

Syllabus

Instructor

Andy Reagan (areagan@smith.edu, Bass 218). My office hours will be held after class Tuesdays and Thursdays Noon - 12:30, and by appointment, in Bass 218 or online.

Description

In this course students will learn how to use statistical models to analyze real-world data. Students will work through examples using statistical software and practice using linear models where the response variable is either quantitative or categorical, and the explanatory variables are quantitative or categorical or of both types. They will also consider logistic models for binary response data. This course will emphasize the importance of communicating technical concepts to a possibly non-technical audience with clarity and precision.

Perspective

This is an applied statistics course. As such, the course work emphasizes using statistical software on the computer to perform data analysis. Homework and examinations will focus primarily on applying the methods learned in class to real data and interpreting the results. One of the goals of this course is for you to be able to clearly communicate the results of your data analysis in non-technical language in both written and oral forms.

Prerequisites

A college-level introductory statistics course or a score of 4 or 5 on the AP Statistics Exam or permission of the instructor. Students need to be familiar with basic statistical concepts including graphical and numerical descriptive statistics, the normal and t -distributions, hypothesis tests, and confidence intervals.

Textbook

(required) *STAT 2: Building Models for a World of Data*, Preliminary Edition v.2 (August 2010), by Cannon, Cobb, Hartlaub, Legler, Lock, Moore, Rossman, and Witmer. The book is available at the Greccourt bookstore in the basement of the Campus Center for about \$150.

Classes

We meet on Tuesdays and Thursdays from 10:30-11:50 AM. Class time will consist of a mixture of lectures, lab exercises, quizzes, discussing material in the textbook, homework, and projects. This is a hands-on course, where you will learn by doing, so it is imperative that you, unless otherwise instructed, bring your book and a computer to class.

Policies

Attendance

Your attendance in class is crucial, as is your punctuality. We are all going to learn this material together, so we need to have everyone present and working. I will make accommodations for an unavoidable absence if you notify me. Our Honor Code means that you will be the judge of whether or not an absence was unavoidable. (For instance, staying in bed because you had the flu would be an unavoidable absence, but oversleeping because you stayed up late to write a paper would be an avoidable absence.) One necessary absence during the semester is not unusual; having more than two is uncommon.

Collaboration

Much of this course will operate on a collaborative basis, and you are expected encouraged and to work together with a partner or in small groups to study, complete homework assignments, and prepare for exams. However, every word that you write must be your own. Copying and pasting sentences, paragraphs, or blocks of R code from another student is not acceptable and will receive no credit. All students are bound by the Smith College Honor Code.

Academic Honor Code Statement

Smith College expects all students to be honest and committed to the principles of academic and intellectual integrity in their preparation and submission of course work and examinations.

Students and faculty at Smith are part of an academic community defined by its commitment to scholarship, which depends on scrupulous and attentive acknowledgement of all sources of information, and honest and respectful use of college resources.

Assignments

Homework

Problems (mainly from the book) will be assigned each week, and posted on Moodle. The homework and the project are the backbone of this course and must be completed satisfactorily to receive an A or B grade in the course. Homework will be graded, and you will be occasionally be asked to present your solution to a homework problem to the class. Homework must be handed in before class starts in the period in which it is due.

Homework should be completed in RMarkdown. All assignments will be due by 5pm of the due date. A hard copy of the knitted file should be submitted. The RMarkdown file should also be submitted via Moodle. Late homework turned in within 48 hours will be accepted with a 25% penalty. No credit will be given after 48 hours unless you are in extenuating situations with verifiable documents.

Although you are encouraged to discuss problems with each other, I expect each person to submit his or her own work. You may choose to work with a partner on data analyses — that's okay, but I want written interpretations and responses to be in your own words. You should always write up your own solutions. It is not acceptable to copy down your classmates' work and in fact, doing this constitutes academic dishonesty.

And remember, an answer to a statistics question is almost never just a number — I want thoughtful explanations and conclusions! In order to receive full credit on a problem, you will need to provide clear explanations and necessary steps to justify your answers. If your work is mostly correct, you will receive most credits. On the other hand, if you do not provide enough work, or follow wrong steps but happen to get a correct answer in the end, then it is likely that you receive no credit or very few partial credit.

Final Project

Your major achievement in this class will be to conduct (in groups of 3) a statistical investigation on a question of interest to you. Rather than collect primary data, we will use data available on the Internet or from faculty research. You will prepare a project proposal describing your study and obtain approval from me before you begin the investigation. Please see the attached schedule for due dates for intermediate parts of the project. During the last week of class, you (and your group) will give a 10 minute oral presentation of your study. The project will give you experience planning a statistical study, acquiring data, creating and testing a linear model, and writing a technical report. We will spend time in class looking at what data is available on the web and about writing a project proposal.

Exams

Two midterm examinations will be given, the first on Thursday, Oct. 18 and the second on Thursday, Dec. Several 10-minute quizzes will be given throughout the term. Quizzes will be easy if you have been keeping up with the reading and the homework. There will be no make-up quizzes but your lowest quiz score will be dropped. Project presentations during the last week of classes will take the place of a final exam.

Grading

Course grades will be a weighted average of the grades on your project, tests, homework, and quizzes. Your participation score will include attendance and participation in class discussions.

Project & Presentation	30%
Homework	25%
First exam	20%
Second exam	20%
Participation	5%

Resources

Course Website

Course information, including this syllabus, will be available on the course Moodle site. I will update information posted there regularly with class handouts, homework, project assignments, and announcements.

Computing

We will use the R statistical software package extensively, and exclusively. R is open source software that is available for free on Mac, Windows, and Linux operating systems. We also have an RStudio server set up (<http://rstudio.smith.edu:8787>), which you are strongly encouraged to use. Most homework assignments will require data analysis using the computer and short written reports of your findings. Documentation for R is online and I will give an introduction to the software.

Writing

Your ability to communicate results, which may be technical in nature, to your audience, which is likely to be non-technical, is critical to your success as a data analyst. The assignments in this class will place an emphasis on the clarity of your writing. The Jacobson Center for Writing (Seelye 307) offers a variety of services designed to help students improve drafts of their papers.

Extra Help

There are Stat TAs available from 7 to 9 pm on Sunday – Thursday evenings in McConnell 301. Although this course may be beyond the scope of what the TAs are prepared to handle, they should be able to help you review material from your intro course, or with R. In addition, the Spinelli Center for Quantitative Learning (McConnell 301, <https://www.smith.edu/qlc/tutoring.html>) supports students doing quantitative work across the curriculum, and has a Statistics Counselor available for appointments. Your fellow students are also an excellent source for explanations, tips, etc.

Schedule

Tentative Schedule

The following outline lists each class date and gives the topic that will be discussed in that class. The reading assignment from the textbook is also given for each class date. Please complete the reading assignment *before* coming to class so that you can participate fully in the discussion. I reserve the right to revise this schedule — in particular whether Mountain Day falls on a class day — updates will be posted on Moodle.

N	Date	Day	Topic	Reading	Due
1	Sep-06	Th	Intro, review terms and notation. Four-step process		
2	Sep-11	T	SLR: model, estimation, conditions	0, 1.1-1.3	
3	Sep-13	Th	SLR: transformations, outliers, influence	1.4-1.6	HW 1
4	Sep-18	T	SLR: inference for slope, ANOVA table	2.1-2.2	HW 2
5	Sep-20	Th	SLR: finishing SLR, more R demos		
6	Sep-25	T	MLR: correlation, conf. & prediction intervals	2.3-2.5	HW 3
-	Sep-27	Th	Mountain Day!	3.1-3.2	
7	Oct-02	T	MLR: comparing two regression lines	3.3	HW 4
8	Oct-04	Th	MLR: second-order models	3.4	
-	Oct-09	T	Fall Break		HW 5 due Oct 10.
9	Oct-11	Th	MLR: multicollinearity; F-tests & partial SS	3.5-3.6	Proposal Outline
10	Oct-16	T	MLR: Review and Summary	3.7-3.8	HW 6
-	Oct-18	Th	First Exam		
11	Oct-23	T	MLR: nested (partial) F tests; added variables plots	4.1	
12	Oct-25	Th	MLR: subset selection; leverage and influence	4.2-4.3	Final Proposal
13	Oct-30	T	MLR: coding categorical predictions	4.4	HW 7
14	Nov-01	Th	MLR: randomization tests	4.5	
15	Nov-06	T	MLR: bootstrap for regression (with Professor Kim)	4.6	HW 8 <i>due Nov-09</i>
16	Nov-08	Th	AOV: 1-way model, assessing model	5.1-5.2	
17	Nov-13	T	AOV: inferences	5.3	HW 9 due Wed, Progress Report
18	Nov-15	Th	AOV: multiple comparisons issues, LSD	5.4	
-	Nov-20	T	Thanksgiving	9.1-9.2	Data Due
-	Nov-22	Th	Thanksgiving		HW 10 due following Mon Nov-26
19	Nov-27	T	LOG: odds ratios & assessing model; 2 x 2 tables	7.2, 9.3-9.5	
20	Nov-29	Th	LOG: multiple predictors, LRT test	10.1-10.3	
21	Dec-04	T	End of Semester Wrapup	11.1	HW 11 due Wed Dec-05
-	Dec-06	Th	Second Exam		
-	Dec-11	T	Project Presentations		
-	Dec-13	Th	Project Presentations		Final Paper