STAT 4130J Applied Regression Analysis Lecture 1: Course overview

We provide online learning & lecture recording!

Agenda

- Meet your teaching team

- Course Introduction

- Course Logistics

- Meet your classmates

Ailin Zhang
- Background: PhD in Geophysics from UCLA (2019). Formerly data scientist @ ExxonMobil
- Research Interests: Data-driven solutions to geoscience problems: earthquake rupture, oil and gas exploration, seismic signal processing.
- Office Hour: Tuesday 2-4 pm @ Longbin building 437B or by appointment.

TA: Haohong Shang

- STATS 413 Applied Linear Regression
- STATS 415 Data Mining
- STATS 315 Deep Learning
- EECS 445 Intro to Machine Learning
- EECS 484 Database Management System
- Research: ML related

- Email: shanghaohong@126.com
- Phone: 18916905957
- Wechat: shhh2001(Preferred)
- RC/OH: TBD

ECE + DS Junior

- **Regression analysis** is used to estimate the relationships between variables and includes many different techniques.

- We will determine if the one of more predictor/independent variables have a significant effect on the response/ dependent variable.

- **Linear regression** serves as the basis for other types of regression.

- It is used extensively in many fields, such as medical science, social science, environmental science, economy and so on.

- This course covers tow major topics:

  - Linear regression:

    - Simple and multiple linear regression, analysis of variance and covariance, and nonparametric analysis. Model building techniques, evaluating model fit, and dealing with violations of model assumptions.

  - Modern regression methods:

    - Logistic regression, Poisson regression, Log-linear models etc.,

Application Scenarios

- You want to characterize the relationship between the dependent and independent variables by determining the extent, direct, and strength of the association

- You would like to build a quantitative equation to predict the dependent variable Y as a function of the independent variables X1, X2, ….

- Apart from X1, X2,….and Y, you might also find other variables relevant to Y, and you would like to study the effects of other variables C1, C2, ….

- You want to determine which of the several independent variables are important and which are not in predicting Y

Regression Analysis is widely used

Regression analysis is a powerful statistical tool used to model the relationship between a dependent variable and one or more independent variables. Here are some examples of regression analysis applications:

1. Predicting stock prices: Regression analysis can be used to predict stock prices based on historical data and other relevant factors such as company earnings, market trends, and economic indicators.

2. Forecasting sales: Regression analysis can help businesses forecast sales based on factors such as advertising spend, seasonality, and consumer demographics.

3. Analyzing customer behavior: Regression analysis can be used to model customer behavior and predict how customers are likely to respond to changes in product features or pricing.

4. Predicting healthcare outcomes: Regression analysis can be used to predict healthcare outcomes based on patient characteristics, medical history, and treatment options.

5. Analyzing economic trends: Regression analysis can be used to analyze economic trends and forecast economic growth based on factors such as inflation, interest rates, and government policies.

6. Predicting weather patterns: Regression analysis can be used to model and predict weather patterns based on historical weather data and other environmental factors.

7. Analyzing sports performance: Regression analysis can be used to analyze sports performance and predict how changes in training, diet, or other factors are likely to affect athlete performance.
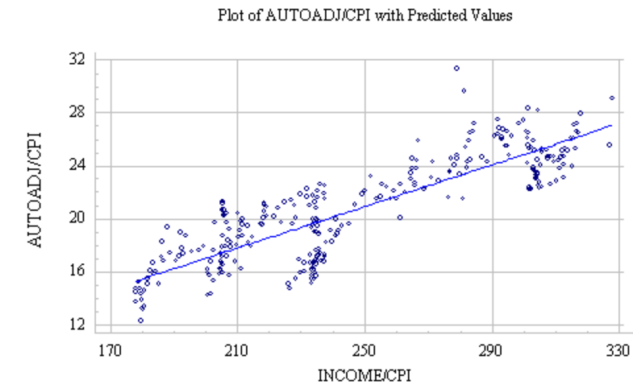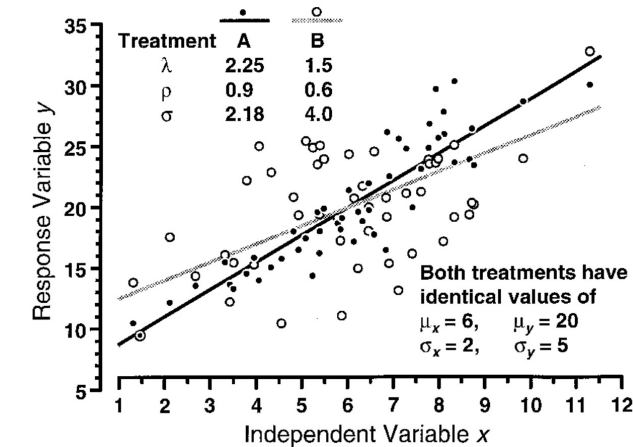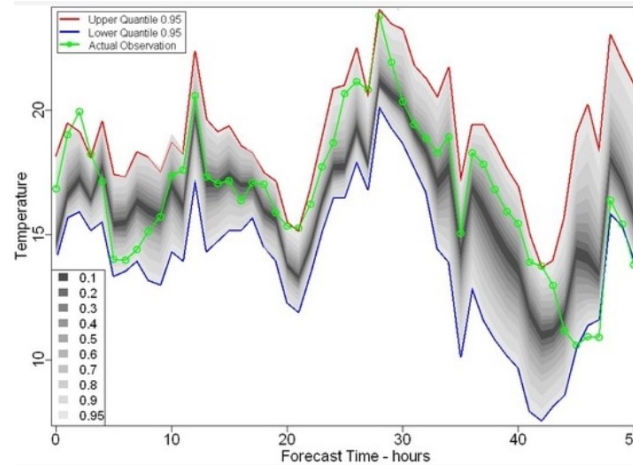
Topics to be discussed

- Simple linear regression

- Linear Algebra

- Multiple linear regression

- Dummy variables

- analysis of variance and covariance

- Model building techniques, evaluating model fit, and dealing with violations of model assumptions.

- Model Selection

- Collinearity

- Other Regression models
  - Generalized linear models
  - Nonlinear regression
  - Time-series regression
  - Nonparametric regression

- Data analysis with R/ Python/SPSS/SAS

Commonly Asked Questions

- What is the difference between STAT 4130 and STAT 4060?
  - In STAT 4060, we are more focused on the computational aspects of different algorithms, i.e. code up linear regression from scratch, how to numerically address the inverse of a matrix etc.,
  - In STAT 4130, we using computational languages as a tool to help us explain the relationship between variables: you are welcome to use all built-in libraries! But you need to very precise about the interpretation of models.

- I am interested in machine learning, can I learn machine learning in STAT 4130?
  - Yes, but probably not the best fit. We will provide high-level introduction to some modern regression methods (machine learning models) in the second half of the lecture, but that is not our primary focus.
  - We will cover 60% linear regression and 40% other regression models in this course.

- Prerequisites
  - Probability and statistics
  - Linear algebra
  - Coding

Course Logistics (Things are evolving!)

- References: We don't have a fixed textbook, but here are some references
  - Fox, John. *Applied regression analysis and generalized linear models*. Sage Publications, 2015.
  - Weisberg, Sanford. *Applied linear regression*. Vol. 528. John Wiley & Sons, 2005.
  - Hadi, Ali S., and Samprit Chatterjee. *Regression analysis by example*. John Wiley & Sons, 2015.
  - James, Gareth, et al. *An introduction to statistical learning*. Vol. 112. New York: springer, 2013. (Recommended for the second half)
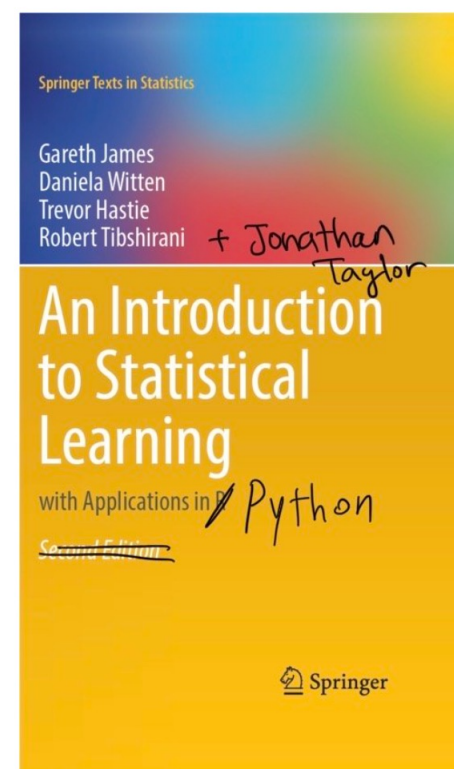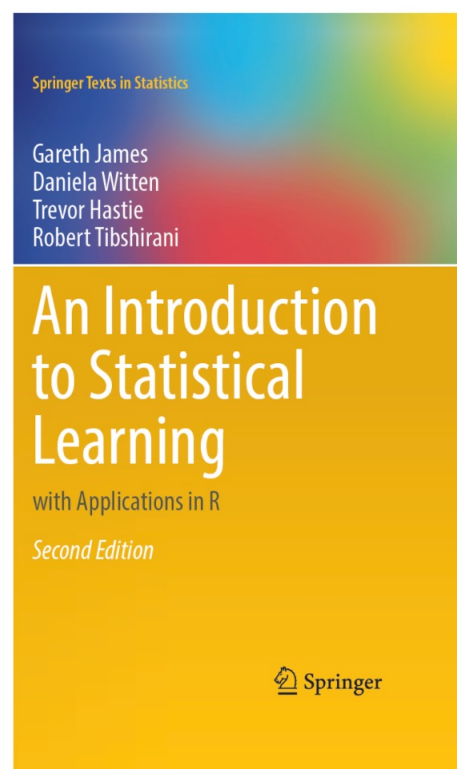
Grading

- I reserve the right to curve the scale if there are less than 30% of students with grades ≥ A. Median grade will be no worse than B − B+

  - **25%** Homework

  - **25%** Quiz

  - **30%** Midterm

    - Around week 8

    - Closed book, closed note

  - **20%** Final Project

    - Due by the end of week 11

  - **2%*** Extra Credit

    - 1 pt for course evaluation

    - 1 pt for Piazza participation
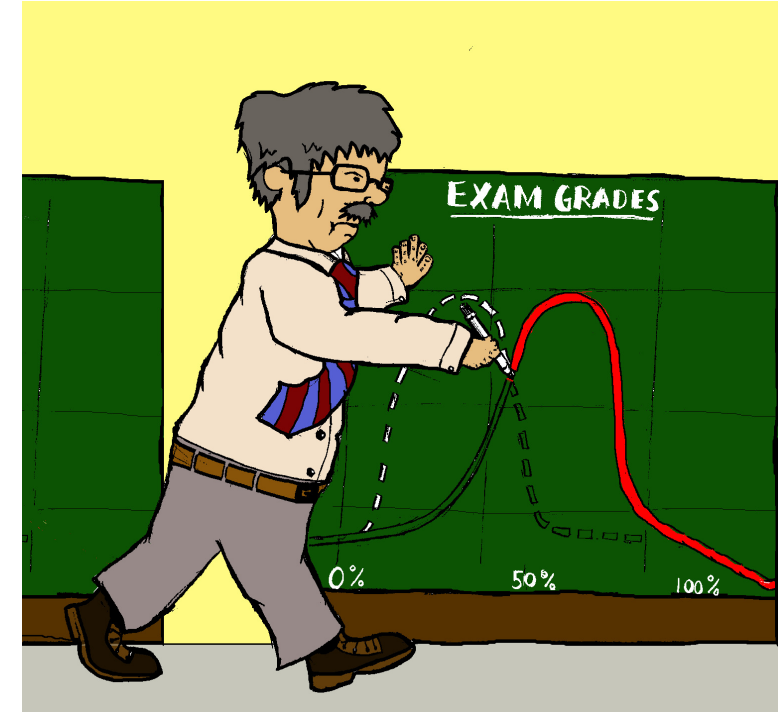
Using chatgpt in your homework and project is fine, but you need state it somewhere in your work.

Quiz and Project

- We will have bi-weekly quizzes scheduled on Tuesdays for odd weeks.

- The quiz will be distributed on Canvas (You need to bring your laptop).

- Open-book, open-note, no internet access 15-20 min.

- If you are working on the quiz in the classroom, you can turn in your work on paper. (Recommended)

- If you are taking the quiz online, please turn in your work in a pdf. (Note: we don't give you extra time for scanning and uploading, therefore please schedule the time accordingly)

- Unable to submit the quiz on time will lead to a point deduction.

- No make-up quiz will be given unless you have a reasonable excuse and notify us beforehand.

Online learning

- We provide online learning & lecture recording! (for lectures)
- For exams, if you want to apply for online midterm and online final presentation, you need to provide an excuse note. Accepted excuses are:
    - You are still physically abroad (only if you are a GDP student)
    - You are phyically discomfort
    - Note: Internship is not an excuse for taking the exam online

We are here to help!

- We really want you to succeed in this class.
- Feel free to reach out with any questions or concerns you have.

Meet you!

- Name, Year, Major

- Why STAT 4130?

- What is your expectation for taking this course?

- Your background in R and Linear Algebra

- Any other questions?