36-401/607 Modern Regression

Course Policies and Syllabus, Fall 2021

Instructor: Edward Kennedy (`edward@stat.cmu.edu`)
Office hours: Tuesday 1:10-2:00pm (Zoom link TBA)

Lectures: Tuesdays and Thursdays 11:50-1:10 pm, DH 2210

Teaching assistants:

Raghav Bansal (`raghavb@andrew.cmu.edu`)
James Carzon (`jcarzon@andrew.cmu.edu`)
Kenta Takatsu (`ktakatsu@andrew.cmu.edu`)
Konrad Urban (`kurban@andrew.cmu.edu`)

Office hours: TBA

Course website: All lecture notes are posted on http://www.cmu.edu/canvas. Homeworks are submitted on Gradescope at https://www.gradescope.com/courses/302308

Recommended Texts:

All of Statistics by L. Wasserman. Springer, 2004.
Statistical Models: Theory & Practice by D. Freedman. Cambridge Press, 2009.
Applied Linear Regression by Kutner, Nachtsheim, Neter. McGraw-Hill, 4th Ed, 2004.

Lecture Notes: Download from Canvas and – this is important! – read them before class.

Most Important Points

1. Read lecture notes before class!
2. Homeworks are due Fridays 3pm on Gradescope: https://www.gradescope.com/courses/302308
3. *No late HW assignments are accepted*, but your lowest score will be dropped.
4. Midterm exams are Thurs 9/30 and Thurs 11/4 during class. There are no make-ups! (The final is scheduled by the university during CMU's usual final exam period.)
5. Final grades: two exams (20% each), final exam (25%), and homework (35%).

Tentative Course Schedule

(Note: exam dates are fixed)

Lecture	Date	Topic
1	Aug 31	Introduction
2	Sept 2	Prob & stat review
3	Sept 7	Linear prediction & regression intro
4	Sept 9	Simple linear regression
5	Sept 14	Bias
6	Sept 16	Variance & interpretation
7	Sept 21	Predicted values
8	Sept 23	Confidence intervals
9	Sept 28	Confidence intervals
10	Sept 30	**EXAM 1**
11	Oct 5	Testing
12	Oct 7	Prediction
13	Oct 12	Matrix review
14	Oct 14	Multiple regression
15	Oct 19	Prediction & inference
16	Oct 21	Diagnostics & polynomial regression
17	Oct 26	Polynomial & categorical regression
18	Oct 28	Interactions
19	Nov 2	F tests
20	Nov 4	**EXAM 2**
21	Nov 9	Model selection
22	Nov 11	Cross-validation
23	Nov 16	Regression without assumptions
24	Nov 18	Nonparametric regression
25	Nov 23	High-dimensional regression
26	Nov 25	NO CLASS
27	Nov 30	Causal inference
28	Dec 2	Causal inference

(FINAL EXAM DATE TBA)

Course Overview

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 generalized linear models, additive models, nonparametric regression, and causal inference.

A minimum grade of C in one of the pre-requisites (36-226 or 36-700) *and* 21-240 or 21-241 is required. A grade of C in this course is required to move on to 36-402 or any 36-46x course.

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.

Homework

There will be approximately weekly homework assignments due **Fridays 3:00 pm EST**, unless otherwise stated. *No late homeworks will be accepted*!

Homeworks will be submitted on Gradescope: www.gradescope.com/courses/302308

Some details on Gradescope:

- You will need to create a single pdf for your submission (which can be a few screenshots / files concatenated). Make sure it is legible! For questions that require coding, it is fine to copy and paste the code in LaTex or take a screenshot of your code. We will let you know if you need to submit a separate R file in case the coding questions are more involved.
- You will upload your HW to Gradescope under the proper assignment. Gradescope will guide you to link each of your answers to the relevant questions.

Other details:

You are not allowed to copy or "consult" solutions from previous years, or another section, or online.

You are encouraged to back up your work by regularly uploading earlier versions of your homework; only your latest submitted version will be graded.

Assignments will be posted at least a week before the due date, and solutions a week or two after the due date.

You may have to read ahead to do some of the problems. You are allowed to discuss the assignments with other students in the course, but the work that you hand in, both written work and code, **must be your own** and written up independently. When you collaborate with another student, please write: "I worked with ..." on this assignment.

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?

Exams and Grading Policy

All exams can be cumulative (usually weighted towards more recent material) and you are not allowed to discuss the content of the exams with other students until the solutions have been posted. The dates of the exams will not be moved so schedule job interviews and other extra-curricular activities around them.

In-Class Exams: There will be two in-class exams during the semester, on:

Thursday, September 30 and **Thursday, November 4**.

Final Exam: TBA (assigned by registrar during semester)

There will not be any makeup exams.

Final Grades: Final scores will be calculated based on a weighted average of your scores for 2 midterms (20% each), homeworks (35%), and final exam (25%).

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

Grades will be computed using the usual scale: 90% and up guarantees an A, 80% to 90% guarantees a B, and so on. I may adjust this scale (only downwards) depending on the grade distribution at the end of the semester.

Lectures

You are expected to attend class and actively take notes. I will prepare lecture notes and slides for you to bring to class; these will be posted by the end of the evening before lecture. *You are responsible for printing out and reading these materials before class.*

Academic Integrity and Plagiarism

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.

Other Policies

1. It is assumed that you check your andrew email daily. Email may be used to communicate important details and announcements regarding the course.
2. Students are expected to be attentive during lecture; *laptop use should pertain directly to the class*, and no cellphones allowed.
3. Sending email to your professor or TAs should be treated as professional communication. Email sent within 24 hours of a due date may not receive a response.
4. No student may record any classroom activity without the consent of the instructor.
5. If you have a disability and are registered with the Office of Disability Resources, please use their online system to notify me of your accommodations and contact me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.

Study & Other Tips

1. **Actively read lecture notes before class.** Highlight and make notes, fill in gaps with extra explanations, re-do examples step-by-step, write down questions about things you don't understand.

   This will help you immensely, and will make lecture much more useful. It will become a time for you to gain a deeper understanding, ask clarifying questions, etc., rather than trying to digest completely new material for the first time.
2. *Take advantage of office hours and use them productively.* Office hours can be a huge benefit to you – you have a direct line to the professor and TAs to improve your understanding and/or get help.
3. *Do homework problems.* Actively doing problems is the *only* way to learn the material. Try to do the problems yourself before discussing them with other people.
4. *Take care of yourself.* Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

   If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.