. Each problem set will have components that you will hand in for grading. Typically, each assignment will be worth the same amount of points. Assignments, total possible point values, and due dates will be posted in advance on Moodle. Some assignments will require you to submit a digital file with reproducible solutions, i.e. an RMarkdown file that reproduces your answers. Late assignments will not be accepted under any circumstances. This is a strict policy. If a problem set (either a lab or a coding challenge) is not handed in on time, it will receive a grade of zero. I will drop your lowest two grades when calculating your final problem set grade, so if you miss one or two assignments and do well on the others it should not impact your grade substantially.

Midterm exam (25%): There will be a mid-term exam in this course. Exact format and timing will be posted on Moodle.

Final Project (30%): In the second half of the course, you will develop and write your own data story as part of a small team. This project will be presented to your classmates. Details will be posted on Moodle during the semester.

Citizenship (10%) : Being a good class "citizen" also plays a large role in your final grade. A partial list of the characteristics of good class citizens are: attending all course meetings, using office hours, asking questions, offering to answer questions, actively listening when others are talking, not interrupting others, helping to foster a non-judgmental and inclusive classroom environment, and participating on the Discord and Moodle forums (both asking and answering questions). Citizenship, unlike participation, is more a function of quality than quantity and can't just be earned by "showing up". The default citizenship score is 5 out of 10.[1]

[1] Acknowledgments to Aaron Swoboda for introducing me to the concept of course citizenship and for some of this text.

Course Policies

Working together: Collaboration on all assignments is expected and encouraged, although you must write up your own assignment. **No copying or cutting and pasting.** Each project will contain an independent component which 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 or language must be cited or acknowledged appropriately. Copying of assignments will result in a zero grade for the assignment and appropriate consequences will be assessed in line with the campus-wide academic honesty policy (see below).

Digital devices: Students can be easily distracted from their work by mobiles devices. Mobile devices will be distracting to you or others during class. Therefore, the policy for this class is that all mobile devices must be stored and turned off or switched into "do not disturb" mode at the start of class. Small mobile devices may not be used during class time for any reason. You may use laptops and tablets during classtime for class purposes only. In the past, this has been a very hard policy for students (and myself) to follow. To help me, I have set an automatic and repeating "do not disturb" rule on my phone so that it never releases any notifications during my classtimes, even when I forget to turn my phone off before class. I also put my phone somewhere where I cannot easily access it or turn it off during class so I am not tempted to check it. I encourage you to consider taking actions such as these to avoid feeling the urge to use your phone during class. Use of devices during class time for non-class purposes will impact your citizenship grade for the class. If you send a direct message via email or Discord to the TA Weidong Wang with the text "I read the syllabus" by the beginning of the second class, you will receive one point of extra credit on your final grade.

Attendance: In-person attendance is required. Multiple absences (excused or not) will impact your citizenship grade. You can earn attendance credit even when you miss class by emailing the lead TA for the course ahead of class-time with the reason for your absence, watching the class recording on Echo360 and completing any in-class work, such as a note-catcher, on your own, by 12pm (noon) ET on the day after the class was held.

COVID-19: Students are expected to follow the UMass Guidance on Isolation and Precautions for Individuals with COVID-19 or Exposed to COVID-19. If you must miss class for a COVID-related issue, you should follow the instructions in the Attendance section above to earn attendance credit for class. Note that while we anticipate that there may be unavoidable reasons to miss class perhaps once or twice a semester, we expect that when you are absent you follow the above guidelines completing in-class work.

Respect: Due to the team-based learning setup of this course, there will be times throughout the semester when you will be listening to presentations from other students within your team or from other teams. Act respectfully during these presentations by listening carefully, not interrupting, and waiting until presentations are over to ask questions or excuse yourself for a water or bathroom break.

Grading Scale

Grade	Percentage
A	93-100
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	67-69
D	63-66
D-	60-62
F	0-59

Course Schedule

Please see the most up-to-date schedule on Moodle.

Council on Education for Public Health (CEPH) Course Competencies

- 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. 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.