MATH 4780 (MSSC 5780) Section 101: Regression Analysis

Instructor: Dr. Cheng-Han Yu

Fall 2023

E-mail: cheng-han.yu@marquette.edu
Office Hours: TuTh 3:20 - 4:50 PM or by appointment
Office: Cudahy Hall 353
Teaching Assistant (TA): `No TA :(`

Web: math4780-f23.github.io/website
Class Hours: TuTh 5:00 - 6:15 PM
Class Room: Cudahy Hall 120

Course Objectives

Starting from simple linear regression, this course provides a brief, yet complete foundation for the understanding of basic regression theory and applications that is a core of supervised learning in statistical machine learning. Topics include multiple linear regression, diagnostic analysis, multicollinearity, nonparametric regression, variable selection, generalized linear models (GLMs) and other selected topics.

Prerequisites

- MATH 2780 (Intro to Regression and Classification), MATH 4720 (Intro to Statistics) or equivalent.
- Programming experience is helpful because the course involves doing regression analysis using R programming language.
- Some linear algebra exposure at the level of MATH 3100 (Linear Algebra and Matrix Theory) also helps as some matrix operations are used for modeling and computation.
- Having taken MATH 4700 (Probability) and MATH 4710 (Statistical Inference) would help too.
- Talk to me if you are not sure whether or not this is the right course for you.

Textbook (Optional)

- (**IS**) *Introduction to Statistics* by Cheng-Han Yu. (Good resource for brushing up your basic probability, statistics and simple linear regression knowledge.)
- (**LRA**) *Introduction to Linear Regression Analysis, 6th edition*, by D. C. Montgomery, E. A. Peck and G. G. Vining. Publisher: Wiley. (Graduate-level book that requires knowledge of matrix algebra.)
- (**CAR**) *An R Companion to Applied Regression*, by John Fox and Sanford Weisberg. Publisher: SAGE. (Doing regression much easier by the `car` and `effects` R packages.)
- (**RD**) *Regression Daignostics*, by John Fox. Publisher: SAGE. (Great concise book for learning regression diagnostics.)
- (**CMR**) *Classical and Modern Regression with Applications*, by Raymond Myers. Publisher: Duxbury Press. (The textbook when I was a master student.).

Programming Languages

The default language of this course is R. Due to time constraint, I will NOT teach you R programming from A to Z. If you never use R and would like to grasp its syntax quickly, study my MATH 3570 R programming slides. I am more than happy to answer any questions about R if you have any. There are a lot of online resources for learning R including the sites below.

- *Quick-R*
- *Hands-On Programming with R*

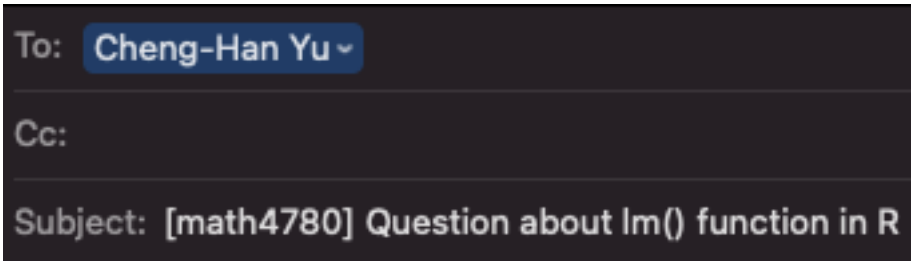

Figure 1: Email Subject Line Example

There are other popular languages for data analysis and machine learning such as Python, Julia, MATLAB, SAS, etc. You may use any other language to complete your homework assignments and/or exams. However, I would not debug your code or provide comments on any technical issues of those languages.

Office Hours

- My in-person office hours are TuTh 3:20 - 4:50 PM in Cudahy Hall room 353.
- We can schedule an online meeting via Microsoft Teams if you need/like.

E-mail Policy

- I will attempt to reply your email quickly, at least **within 24 hours**.
- **Expect a reply on Monday if you send a question during weekends**. If you do not receive a response from me within two days, re-send your question/comment in case there was a "mix-up" with email communication (Hope this won't happen!).
- Please start your subject line with [**math4780**] or [**mssc5780**] followed by a clear description of your question. See an example in Figure 1.
- Email etiquette is important. Please read this article to learn more about email etiquette.
- I am more than happy to answer your questions about this course or statistics in general. However, due to time constraint, I may choose **NOT** to respond to students' e-mail if
  1. The student could answer his/her own inquiry by reading the syllabus or information on D2L.
  2. The student is asking for an extra credit opportunity. The answer is "no".
  3. The student is requesting an extension on homework. The answer is "no".
  4. The student is asking for a grade to be raised for no legitimate reason. The answer is "no".
  5. The student is sending an email with no etiquette.

Grading Policy

- Students in this class range from juniors to second-year PhD students, so it makes sense to evaluate student performance using a different scale.
- For students in **MATH 4780**, the final grade is earned out of **1000 total points** distributed as follows:
  - **Homework 1 to 6: 480 pts (80 pts each)**
  - **In-class Exam: 160 pts**
  - **Final project: 300 pts**
  - **Class Participation: 60 pts**
- For students in **MSSC 5780**, the final grade is earned out of **1200 total points** distributed as follows:
  - **Homework 1 to 6: 480 pts (80 pts each)**
  - **In-class Exam: 160 pts**
  - **Final project: 300 pts**
  - **Class Participation: 60 pts**
  - **MSSC work: 200 pts**
- You will **NOT** be allowed any extra credit projects/homework/exam to compensate for a poor average. Everyone must be given the same opportunity to do well in this class. Individual exam will **NOT** be curved.
- The final grade is based on your percentage of points earned out of 1000 or 1200 points and the grade-percentage conversion Table. $[x, y)$ means greater than or equal to x and less than y. For example, 94.1 is in $[93, 100]$ and the grade is A and 92.8 is in $[90, 94)$ and the grade is A-.

Table 1: Grade-Percentage Conversion

Grade	Percentage
A	[94, 100]
A-	[90, 94)
B+	[87, 90)
B	[83, 87)
B-	[80, 83)
C+	[77, 80)
C	[73, 77)
C-	[70, 73)
D+	[65, 70)
D	[60, 65)
F	[0, 60)

- This is not a course that gives most of students grade A. If you want to obtain a good grade, study hard. No pain, no gain.

Homework

- Homework will be assigned through the course website in weekly schedule.
- To submit your homework, please go to **D2L > Assessments > Dropbox** and upload your homework in **PDF** format.
  - Combine images into a PDF file
  - Microsoft Word to PDF in 10 Seconds
  - Convert images to PDF in Macbook/iMac
  - Combine two or more images to get a single PDF file
- There will be 6 homework sets to be graded, Homework 1 to 6.
- Some questions are required for MSSC students.

- You will get a better understanding of the material if you discuss it with others. However, you must submit YOUR OWN work.
- **NO LATE HOMEWORK WILL BE ACCEPTED NOR WILL YOU BE ALLOWED TO MAKE UP MISSED HOMEWORK!**
- **Handwriting is not allowed for data analysis part**.

In-Class Exam

- The in-class exam covers materials in Week 1 to Week 7 on simple linear regression and multiple linear regression.
- Some questions are required for MSSC students.
- **NO make-up midterm exam** unless you miss exams due to COVID-19 symptoms, exposure, diagnosis, quarantine, and/or isolation, or you have an excused absence as defined in Attendance in Academic Regulations.

Class Participation

- You will be presenting exercise problems.
- More details will be released later.

Final Project

- A detailed final project guideline will be released later.

Academic Integrity

- This course expects all students to follow University and College statements on academic integrity.
- **Honor Pledge and Honor Code**: *I recognize the importance of personal integrity in all aspects of life and work. I commit myself to truthfulness, honor, and responsibility, by which I earn the respect of others. I support the development of good character, and commit myself to uphold the highest standards of academic integrity as an important aspect of personal integrity. My commitment obliges me to conduct myself according to the Marquette University Honor Code.*
- You know what I am talking about. Yes, **DO NOT CHEAT**.

Accommodation

If you need to request accommodations, or modify existing accommodations that address disability-related needs, please contact Disability Service.

Attendance and COVID-19

- It is your responsibility as a Marquette University student to protect the health and safety of our community in this course. The University has mandated vaccination for all students for the academic year 2021-2022. COVID Cheq will still be used, but those fully vaccinated that provide documentation will no longer receive the daily symptom screening survey. Instead, they will receive a daily email providing a green check. You may be required to show your automated green check to show clearance for entry into campus buildings. If you are experiencing symptoms related to COVID-19, you must follow the link in the automated email to report symptoms.
- Visit What to do if you are exposed to COVID-19 or test positive website for university guidelines on the best course of action.
- Visit guidance on Fall 2023 Class attendance, withdrawal, and grading
  - Students are responsible for contacting instructors prior to the missed class session to indicate absence and the need to make up classwork/assignments.

- Students requesting make up classwork/assignments are required to provide the COVID Cheq "stop sign" to confirm inability to attend class.

Tentative Course Schedule

Week 1, 8/28 - 9/3: Syllabus, Overview of Regression, Probability and Statistics Review

Week 2, 9/4 - 9/10: Simple Linear Regression

- **Labor day 9/4**
- **Drop deadline 9/5 11:59 PM**
- **Homework 1 due Friday, 9/8 11:59 PM**

Week 3, 9/11 - 9/17: Simple Linear Regression

Week 4, 9/18 - 9/24: Multiple Linear Regression

- **Homework 2 due Friday, 9/22 11:59 PM**

Week 5, 9/25 - 10/1: Multiple Linear Regression

Week 6, 10/2 - 10/8: Multiple Linear Regression in Matrix Form (MSSC)

- **Homework 3 due Friday, 10/6 11:59 PM**

Week 7, 10/9 - 10/15: Regression Diagnostics: Residuals and Unusual Data

Week 8, 10/16 - 10/22: Regression Diagnostics: Symmetry, Normality and Constant Variance

- **In-class Exam on 10/17**

Week 9, 10/23 - 10/29: Regression Diagnostics: Linearity and Lack of Fit

- **Midterm grade submission 10/24 by noon**
- **Homework 4 due Friday, 10/27 11:59 PM**

Week 10, 10/30 - 11/5: Bootstrapping and Simulation-based Inference

Week 11, 11/6 - 11/12: Polynomial and Nonparametric Regression

Week 12, 11/13 - 11/19: Categorical Variables

- **Homework 5 due Friday, 11/17 11:59 PM**
- **Withdrawal deadline 11/17**

Week 13, 11/20 - 11/26: Collinearity

- **Thanksgiving holidays: 11/22 - 11/26**

Week 14, 11/27 - 12/3: Variable Selection

Week 15, 12/4 - 12/10: Logistic Regression

- **Homework 6 due Friday, 12/8 11:59 PM**

Week 16, 12/11 - 12/17:

- **Project Presentation, 12/12, 5:45 - 7:45 PM**
- **Final Grade submission 12/19 by noon**

*** I reserve the right to make changes to the syllabus.**