SDS 291 (Multiple Regression) Syllabus

About the Course

Instructor

- Amelia McNamara (`amcnamara@smith.edu`, Burton 215, 413-585-3851).
- Standard office hours: Tuesdays from 2:00-3:30 pm, Fridays from 10:30 am-12:00 pm, and by appointment, in Burton 215.
- R-focused office hour: Mondays from 4:00-5:30 pm in Sabin Reed 301.

Description

Same as MTH 291. Formerly MTH 247. Theory and applications of regression techniques; linear and nonlinear multiple regression models, residual and influence analysis, correlation, covariance analysis, indicator variables and time series analysis. This course includes methods for choosing, fitting, evaluating and comparing statistical models and analyzes data sets taken from the natural, physical and social sciences.

Website

The course website and Moodle will be updated regularly with lecture handouts, project information, assignments, and other course resources. Homework and grades will be submitted to Moodle. Please check both regularly.

Prerequisites:

- an introductory statistics course (e.g. MTH/PSY 201, GOV 190, MTH 219 (formerly MTH 241), MTH/SDS 220 (formerly MTH 245), ECO 220, or the equivalent or a score of 4 or 5 on the AP Statistics examination)

Textbooks

One textbook is required:

- *STAT 2: Building Models for a World of Data*, by Cannon, Cobb, Hartlaub, Legler, Lock, Moore, Rossman, and Witmer. W.H. Freeman and Company, 2012. Note that you do **not** need the online access!

Another (slightly more mathy) book suggested as a supplementary text or reference:

- *A Modern Approach to Regression with R*, by Sheather, Springer, 2012.

Classes

Attendance is recommended, but not required. You are an adult and make your own priorities. If you are not in class, I will assume you understand the material that was covered. In the case where you may need an extended absense and feel it will impact your learning (e.g. an illness, death in the family, conference, etc), please let me know so we can find a way for you to make up the material.

Policies

Respect

Please plan to treat me and your classmates with respect. This includes things like arriving at class on time, coming in quietly if you are late, and focusing on the task at hand.

Collaboration

Much of this course will operate on a collaborative basis, and you are expected and encouraged 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 code from another student is not acceptable and will receive no credit. No interaction with anyone but the instructors is allowed on any exams or quizzes. All students, staff and faculty are bound by the Smith College Honor Code, which Smith has had since 1944.

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.

Cases of dishonesty, plagiarism, etc., will be reported to the Academic Honor Board.

Assignments

1. Homework [20%]: There will be several problem sets over the course of the semester. Problem sets will involve computational assignments in R with written explanations. In order to streamline your workflow and my ability to comprehend your work, you should complete all of your homework assignments in R Markdown. All homework will be submitted electronically by midnight (11:55 pm) of the due date. Late assignments will lose points at the rate of 20% per day.

2. Quizzes [5%]: There will be bi-weekly reading quizzes on Moodle. These will provide an opportunity for you to get instant feedback on what you are learning.

3. Project & Presentation [25%]: You will work on a term project in a group of three over the course of the semester. This is an opportunity for you to demonstrate your understanding of the material and put it into practice. More details about the project will follow.

4. Participation [5%]: Engagement with your group work (both in class and during the project). You are expected in fully participate. If you do so, you should recieve full credit for this portion of the grade.

5. Exams [45%]: There will be **two**, self-scheduled, closed-book exams. You may bring a calculator and one piece of paper of hand-written notes (double-sided). After the exam, you have the opportunity to raise your exam score by submitting corrections for any questions you got wrong. Smith College has had an academic honor code since 1944, and all students, staff and faculty are bound by this code. Cases of dishonesty, plagiarism, etc., will be reported to the Academic Honor Board.

Grading

When grading your written work, we are looking for solutions that are technically correct and reasoning that is clearly explained. Explanation and context of an answer are the most important component. Numerically correct answers, alone, are not sufficient on homework, tests or quizzes. Neatness and organization are valued, with brief, clear answers that explain your thinking. If I cannot read or follow your work, I cannot give you full credit for it.

Resources

Computing

The use of the `R` statistical computing environment with the RStudio interface is thoroughly integrated into the course. You have two options for using `RStudio`:

- The **server** version of `RStudio` on the web. The advantage of using the server version is that all of your work will be stored in the cloud, where it is automatically saved and backed up. This means that you can access your work from any computer with a web browser and an internet connection. The downside is that there are only 2 CPUs and 8 GB of RAM allocated to this machine, and you have to share those resources with each other and with all of the MTH220 students!
- A **desktop** version of `RStudio` installed on your machine. Your laptop likely has at least 2 CPUs and at least 4 GB of RAM. The downside to this approach is that your work is only stored locally, but I get around this problem by keeping all of my work in a Dropbox folder.

Note that you do not have to choose one or the other – you may use both. However, it is important that you understand the distinction so that you can keep track of your work. Both `R` and `RStudio` are free and open-source, and are installed on most computer labs on campus. Please see the Resources page for help with `R`.

Unless otherwise noted, you should assume that it will be helpful to bring a laptop to class. It is not *required*, but since there are only three workstations in the classroom, we will need a critical mass (i.e. at least $n/2$) computers in the classroom pretty much every day.

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.

Extra Help

There are Statistics TAs available from 7:00-9:00pm on Sunday–Thursday evenings in Burton-301. In addition, the Spinelli Center for Quantitative Learning (2nd Level of Neilson Library) 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.

Tentative Schedule

The following is a brief outline of the course. Please refer to the complete day-to-day schedule for more detailed information.

Week	Reading	Topic
1	Ch. 0, 1	Introduction & modeling. Understanding statistical models, and the four-step process (Choose, Fit, Assess, Use).
2	Ch. 1, 2	Simple Linear Regression. Least squares estimates, fitting models, checking conditions, making transformations.
3	Ch. 2, 3	Inference in simple linear regression, hypothesis tests and confience intervals. Regression and correlation.
4	Ch. 3	Multiple Linear Regression. Fitting models, checking conditions, comparing models.
5	Ch. 3	**Exam 1**

Week	Reading	Topic
6	Ch. 3	Second-order models. Interaction terms and polynomials. Multicolinearity.
7	Ch. 4	Model Selection. Using nested F-tests, added variable plots and stepwise regression.
3/12-3/20		**Spring Recess**
8	Ch. 4	Randomization & the Bootstrap as methods of inference.
9	Ch. 5	ANOVA. Fitting models, checking conditions, interpretation of results.
10	Ch. 5	Multiple Comparisons. Bonferroni corrections, Fisher's LSD, Tukey's HSD.
11	Ch. 9, 10	Logistic Regression and multiple logistic regression. Fitting and interpreting models. Checking conditions, formal inference.
12		Data wrangling and time series analysis. Autoregression and ARIMA models, estimation and forecasting with time series data. Errors and confidence intervals on forecasts.
13		**Exam 2** and **Project Presentations**
5/6		**All work due**