STAT 411/616: Advanced Statistical Methods

Date: Spring 2019; Course 20318/22510
Time: TTh 1:00 -2:15 pm
Location: Duncan Hall DCH 1064

Instructor

Prof. John A. Dobelman
dobelman@stat.rice.edu
Duncan Hall 2100: 713 348 5681
Office Hours: By appointment.

Course Website

http://dobelman.rice.edu (see courses)
We will also be using Canvas

Teaching Assistant (TA)

Huiming Lin
hl68@rice.edu
Duncan Hall 2091:
Office Hours: Mon 2-3 pm or by appointment

Every.Sentence.Here.Is.Important™

Syllabus Outline

Course and Outcomes

This course is topical from semester to semester. This semester we will cover study design, ANOVA modeling and some computational methods for inference.

Study design addresses the most important part of a statistical study, namely, how the data are collected. Indeed, the maxims, "bad data in, bad data out" and "garbage in, garbage out," make clear that if careful attention is not given to how one collects data, then there is little to be done to obtain valid statistical results (e.g., statistical summaries, p-values, confidence intervals, model fits, etc.). As part of the story we also distinguish between observational studies and experiments; the former are popular in many disciplines (especially sociology and epidemiology), but they can only point to association while experiments may point to causation.

A prerequisite to STAT 411 is linear regression. We will extend linear regression to linear models, including ANOVA (factorial) models, random effects and mixed effects. Generalized linear models (GLM) will be introduced through binomial and Poisson regression to model

counts. For example, we may want to develop predictive models for a certain disease based on a number of predictors such as age, sex, and BMI.

We also consider some inferential methods that are made possible by modern computing power. The bootstrap is a methodology that is typically used to estimate standard errors or confidence intervals, especially in nonparametric settings. Permutation methods are useful for generating reference distributions in hypothesis testing. The Expectation-Maximization (EM) algorithm is a method for finding maximum likelihood estimates and their standard errors. When many hypothesis tests are evaluated, traditional error rates can be misleading and multiple testing corrections can be applied to obtain more valid decisions.

Prerequisites: STAT 310/312 or ECON 307/382, Probability and Statistics and either STAT 410 Linear Regression or STAT 615 Regression and Linear Models. Familiarity with matrix analysis or linear algebra is strongly recommended but not required. However, the course will make use of aspects of linear algebra that were covered in STAT 410 and/or STAT 615.

Course Text

Given the nature of the course, there are no required textbooks. Notes or references to course material will be given throughout the semester; these are in the Canvas Files areas (Reference & Texts or Handouts). Additionally, we have compiled a list of good source materials and some standard textbooks that cover various parts of the course. This reading list is available at ReadingList.411_616.pdf.

Course Content

Topics/chapters include

- Likelihood and sometimes vaguely understood math/stat and inferential concepts
- Bootstrap and Sampling
- ANOVA
- Alternative regression models (Things they don't tell you about regression)
- Generalized Linear Models (GLM)
- Experimental design
- Useful nonparametric statistics and hypothesis testing
- Simulation
- Statistical process control (SPC)
- Survival analysis with explanatory variables

Course Policies

Undergraduate/Graduate Course Designation

The STAT 616 designation is intended for graduate students duly admitted into a graduate course of study at Rice University. In certain exceptional cases, undergraduates may petition to take the STAT 616 version, but GPA and other criteria will be applied. STAT 616 requires additional, more theoretical coursework and will be evaluated at a graduate level.

Grading

Grading for this course will consist of homework exercises from the text (45%), graded examinations (45%), and class participation (10%). Homework and exams are all equal weighted within their category. Students taking the graduate-level STAT 616 will have additional assignments, with the same grading breakdown as STAT 411. Note that an "A" average in classwork can be offset by low class participation grades!

Professional Standards (Lite)

All assignments, with exception of handwritten math proofs and things of that nature, must be prepared nicely (like this syllabus, *más o menos.*) This means you need to organize it neatly with appropriate layout, graphics, citations, etc. Unnecessary printouts of numbers are not acceptable, nor are meaningless digits of precision, etc. The font should be large enough for management to easily read (no less than 10-point font, preferably 12).

You should include a cover page indicating your name, the class, assignment, and date submitted. In the real world, if your management/customers or clients cannot read the report, or if it is filled with superfluous information, they will just return it to you with a request to resubmit. If you are unlucky, they will just pass you over for someone who does these things better.

You are encouraged to make appointments with the peer consultants at the Center for Written, Oral, and Visual Communication for your assignments in this course. These consultants do not proofread or edit your work, but they will provide feedback on topics such as the organization of your paper or presentation, the coherence of your argument, appropriate sentence structure, and grammatical errors. You can make an appointment at the Center's website: http://cwovc.rice.edu/.

Originality

Although some assignments are not "pledged", keep in mind that plagiarism is a serious problem, and is especially problematic for the student or researcher. Plagiarism will be treated, for pledged assignments, in accordance with the honor code provisions; and for non-pledged assignments, your paper will be returned for a rewrite with the automatic deduction of one letter grade. By plagiarism we mean "quoting, paraphrasing, or otherwise using another's words or ideas as one's own without properly crediting the source." Unfortunately, with the prevalence of internet sources, it is sometimes easy to inadvertently commit plagiarism. In order to avoid this problem, you should consult this white paper http://www.stat.rice.edu/~dobelman/courses/Plagarism.Hewitt.2016.pdf and also familiarize yourself with http://futureowls.rice.edu/futureowls/Honor_Code.asp.

Citation

Any cited works in your reports must have citations listed, either as a separate section, or (less desirably) as footnotes. Exact citation format is your choice[1], but be consistent.

Use of Canvas

The Canvas system is the course management tool for announcements, assignments, resources, etc. ***Do not email the instructor/TA*** questions about the course or assignments, but rather post as a discussion on Canvas so that all can see the conversation. Any such emails will be ignored. If online submission for assignments is required, please upload in the appropriate area (usually in the assignments section).

Assignment Submission, Lateness and Grading Policy

All assignments must be submitted by the date due as **hard copy** to the instructor or designee. Assignments will be collected at the beginning of class on the due date. Late penalties begin after the assignment is collected. Hard copy papers must be properly bound; only properly bound assignments will be accepted. Do not use a paper clip. The

---

[1] See for example https://owl.english.purdue.edu/owl/section/2/

instructors/TA will not be able to bind or print out e-mailed or online-only assignment submissions. Soft copy postings may be made as of COB the day the assignment is due. Both hard and soft copies of the assignments **are required in order to receive credit** for the assignment.

Late Homework Policy: You may submit late homework through 3 calendar days after the due date. Each day will incur a 10-percentage point penalty. You need not inform me that your homework will be late, but you must submit your late homework to me in person. The late homework policy is in place to allow you some flexibility in submitting homework; special permission to submit late homework without penalty is not allowed. Plan travel, job interviews, and other such activities accordingly. If you are away or can't make class, you may submit your homework via email (before the due date and time) or have someone submit it for you. Late penalties begin after the assignment is collected in class; if I collect homework at 1:10pm, then late homework starts at 1:11pm. Please be on time. All issues regarding homework will be addressed on an individual student basis and the final decision rests with the instructor. Please do not ask to submit homework without penalty; this is the only way I know to treat all circumstances equally.

Laptops and Wearable/Portable Electronic Devices (PED)

Do not use laptops, tablets or cell phones in class unless so requested by the instructors; they are a distraction to other students. Prohibited devices include laptops, earbuds, gaming devices, mp3/music/media players, cell/smart-phones, PDA's, Kindle/e-book readers, tablet computers, Apple/Google watch, i* devices, multi-purpose wrist communicators, cameras, GPS/GIS devices, Google glasses, VR headsets, etc. Hearing aids are acceptable to use in accordance with the last item (disabilities) and procedures herein. From time to time the instructors may ask a student to look something up, but in most cases this is not necessary. If the student requests, such devices may be permitted, if the purpose is clearly articulated in advance. The student will be asked to put away their PED's if they are taken out in class; upon the second request, the student will be dismissed from class.

Graded Material

Homework

You can expect 5-7 assignments, which will include both theory and data applications. You are allowed to work with other students but your submitted homework must reflect your own thinking and expression. Complete all computational work with appropriate statistical software (this will usually be the R programming language). Use only results you need from your computer work; do not cut and paste all output into your homework submissions; use just those results that answer the questions.

Examinations

There will be two in-class exams, each covering half of the course. There is no comprehensive final exam. Exams are usually closed book, with permitted calculators, etc., under pledged conditions; NO INTERNET is permitted. Students are allowed one 8.5"x11" (or smaller) handwritten note sheet per exam.

Attendance

Students are expected to attend class. Much material is presented in class which might not otherwise be in texts, notes, etc. Attendance will be reflected in the Class Participation portion of the course grade. Although we plan on keeping the course website up to date, if a student

misses a class, then s/he is responsible for keeping up with the course material and finding out if any exams, quizzes, or homeworks have been assigned or scheduled. Similarly, important due date changes might sometimes be made in class to your benefit which might not be immediately posted on Canvas.

Other Provisions

Software

It is impossible to perform statistical/quantitative data analysis without some sort of computer software, and it is expected that the student will become proficient with one or more statistical software packages. The most widely used data analysis software in the real world today is Microsoft Excel, and its capabilities are impressive. However, more specialized software is sometimes needed, such as Matlab, R/S-Plus, SAS, Python, SQL, Resampling Stats, SPSS, Stata, StatTools, StatExact, Lisrel, @Risk, Maple, Mathematica, C-Plex, etc. Of these, R and Python have become the choice of many because of their relative ease of use and low cost (FREE!) Download information for some of these packages is available on the course website(s).

One can also program most statistical procedures in programming language such as Java, C/C#/C++, Fortran, VB, etc., along with specialized add-in routine libraries, but these require a lot of work to code and debug. Additionally, you will find that most corporate employers will not be paying for the nice software that is available for you here on campus. Consequently, to enhance your value to your future employer, we suggest that you become proficient in Excel, R/Python and SQL/SAS. Althought this course usually requires the use of R, like in the real world, you may need to use several in the production of your reports.

Note that a good word processor will also be required in order to prepare reports and presentations; Microsoft Word, OpenSource, LaTeX, etc., are candidates for use in preparing these documents, although use of LaTeX in the business world is sparse.

Data Access

Many assignments are data-oriented. In many cases the data is provided as part of the assignments; in others, the student will have to find data on their own. This is an important skill and will be used repeatedly.

Rice Honor Code

Before enrolling in this course, you must understand and agree to abide by the Honor System in place at Rice University which protects the academic integrity of all coursework. All students (including graduate students) at Rice are bound by the Rice Honor Code. The Honor Code is a unique feature at Rice, one that is valued highly, and is of profound importance. New students should familiarize themselves with the Honor Code before starting classes. Honor Code violations are very serious, and can lead to dismissal from the University. Suspected violations will be processed in accordance with (http://honor.rice.edu).

Examinations are conducted under pledged conditions. Proper recitation and use of the Honor Pledge on examinations will be required to avoid a possible penalty of up to 5% of the test grade. Note that the use of prior years' and other solutions to pledged exercises is considered UNAUTHORIZED AID and is not permitted under the Rice Honor code.

If homework and individual assignments/projects may be worked on with other class members, each student must submit their own work for credit. Homeworks should be submitted by each person, but you should indicate with whom you worked when applicable. No direct copying is allowed. Group projects and assignments may be submitted by the group, per submission requirements above.

Sponsored Message Regarding Student Responsibility:

The Committee on Examinations and Standing has asked that we reiterate the responsibilities of the student to comply with deadlines affecting their status and standing. Essentially this means that you need to be aware of the deadlines for which you alone are responsible (i.e., not your advisor, etc.). This data is maintained on http://registrar.rice.edu/calendars/.

Religious Holidays

If you wish to be excused from class on any of these days which do not coincide with University holidays already on the official academic calendar, you must notify the instructor in writing no later than the second week of class. Absence from classes or examinations for religious reasons does not relieve students from responsibility for any part of the course work required during the period of absence

Disabilities

Any student with a documented disability wishing academic adjustments or accommodation is required to speak with the instructor about it during the first two weeks of class. All discussions will remain confidential. Additionally, you must make sure this documentation is on file with Disability Support Services (Allen Center, Room 111 | adarice@rice.edu | x5841) in order to register your disability and to determine the accommodations you need. The instructor cannot be make accommodation without the appropriate letter from Rice DSS. Additional program and documentation requirements and responsibilities are spelled out at http://dss.rice.edu/.

Changes to Syllabus

Changes to this syllabus may be published from time to time, with notice and explanation given in class and via Canvas (if used).

Rev: 1/4/2018