Agenda

1. Questionnaire
2. Syllabus
3. Course objectives
4. What is a model? Why do we use models?
5. Terms: categorical/quantitative, response/explanatory, parameter/statistic
6. Motivating examples
7. The four-step process of statistical modeling, C-F-A-U

Questionnaire You should all have copies of the questionnaire I handed out. It starts something like this:

Full Name	Prefer to be called	Pronouns	Email Address
Amelia McNamara	Amelia (or, Professor McNamara)	she/her	amcnamara@smith.edu

Please fill out the green sheet and return it to me by the end of class. This is how I get to know you, and keep track of how many people are interested in the class.

Syllabus You should also have a paper copy of the syllabus, or you can look online at `http://bit.ly/sds291`. We'll go over some important pieces, but please plan to read through it in detail on your own.

Course Objectives By the end of this course I want you to be able to:

- Construct an appropriate model based on data
- Modify your model to adapt to new data or knowledge
- Understand and verify the assumptions upon which your model is based
- Understand and explain the types of inference justifiable based on your data
- Present your findings to a general audience in both written and oral forms
- Work with messy data, perform exploratory data analysis, and use `R`

What is a model? Why do we use models?

Some terms

- Variable types:
  - Categorical (special type: binary)
  - Quantitative
- Model inputs and outputs:
  - Explanatory
  - Response
- Language from sampling:
  - Parameters
  - Statistics

Motivating examples Even though regression models are relatively simple, they are very widely used. The questions below set up a lot of *binary* distinctions (see what I did there?) but usually in statistics you can argue both ways.

- The lane effect in the Rio Olympics can be modeled using regression. `https://www.washingtonpost.com/news/wonk/wp/2016/09/01/these-charts-clearly-show-how-some-olympic-swimmers-may-have-gotten-an-unfair-advantage/`

  How many explanatory variables are they using in their model? Are the explanatory variables categorical or quantitative? Is the response categorical or quantitative? Is the model being used for prediction or description?

- How to avoid boring sunsets. `http://fivethirtyeight.com/features/how-to-avoid-boring-sunsets/`

  How many explanatory variables are they using in their model? Are the explanatory variables categorical or quantitative? Is the response categorical or quantitative? Is the model being used for prediction or description?

- (Old!) Predicting who will win a particular county. `http://www.nytimes.com/interactive/2016/02/09/us/politics/campaign-ad-tracking.html?_r=0`

  How many explanatory variables are they using in their model? Are the explanatory variables categorical or quantitative? Is the response categorical or quantitative? Is the model being used for prediction or description?

- Recitivism risk predicted using a model. `https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing`

  How many explanatory variables are they using in their model? Are the explanatory variables categorical or quantitative? Is the response categorical or quantitative? Is the model being used for prediction or description?

- (Old!) What is the jobless rate for people like you? NY Times article, `http://www.nytimes.com/interactive/2009/11/06/business/economy/unemployment-lines.html`

  How many explanatory variables are they using in their model? Are the explanatory variables categorical or quantitative? Is the response categorical or quantitative? Is the model being used for prediction or description?

The four-step process

1. **C**hoose

2. **F**it

3. **A**ssess

4. **U**se