# DATA 5600: Introduction to Regression and Machine Learning for Analytics

**Spring Semester, 2022**

## Course Information

- Course Dates: Jan. 10 - May 4, 2022
- Course Time: MW 3:00 - 4:15 PM
- Course Room: Huntsman Hall 326
- Slack Channel
- Course Canvas

## Instructor Information

- Tyler J. Brough
- Office Hours: By Appointment
- Office: BUS 512
- 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. **Economic Logic:** Arbitrage is the central underlying concept all of economics and finance. Developing skills in economic 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:

- Introductory Econometrics: A Modern Approach 5th Edition

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

- An Introduction to Modern Bayesian Econometrics by Lancaster.

- Computational Statistics Handbook with MATLAB by Martinez & Martinez.

- Introducing Monte Carlo Methods with R by Robert & Casella

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 this course what we individually and collectively put into it.

We will follow the Socratic Method during lectures. A good amount of class time will also be dedicated to 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 the weighted total points accumulated. The weights given to each part of the class are as follows:

- *Class Preparation & Participation (100%)* - In a Socratic style seminar student preparation and participation is essential!

- *Computational exercises (10%):* There will be four main computational exercises to develop software architecture skills that are essential for the practicing data analyst.

- Chicken nuggest/coin problem
- Guess my number problem
- Monte Carlo simulation of pi problem
- Cherry tootsie rolls bandit problem

- *Preliminary exam (20%):* A take-home preliminary exam covering the essentials of probability and mathematical statistics that are essential for regression analysis.

- *Student Presentations (10%)* - Throughout the course a set of specialized readings will be distributed. Each student will have an opportunity to prepare a presentation to the class on one or more of the assigned readings.

- *Annotated Bibliography & Wiki-style Articles (10%)*

  - Throughout the course, a set of specialized readings will be distributed. Each student will prepare an annotated bibliography of each reading.
  - Students will prepare a Wikipedia-style summary article on one specialized topic from each of the four areas of concentration. The following are examples (not exhaustive) of topics that might be chosen:
    1. Probability & Math Stats: Bayes rule, null hypothesis testing, maximum likelihood, etc.
    2. Computational statistics: Bootstrap methods, sampling-importance resampling (SIR), variance reduction technique, etc.
    3. Regression: least squares, Bayesian regression, etc.
    4. Advanced topics: supervised learning, classification, loss functions, non-stationary time series, etc.

- *Projects (30%):*

  - Simple linear regression: futures hedging data (15%)
  - Multiple linear regression: stock returns data (15%)

- *Take home final exam (20%):* there will be a applications-based take-home final exam.

- *Extra Credit:* There will several opportunities for extra credit work.

### 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 main modules:

1. Review of Probability and Mathematical Statistics: Appendices A - C
2. Some basic tools from Computational Statistics: handouts
3. Linear regression: Chapters 2 - 8
4. Advanced topics:

- Regression with time series variables: Chapters 10 - 12
- Limited dependent variable models: Chapter 17

***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:

- **First day of classes** - Jan 10
- **Martin Luther King, Jr. Day** - Jan 17
- **Presidents' Day** - Feb 21
- **Spring Break** - Mar 7 - 11
- **Last day of classes** - Apr 26
- **Final exams** - Apr 28 - May 4