UT-Austin iSchool SyllabusI306 Statistics for InformaticsSpring 2023

1/3/23

Description (from the catalog)

Restricted to informatics majors and students pursuing the informatics minor. Examine fundamental principles of probability and statistics. Cultivate an understanding of descriptive and inferential statistics. Conduct and interpret statistical analyses using statistical analysis software, and apply these analyses to common issues in informatics. Three lecture hours a week for one semester. Offered on the letter-grade basis only.

Details

Important note: The information presented in this syllabus is subject to expansion, contraction, change, or stasis during the semester. In case of conflict between versions, the copy on Canvas takes precedence.

Course Number

28075

Prerequisites

none, but knowing math and programming will help

Time

TTh 1700–1830

Place

UTC-1.116

Dates

January 9–May 1, 2023

Final Exam

Take-home, due at our official exam time

Instructor

Mick McQuaid

Email

mcq@utexas.edu

Office

1616 Guadalupe St, Room 5.402

Office Hours

TUE 1430–1630, WED 1300–1500 or by appointment

Materials

Our textbook is the freely available Diez, Çetinkaya-Rundel, and Barr (2019). Additional textbooks you can use include the freely downloadable Wickham, Çetinkaya-Rundel, and Grolemund (2023) and James et al. (2021). A more advanced textbook is the freely available Kuhn and Silge (2022), the full text of which is available at tmwr.

Learning Outcomes

- learn to describe data using statistics and contingency tables to summarize
- learn to use probability distributions
- learn to visualize data using R
- learn to develop confidence intervals
- learn to conduct hypothesis tests
- learn to conduct single and multiple regression using R
- learn to write reproducible reports using Quarto

Class Format

The class will be half lecture (Tuesdays), half work time on a computer (Thursdays). You should definitely bring a laptop computer to class, especially on Thursdays. On Tuesdays we will go through the Diez, Çetinkaya-Rundel, and Barr (2019) book, while on Thursdays we will mostly work through the exercises in the Wickham, Çetinkaya-Rundel, and Grolemund (2023) book.

Schedule

Week 1 (10 Jan, 12 Jan) Introduction to data — Introduction to R — Introduction to R Studio — Introduction to R markdown

Week 2 (17 Jan, 19 Jan) Summarizing data — Examining numerical data — Considering categorical data — More on R markdown — Introduction to Quarto

Week 3 (24 Jan, 26 Jan) More on R and quarto — Visually summarizing data — Milestone 1 due

Week 4 (31 Jan, 2 Feb) Probability — Defining probability — Conditional probability — Sampling from a small population — Random variables — Continuous distributions

Week 5 (7 Feb, 9 Feb) Distributions of random variables — Normal distribution — Geometric distribution — Binomial distribution — Negative binomial distribution — Poisson distribution — Milestone 2 due

Week 6 (14 Feb, 16 Feb) Foundations for inference — Point estimates and sampling variability — Confidence intervals for a proportion — Hypothesis testing for a proportion

Week 7 (21 Feb, 23 Feb) Inference for categorical data — Inference for a single proportion — Differences of two proportions — Testing for goodness of fit using chi-square — Testing for independence in two-way tables — Exercise 1 due

Week 8 (28 Feb, 2 Mar) Inference for numerical data — One-sample means with the t-distribution — Paired data — Difference of two means — Power calculations for a difference of means — Comparing many means with ANOVA

Week 9 (7 Mar, 9 Mar) Introduction to linear regression — Fitting a line, residuals, and correlation — Least squares regression — Types of outliers in linear regression — Inference for linear regression — Exercise 2 due

Spring Break

Week 10 (21 Mar, 23 Mar) Multiple and logistic regression — Introduction to multiple regression — Model selection — Checking model conditions using graphics — Multiple regression case study: Mario Kart — Introduction to logistic regression — Milestone 3 due

Week 11 (28 Mar, 30 Mar) More on R and the tidyverse

Week 12 (4 Apr, 6 Apr) In class work on milestone 4 and exercise 3 — Milestone 4 due

Week 13 (11 Apr, 13 Apr) Typical UX experiments

Week 14 (18 Apr, 20 Apr) Typical data science experiments — Exercise 3 due

Grading

I plan to grade assignments within two weeks of their due date except where circumstances interfere. The grading scale used along with the grade components follow.

- A >= 94.0%
- A- >= 90.0% & < 94%
- B+ >= 87.0% & < 90%
- B >= 83.0% & < 87%
- B- >= 80.0% & < 83%
- C+ >= 77.0% & < 80.0%
- C >= 73.0% & < 77.0%
- C- >= 70.0% & < 73.0%
- D >= 60.0% & < 70.0%
- F < 60.0%

Group work: 4 milestones, each 10 points

- Milestone 1: description (tables, summary stats)
- Milestone 2: description (visualization)
- Milestone 3: regression
- Milestone 4: regression diagnostics

Individual work: 3 assignments, each 10 points + 1 exam 30 points

- Exercise 1: description (tables, summary stats, visualization)
- Exercise 2: regression
- Exercise 3: regression diagnostics
- Final exam: take-home, due at official exam time

Attendance

I will take attendance every day and your final grade for the class will be dropped one-half letter grade if you are only present for 70 to 80 percent of classes. It will be dropped a full letter grade if you are only present for 60 to 70 percent of classes. If you are present less than 60 percent of the classes, your final grade will be dropped by two letter grades and it will be difficult for you to achieve a passing grade so, in that case, you should drop the class.

If you have a legitimate need for absence, such as illness or job interview, notify the instructor by email as soon as possible and you may receive an excused absence.

POLICIES

Accessible, Inclusive, and Compliant Statement

The university is committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state law. Please let me know if you experience any barriers to learning so I can work with you to ensure you have equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, and need accommodations please contact Disability and Access (D&A). Please refer to D&A's website for contact and more information: http://diversity.utexas.edu/disability/. If you are already registered with D&A , please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations and needs in this course.

Policy on Academic Integrity

Students who violate University rules on academic misconduct are subject to the student conduct process and potential disciplinary action. A student found responsible for academic misconduct may be assigned both a status sanction and a grade impact for the course. The grade impact could range from a zero on the assignment in question up to a failing grade in the course. A status sanction can range from probation, deferred suspension and/or dismissal from the University. To learn more about academic integrity standards, tips for avoiding

a potential academic misconduct violation, and the overall conduct process, please visit the Student Conduct and Academic Integrity website at: http://deanofstudents.utexas.edu/conduct.

Class Recordings

Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

Personal Pronouns

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender identity & expression, and nationalities. Class rosters are provided to the instructor with the student's legal name, unless they have added a "chosen name" with the registrar's office, which you can do so here: https://utdirect.utexas.edu/apps/ais/chosen_name/. I will gladly honor your request to address you by a name that is different from what appears on the official roster, and by the pronouns you use (she/he/they/ze, etc). Please advise me of any changes early in the semester so that I may make appropriate updates to my records. For instructions on how to add your pronouns to Canvas, visit https://utexas.instructure.com/courses/633028/pages/profile-pronouns. More resources available on the Gender and Sexuality Center's website, https://www.utgsc.org.

Basic Needs Security

Any student who faces challenges securing their food or housing and believes this may affect their performance in the course is urged to contact the Dean of Students for support. UT maintains the UT Outpost (https://deanofstudents.utexas.edu/emergency/utoutpost.php) which is a free on-campus food pantry and career closet. Furthermore, please notify the professor if you are comfortable in doing so. This will enable him to provide any resources that he may possess.

Mental Health Information

I urge students who are struggling for any reason and who believe that it might impact their performance in the course to reach out to me if they feel comfortable. This will allow me to provide any resources or accommodations that I can. If immediate mental health assistance is needed, call the Counseling and Mental Health Center (CMHC) at 512-471-3515 or you may also contact Bryce Moffett, LCSW (iSchool CARE counselor) at 512-232-2983. Outside

CMHC business hours (8am-5pm, Monday-Friday), contact the CMHC 24/7 Crisis Line at 512-471-2255.

References

Diez, David, Mine Çetinkaya-Rundel, and Cristopher D Barr. 2019. *OpenIntro Statistics, Fourth Edition.* self-published. https://openintro.org/os.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2021. *An Introduction to Statistical Learning.* Springer New York.

Kuhn, Max, and Julia Silge. 2022. *Tidy Modeling with R.* Sebastopol, CA: O'Reilly.

Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* 2nd ed. O'Reilly Media, Inc. https://r4ds.hadley.nz/.