

Modern Regression, 36-401 and 36-607, Section B (Fall 2017)

This course is an introduction to applied data analysis. We will explore data sets, examine various models for the data, assess the validity of their assumptions, and determine which conclusions we can make (if any). Data analysis is a bit of an art; there may be several valid approaches. We will strongly emphasize the importance of critical thinking about the data and the question of interest. Our overall goal is to use a basic set of modeling tools to explore and analyze data and to present the results in a scientific report.

The course includes a review and discussion of exploratory methods, informal techniques for summarizing and viewing data. We then consider simple linear regression, a model that uses only one predictor. After briefly reviewing some linear algebra, we turn to multiple linear regression, a model that uses multiple variables to predict the response of interest. For all models, we will examine the underlying assumptions. More specifically, do the data support the assumptions? Do they contradict them? What are the consequences for inference? Finally, we will explore extra topics such as nonlinear regression or regression with time-dependent data.

A minimum grade of C in any one of the pre-requisites is required. A grade of C in this course is required to move on to 36-402 or any 36-46x course.

Prerequisites: At least a C grade in (36-226 or 36-625 or 73-407 or 36-310) and (21-240 or 21-241).

Corequisites: None.

Course Web Site: Course information will be posted on www.stat.cmu.edu/~larry/=stat401 and on Canvas.

Textbook: *Applied Linear Regression* by Kutner, Nachtsheim, and Neter. McGraw-Hill, Fourth Edition, 2004.

Lecture Notes: We will use lecture notes written by Professor Cosma Shalizi which provide a more modern view of the material. The notes are on the website. Download them and bring them to class.

Lectures: Tuesdays and Thursdays, 12:00 - 1:20, PH 100.

Instructor: Larry Wasserman larry@cmu.edu, Office: BH 132G

Instructor Office Hour: Tuesdays 1:30 - 2:30, BH 132G

TAs: TBD

TA Office Hours: TBD

Learning Objectives:

The goal is for students in this course to do the following:

- Demonstrate how/when to use exploratory data analysis tools (e.g., graphical displays).
- Develop model-building skills including evaluation of assumptions and interpretation of model-fitting results for linear regression models.
- Learn and apply the basic mathematical theory underlying linear regression models.
- Develop written and verbal communication skills for discussing conclusions and limitations of statistical evidence; present data analysis appropriately in a scientific report.
- Effectively use R, a widely-used statistical package, in data analysis.

Calendar:

A course calendar will be maintained on www.stat.cmu.edu/~larry/=stat401. Keep an eye on the calendar as it can change.

Homework:

There will be regular homework assignments, due **Fridays by 3:00**. Homeworks should be turned electronically using Canvas.

NO LATE HOMEWORKS WILL BE ACCEPTED FOR ANY REASON.

Assignments and solutions will be posted on the website.

You are allowed to discuss the assignments with other students in the course, but the work that you hand in **must be your own**. This means that each student must perform all of the work and write up the results independently. Failure to do so will constitute a violation of the University's policy on academic integrity.

Note that if only the correct answer is provided, but no relevant derivations, then zero points will be awarded. *Some advice:* With each exercise, look at your solution, and ask yourself: "Suppose I had been provided in advance with the correct answer to this exercise. Would it be clear to the grader that I understand how to reach that answer?" Keep in mind that "clarity" extends to your handwriting.

Please see the TA or instructor during office hours for help with homework problems.

In-Class Exams: There will be two in-class exams during the semester, on Thursday, October 5 and Tuesday, November 14.

There will not be any makeup exams.

Final Exam: There will not be a final exam.

Data Analysis Exams: There will be two take-home data analysis exams during the semester; each will be distributed at the conclusion of the in-class exams, and due one week later. Your analysis and results will be turned in as a structured report. We will cover the specifics in detail later in the course.

Final Grades: Final scores will be calculated based on a weighted averages of your scores for the homeworks (20%), the two tests (20% each), the two data analysis exams (20% each).

The lowest homework score will be dropped. Each homework will receive equal weight, regardless of the number of points.

Final grades will be assigned based on the same scale: 90% and up is an A, 80% to 89% is a B, etc.

Lectures: Lecture notes will be posted on the website. You are responsible for printing these and bringing them to class.

Computing: We will use the software package R to perform computations and simulations during the course. R is an all-around excellent piece of statistical software, and widely used in the sciences and research fields. You can download R (it is free) from

<http://cran.r-project.org/>

It is also available in many clusters. You may also use the nearly-identical (but not free) program S+, available

on all campus computers. You might also want to use R Studio

<http://www.rstudio.com>

R Studio is a user-friendly interface for working with R.

Academic Integrity: All students are expected to comply with the CMU policy on academic integrity. This policy is online at

http://www.studentaffairs.cmu.edu/dean/acad_int/

Cheating, copying, etc. will not be tolerated; please ask if you are unsure of whether or not your actions are complying with assignment/exam instructions.

Policies: Please note each of the following.

1. It is assumed that you check your andrew email at least once per day. Email will be used to communicate important details and announcements regarding the course.
2. All cellphones (or anything else that makes noise) should be silenced during class.
3. Sending email to your professor or teaching assistants should be treated as professional communication.
4. No student may record or tape any classroom activity without the consent of the instructor. If a student believes that he/she is disabled and needs to record or tape classroom activity, he/she should contact the Office of Equal Opportunity Services, Disability Resources to request appropriate accommodation.
5. If you have a disability and need special accommodation, please contact the Disability Resources office at 8-2013.