Day 1, Morning Session: Sets, Statements, and Proof

Math and Code Bootcamp

Name

PhD Student/Post-Doc, Trinity College Dublin

September 2020

Roadmap for the week

- We're learning how to make inferences about a population from a sample
- Last time: We figured out how to determine if two samples are different or independent (diff-in-means, contingency tables)

Outline for today:

- From scatterplots to correlations
  - i.e. how similar are the data (does variation in one explain variation in the other)
- Bivariate regression
  - Assumptions
  - Estimation (i.e. drawing the "best" line through data)

Step 1: Standardizing variation in variables

$$\frac{x - \bar{x}}{s}$$

Example: Populations of New England states

	x	$\frac{x-\bar{x}}{s}$
CT	3.5mil	?
ME	1.3mil	?
MA	6.6mil	?
NH	1.3mil	?
RI	1.0mil	?
VT	0.6mil	?

Standardizing variables: Mean and SD in R

```r
# create vector
x <- c(3.5, 1.3, 6.6, 1.3, 1, .6)
# get mean and sd
c(round(mean(x), 2), round(sd(x), 2))
```

```
[1] 2.38 2.3
```

$$\bar{x} = 2.38 \quad s = 2.30$$

Wrap-up

Today we learned about...

- Correlations
- Simple linear regression:
    1. Assumptions
    2. Estimation