IS542AO
Data, Statistical Models and Information
Spring 2020
Wednesdays, 6:00pm-8:00pm,Online

Instructor: Jill Naiman
Office hours: <TBD>
Email: jnaiman@illinois.edu

Course Description
An introduction to statistical and probabilistic models as they pertain to quantifying information, assessing information quality, and principled application of information to decision-making. The increasing prevalence of massive data sets and falling computational barriers have rendered statistical modeling an integral part of contemporary information management. With this in mind, this class prepares students to select and properly undertake commonly encountered modeling tasks. The course reviews relevant results from probability theory, emphasizing the merits and limitations of familiar probability distributions as vehicles for modeling information. Subsequent consideration includes parametric and non-parametric predictive models, as well as a discussion of extensions of these models for unsupervised learning. Throughout these discussions, the course focuses on model selection and gauging model quality. Applications of statistical and probabilistic models to tasks in information management (e.g. prediction, ranking, and data reduction) are emphasized.

Learning Objectives
Students will demonstrate an understanding of probability theory and statistical learning by building and evaluating models of a diverse range of data sets. By the end of the course students will have basic concepts of what constitutes a “good” statistical question, what one can feasibly learn and predict with data, and an overview of toolsets and methods to answer elementary statistical questions. In particular, each student will be able to:

- Articulate the role of marginal, joint, and conditional probability in modeling processes involving information.
- Select, parameterize, and compare probability distributions as vehicles for modeling information.
- Specify, estimate and evaluate elementary parametric statistical models.
- Specify, estimate and evaluate elementary non-parametric statistical models.
- Articulate professional responsibilities with respect to creating, describing and using models built from data.

Pre- and Co-requisite
IS452 Foundations of Information Processing is strongly recommended as a prerequisite. A highly motivated student could pass this course without IS452 (so the prerequisite is not enforced), but programming will not be covered in this course. Students who have not completed an introductory course on statistics will need to come up to speed quickly on material covered early in the semester.

Required Texts
James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013) *An Introduction to Statistical Learning.* New York: Springer. [**abbreviated ISL**]
http://www-bcf.usc.edu/~gareth/ISL/

Diez, D., Barr, C., and Cetinkaya-Rundel, M. (2015) *OpenIntro Statistics* Third Edition, [available online, https://www.openintro.org/stat/textbook.php?stat_book=os, **abbreviated OIS**]

Venables, W.N., Smith, D.M and the R Core Team (2012) *An Introduction to R*. [available online, http://cran.r-project.org/doc/manuals/R-intro.pdf, **abbreviated ITR**]

Supplemental Texts
Johnson, R. A. (2009) *Statistics: Principles and Methods.* New York: Wiley. [available on reserve]

Manning, C. D., and Schütze, H. (1999) *Foundations of Statistical Natural Language Processing.* Cambridge, MIT Press. [available on reserve]

Maindonald, J., *Using R for Data Analysis and Graphics - Introduction, Examples and Commentary* [available online, https://cran.r- project.org/doc/contrib/usingR.pdf]

Paradis, E., *R for Beginners* [available online, https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf]

Course Schedule

Note: the precise scheduling of topics is likely to change based on student feedback and interests.

Week	Topic	Reading
1	● Data, Models, and Information ● Elementary statistics: Definitions ● Overview of R	OIS 1 (ISL 1)
2	● Elementary statistics: Applications & Plots	OIS 1 (ISL 1)
3	● Introduction to data analysis with R ● Review of tabular and graphical displays of data	ITR 1, 2, 5, 6, 7, 12
4	● Random variables: expectation and variance ● Joint and conditional probability ● Bayes rule	OIS 2
5	● Random variables: distributions (normal, binomial, poisson)	OIS 3
6	● Modeling data with probability distributions ● Foundations for inference	OIS 4

7	• Inference for numerical data • Inference for categorical data	OIS 5, OIS 6
8	• Linear regression	OIS 7 (ISL 3)
9	• Multiple linear regression	OIS 8 (ISL 3)
10	• Logical regression	OIS 8 (ISL 4)
12	• *k*-Nearest neighbor classification and regression	ISL 2.2.3, 4.6.5
13	• Intro to Unsupervised linear models: Principle component analysis	ISL 10.0-10.2
14 15	• Case studies • Case studies and review	NA NA

Assignments and Methods of Assessment:

Assignment

1. Weekly homework	50%
2. Midterm exam	15%
3. Final exam	25%
4. Class engagement	10%

Class Participation Policy

Leaders are expected to clearly articulate issues and problems and how analytical tools can help. The way we foster this in the course is that you must participate in the classroom discussion. You are not required to "speak up" during every class session, but you do need to attend and contribute to the class and/or forum discussion over the semester.

Leaders are also expected to foster productive environments for those around them. Those in this class come from a variety of backgrounds and comfort levels with the material and programming - it is expected that all students (and instructor!) will remain cognizant of this fact at all times and any demeaning language or behavior will not be tolerated.

Attendance Policy

Enrollment in this course includes an expectation of regular attendance. If you find you must miss class, contact me as soon as possible to inquire about make- up exercises in lieu of attendance. Students missing more than one class session or who regularly arrive late or leave early will not pass the class unless alternate arrangements have been made with the instructor. According to University policy: "For a graduate level course, attendance is expected, and should not be counted toward the final grade. The Student Code explicitly states that for all students, "(a) Regular class attendance is expected of all students at the University" (http://admin.illinois.edu/policy/code/article1_part5_1-501.html)

Exam and Homework Policy

We do not accept late homework, however we will drop your lowest homework grade when we calculate your final grade. All work must be your own, including all code turned in for assignments or exams.

All assignments and exams, including code and data, must be uploaded as files in Moodle. Your name should appear in the files and the file names as follows lastname-first-module.ext (e.g, naiman-jill-assignment1.pdf). The submission must include:

1) A narrative document as a PDF file (to be read by a human). To preserve the natural flow of the narrative, figures (e.g., screenshots, code snippets) and tables should be embedded into the document near their first mention. Any supplementary files containing R programs or data should be referenced in the text and separately uploaded.

AND
2) All R code as separate files with an .R extension (to be read by a computer).

Academic Integrity
Please review and reflect on the academic integrity policy of the University of Illinois, http://admin.illinois.edu/policy/code/article1_part4_1-401.html to which we subscribe. By turning in materials for review, you certify that all work presented is your own and has been done by you independently.
If, in the course of your writing, you use the words or ideas of another writer, proper acknowledgement must be given. Not to do so is to commit plagiarism, a form of academic dishonesty. If you are not absolutely clear on what constitutes plagiarism and how to cite sources appropriately, now is the time to learn. Please ask me!
Please be aware that the consequences for plagiarism or other forms of academic dishonesty will be severe. Students who violate university standards of academic integrity are subject to disciplinary action, such as a reduced grade, failure in the course, or suspension or dismissal from the University.

Statement of Inclusion
http://www.inclusiveillinois.illinois.edu/chancellordivstmtswf.html - ValueStmt

As the state's premier public university, the University of Illinois at Urbana- Champaign's core mission is to serve the interests of the diverse people of the state of Illinois and beyond. The institution thus values inclusion and a pluralistic learning and research environment, one which we respect the varied perspectives and lived experiences of a diverse community and global workforce. I support diversity of worldviews, histories, and cultural knowledge across a range of social groups including race, ethnicity, gender identity, sexual orientation, abilities, economic class, religion, and their intersections.

Disability Statement

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak St., Champaign, call 333-4603 (V/TTY), or e-mail a message to disability@uiuc.edu.