SDS358: Applied Regression Analysis

Day 1: Introduction and Overview

Dr. Michael J. Mahometa

Agenda for Today:

- Introductions
- What regression is good for
- About the course (general structure)
- The Lab
- The Syllabus for final bits

Hadley Says:

"You can't go from knowing nothing to becoming an expert without going through a period of great frustration and great suckiness."
–Hadley Wickham

Introductions

Me:

- Statistical Consultant
- Regression lover
- MOOC creator
- ex-SDS302 teacher

Introductions

You:

- What's your name?
- What department are you from?
- Why stats have you had so far?
- What's your software experience?
- Why are you here?

This course:

SDS358

Applied Regression Analysis

This course is intended to:

- Show you regression analysis
- Give you skills to perform an analysis on data
- Teach you to interpret the analysis output
- Use regression analysis

Why Regression?

Other than "Because it's awesome"

- We can answer amazing questions.
  - Does a functional relationship exist?
  - Identify the type of relationship (linear hopefully)
  - Make predictions
  - Make inference (Is something important in the prediction?)
  - Does a relationship change given another predictor?
  - What's the liklihood of something happening?

Course structure

- Weekly concepts
  - Reading quiz at the start
  - Lecture, Lecture, Lab
  - Lab assessment at the end
- Three "unit" tests
- Project
- Readings and Text

About the Lab

I wanted to get you skilled on a tool that you can take and actually **use**, so I chose R (and its close companion, RStudio).

Pros:

- It's free
- You can use it on PC, Mac, or Linux
- It's recommended by the ASA
- It provides for "reproducibility"
- It's very flexible

About the Lab

I wanted to get you skilled on a tool that you can take and actually **use**, so I chose R (and its close companion, RStudio).

Cons:

- It's very flexible
- It's syntax based (you write code)
- The learning curve can be steep

So to help, we'll work the Lab together. And I'll give you the syntax.

About the Lab

Part II

The Labs will be in this room.

- You **will need** your own laptop
- And, you **will need** R, Rstudio, and my SDSRegressionR package from GitHub
- Your lab assessments will be completed independently after the Lab.

About the Lab

Part II

The Labs will be in this room.

- You **will need** your own laptop
- And, you **will need** R, Rstudio, and my SDSRegressionR package from GitHub
- Your lab assessments will be completed independently after the Lab.

I recommend:

- Make a specific directory structure
- Back up your work

About the Lab

Part III

- Labs will answer a primary research question using *REAL* data.
  - Reflect on the Question
  - Analyze the Data
  - Draw Conclusions
- Labs will be screen-cast (for you to return to a section and review our logic).
- Lab Assessment will follow the same structure (but you'll be doing it independently).
- Example

Why do we have this Lab?

- It's an *applied* class.
- It's helpful to students.

What folks are saying:

"His method of teaching through lecture, class discussions, lab days, allow students to both learn apply the different techniques. He really cares whether you understand the material or not, and will always make sure to answer any questions."

"One of my favorite classes I have taken during my time at UT. Course was very hands-on and I liked how applicable the topics were to real life situations. I also enjoyed how we used regression techniques on real data."

The Syllabus

Everything's on Canvas…

Homework

- Get R, and RStudio (we'll install SDSRegressionR the package on Friday).
  - *Make sure* to get the most recent version of RStudio.
  - You may as well update R also.
- Fill out the First Day Survey on Canvas by 2pm Thursday.
- Reading Quiz 1 due by Wednesday **10am**