Intro to data science (QR2)

UATX

Introductions

- name, where you're from, and what did you do over break?
- what are you hoping to get out of this class?

Warm up: marijuana

Let's play a game...

- You will flip a coin...
- if heads, you will write down the number 1 if your social security number ends with an even digit, otherwise write down 0.
- if tails, you will write down the number 1 if you ever used marijuana (smoke, edible, ...), otherwise 0.

Question: What percentage of students have used marijuana?

Warm up: marijuana

Law of total probability

$$\begin{aligned} P(1) &= P(1, \text{heads}) + P(1, \text{tails}) \\ &= P(1 \mid \text{heads})P(\text{heads}) + P(1 \mid \text{tails})P(\text{tails}) \end{aligned}$$

The bar "|" means conditional probability – like fixing a known state of the world.

When you're done, tell me whether you wrote down a 1 or 0!

What is this class all about?

We will learn methods and tools for advanced quantitative reasoning, i.e., how to effectively debate with data.

$\rightarrow$ Emphasis on data, uncertainty, and statistical modeling
$\rightarrow$ Entry point into higher level coursework in machine learning

This course will be a mixture of practice and principle, and you will leave course with healthy skepticism and a foundational toolbox.

Class structure

Foundational topics aka principle (weeks 1 through 5)

(1) Data
(2) Probability
(3) Prediction and Regression

Practice (weeks 6 through 10)

(1) Precision
(2) Causality

Class structure

Before class

- readings
- explore our coding tools

During class

- lecture and discussion

After class

- homeworks (free response and computing, one per week)

Class structure

Evaluation

- homeworks (20%)
- in-class midterm (40%)
- research project (30%)
- engagement/participation (10%)

Class structure

Evaluation

– homeworks (20%)
– in-class midterm (40%)
– research project (30%)
– engagement/participation (10%)

Research project

– 2 per group
– conduct your own quantitative analysis. We will provide real-world data and a question prompt. More to come!

Research project

The deliverables will be:

(i) in-class presentation and
(ii) written report of your findings.

Midterm

Written and in-class on 10/23. Mixture of T/F, multiple choice, and free response.

Github

All course material (exercises, computing, resources) will flow from the following github site:

https://github.com/dpuelz/Quantitative-Reasoning-II

Please bookmark this and visit it often. If you're fancy, fork/clone this repo locally on your computer.

Rmarkdown

You will turn in all homeworks in "knitted `Rmarkdown` format."

RMarkdown is an authoring format that allows you to create dynamic documents with R code embedded in them.

- integrate R code, output, and visualizations seamlessly.
- generate reports in various formats (HTML, PDF, Word, etc.).
- reproducible data science: Share your code and results in one document.
- easy to learn and use for data scientists and researchers.
- supports interactive elements like Shiny apps and HTML widgets.

Basic `Rmarkdown` workflow

1. install R and RStudio (if you haven't already).
2. create a new `Rmarkdown` document in RStudio.
3. write your narrative text using Markdown syntax.
4. embed R code chunks to execute code.
5. "knit" (i.e. compile) the document to generate the output in the desired format.

Course expectations

- collaborate with your fellow students
- engage with the readings and in class discussions
- utilize office hours and help sessions!
- I will be asking a lot of you because I know you're excellent students :)
- keep up with the fast pace and have fun!