

Statistics

Statistical Modeling and Learning (STM 2102, 4.5 credits)

THE UNIVERSITY OF AUSTIN

Instructor: David Puelz

Course Webpage: <https://github.com/dpuelz/StatisticalModeling>

Email: dpuelz@uaustin.org

Meeting Schedule: M/W/F from 10:00a-11:15a

Office Hours: Please consult the webpage

Course Description

This course trains students to wield data as both a scientific and predictive tool with an applied focus on business and economics. It develops mastery in two complementary approaches: supervised machine learning, which builds powerful models for prediction and decision making, and causal inference, which uncovers credible cause and effect relationships. Students will learn to design empirical studies, implement algorithms, and critically evaluate when data reveals insight.

Course Objectives

- Understand the fundamentals of statistical modeling and supervised learning
- Fit and interpret regression models for prediction and inference
- Apply statistical methods to real-world data using appropriate software
- Evaluate model performance and understand the bias-variance tradeoff

Required Readings

The first book is available for free online. The second book is supplemental reading for the causal inference parts of class.

- *Introduction to Statistical Learning* (ISL) – Gareth James, et al.
- *Mastering 'Metrics* (MM) – Joshua D. Angrist and Jörn-Steffen Pischke

Course Cadence

There will be 5 quizzes and 5 homework assignments. The quizzes will be on the Fridays of weeks 2, 4, 6, 8, and 10. The homeworks will be due on Fridays at 10:00a (start of class) on the same days as the quizzes. The quiz content will be related to the homework, and we will mark up the quizzes in class directly after finishing the quiz. We will have a final exam during the scheduled exam time (on week 11 of the course).

Evaluation

The class grade is comprised of the following:

- Homework (20%)
- Exam (30%)
- Quizzes (50%)

Topics

This is a course on supervised learning. We will cover the following topics: (i) introduction and bias-variance tradeoff, (ii) inference for regression, (iii) multiple regression, (iv) categorical predictors

and interactions, (*v*) assumptions and diagnostics, (*vi*) nonlinear regression, (*vii*) time series regression, (*viii*) logistic regression, (*ix*) model selection and penalized regression, (*x*) trees, ensembles, and neural networks, and (*xi*) causal inference.

Below is a timeline for the semester. It is subject to change as we make our way through these topics. Please also pay attention to the course website. I will include supplemental readings beyond those in our textbooks.

| Date | Topics | Reading |
|------------------|---|----------------------------|
| Week 1 (Jan 5) | Intro and bias-variance tradeoff | ISL: Ch 1, 2 |
| Week 2 (Jan 12) | Inference for regression + Multiple regression | ISL: Ch 3.1–3.3, MM: Ch 2 |
| Week 3 (Jan 19) | Categorical predictors and interactions | ISL: Ch 3.6–3.7 |
| Week 4 (Jan 26) | Assumptions, diagnostics + Nonlinear regression | ISL: Ch 3.3.3, 7.1–7.4 |
| Week 5 (Feb 2) | Time series regression | ISL: Ch 3 (supplemental) |
| Week 6 (Feb 9) | Logistic regression | ISL: Ch 4.1–4.3 |
| Week 7 (Feb 16) | Model selection and penalized regression | ISL: Ch 6.1–6.5 |
| Week 8 (Feb 23) | Trees, ensembles, and neural networks | ISL: Ch 8.1–8.3, 11.1–11.3 |
| Week 9 (Mar 2) | Causal inference | MM: Ch 1, ISL: Ch 3.2 |
| Week 10 (Mar 9) | Causal inference | MM: Ch 3–5 |
| Week 11 (Mar 16) | <i>final exam week</i> | |

Accessibility Statement

Please review the University Accessibility Statement in the student catalog. Students having special needs should email Accomodations@uaustin.org. Disability Support Services: The University will make reasonable accommodations for students with disabilities in compliance with Section 504 of the Rehabilitation Act and the Americans with Disabilities Act. The purpose of accommodations is to provide equal access to educational opportunities for eligible students with academic and/or physical disabilities.

Attendance Policy

Attendance is mandatory. If you don't show up, you will fail the course. Each student may miss up to 10% of classes for any reason—no excuse required and without penalty—corresponding to 1, 2, or 3 classes in a 1.5, 3.0, or 4.5 credit course, respectively. Any additional absences will result in a final grade penalty of 12% (1.5 credits), 6% (3.0 credits), or 4% (4.5 credits), regardless of the reason. Missing more than 25% of classes—including the allowable absences—without a medical excuse will result in failing the course. Being more than 20 minutes late to class counts as an unapproved absence.

Showing Work

On all course deliverables, there is an expectation of showing how you got to an answer. If you show a partially correct sequence of steps to a wrong answer, you may be awarded partial credit (at the discretion of the instructor). If you only write down the correct answer, you will not receive any credit. I will not award credit to random guessing and shoddy, vague, or unclear supporting work. This is part of the pedagogical design—demonstrate to me that you know what you're doing!

Late Policy

Sometimes we have bad days, bad weeks, and bad terms. In an effort to accommodate any unexpected, unfortunate personal crisis, I have built a grace policy into the course: that is, a one-time, three-day grace period for one homework assignment. You do not have to utilize this policy, but if you find yourself struggling with unexpected personal events, I encourage you to e-mail me as soon as possible to notify us that you are using the grace policy. **All other late assignments will be penalized 20 percentage points per day or partial day that they are late.**

STEM Center Learning Objectives

This is an intellectual foundations course that contributes to the following STEM Center learning objectives:

- Speak, read, and write in mathematical language.
- Formulate and solve quantitative problems using appropriate technical language, models, and techniques.
- Collect and analyze data using appropriate technology tools.
- Communicate technical solutions clearly through writing, speaking, and visualization.

Final Course Notes: AI Help, Academic Misconduct, etc.

I encourage the use of ChatGPT and other AI tools for coding. They are tremendously useful calculators that will continue to grow in importance. I have zero tolerance for submitting work that is not your own, regardless of whether it is plagiarized or copied from a fellow student. If this occurs on either the homework, midterm, or project, I will issue an automatic zero for the deliverable. Finally, if you are reading this sentence, please send me a note stating, “I have read the syllabus!”, and I’ll give you 2 bonus points on your raw final exam grade!