DATA 5600: Introduction to Regression and Machine Learning for Analytics

Fall Semester, 2021

Course Information

- Course Dates: August 30 - December 17, 2021
- Course Time: MW 1:30 - 2:45 PM
- Course Room: Huntsman Hall 360
- Slack Channel
- Course Canvas

Instructor Information

- Tyler J. Brough
- Office Hours: By Appointment
- Office: BUS 512 and 719
- Email: tyler.brough@usu.edu

Syllabus

Course Description

This course is the first in a sequence of courses in data analysis for business and economic decision making. The focus of this course is regression modeling. The approach of the course is applied rather than theoretical. That is, we will focus not just, or even primarily, on the mathematics of regression modeling but on how to apply and understand these data analysis tools. The goal of the course is for you to develop deep intuition for how to build and apply regression models.

The main purpose of the course is to foster three modes of thought to enable you to become thoughtful scholars of data analytics:

1. **Computational Logic:** Computational thinking helps break a problem down and find a practical path to successful implementation. Important techniques from computational mathematics and computational statistics will be introduced as they pertain to data analysis.
2. **Statistical Logic:** The main focus of the course will be learning to think statistically about problems in business analytics. All meaningful problems in data analysis are necessarily embedded in conditions of uncertainty. There exists a core statistical logic that is distinct from the mathematical. Learning to develop this mode of thinking is an essential step in the life of any successful business data analyst. We will find that statistical reasoning is essential for proper business reasoning.
3. **Arbitrage Logic:** Arbitrage is the central underlying concept all of economics and finance. Developing skills in arbitrage reasoning is an

essential element in business data analytics.

These three ways of thinking are not independent. Quite to the contrary, we will find that they are strongly mutually reinforcing.

Textbook and Other Resources

There is only one *required* textbook for the course:

- Regression and Other Stories by Gelman, Hill & Vehtari.

I will also use some other books for lecture material. Some of these are the following:

- Computational Statistics Handbook with MATLAB by Martinez & Martinez.
- An Introduction to Modern Bayesian Econometrics by Lancaster.
- A Guide to Econometrics by Kennedy.

There will be other readings from academic and trade articles assigned throughout the semester. These will be made available on Canvas. ***This course is reading intensive so please be prepared to read thoroughly and come prepared to discuss those readings.***

Method of Teaching and Learning

This course will be taught as a graduate seminar style course. That means that your preparation and participation is crucial. We will get out of the course what we individually and collectively put in to it.

I will use the Socratic method during lectures as much as feasible. A good amount of class time will also be dedicated to instructor guided computational exploration. The idea behind this exploration is to develop a methodology for ***computational thinking***. Mathematical theory will serve as an important background to this enterprise but will not itself be the focus.

Your preparation and participation is absolutely essential to the success of this course!

Assessment

The grade that you will earn will be determined by your ranking in the class based on the weighted total points accumulated. There is not a predetermined percentage of the class that will get an A or that will fail. If you all do excellent work, you will all earn exceptional grades. The weights given to each part of the class are as follows:

- Class Preparation and Participation (15%) - Your preparation is crucial! No student can earn an A without meeting these requirements!

- Computational Exercises (25%) - There will be weekly homework assignments that will consist of numerical and computational problems.
- Midterm Exam (30%) - The midterm will be a take-home exam. You may take the midterm as many times as you like. It will cover foundational concepts from prerequisites as well as the early chapters of the textbook.
- Final Project (30%) - This is your opportunity to build and implement a real-world regression model. There will be two components of your grade:
    1. A written research report.
    2. An oral presentation of your report.

Communications

I will use Canvas somewhat minimally to outline our progress, share documents and other materials, and for uploading content.

For all other communication purposes we will use a course Slack channel. You will receive an invitation to the Slack the first week of class. You should check for communications and announcements often.

Topics

We will cover the following three main modules:

1. Fundamentals: chapters 1 - 5
2. Linear regression: 6 - 11
3. Advanced topics:
    - Logistic regression and classification: 13
    - Causal inference: 18
    - Machine learning: 22

NB: I reserve the right to dynamically alter this list as the course progresses. I will announce any such changes in class and on the course Slack channel.

Import dates:

- **Labor Day** - Sep 6
- **Fall Break** - Oct 15
- **Thanksgiving Holiday** - Nov 24-26
- **No-Test Week** - Dec 6-10
- **Last Day of Classes** - Dec 10
- **Final Examinations** - Dec 13-17