Intro to Sampling

Nate Wells

Math 141, 2/28/22

Outline

In this lecture, we will. . .

Outline

In this lecture, we will. . .

- Discuss random sampling: the heart of statistics!
- Perform a group sampling activity

Sampling Overview

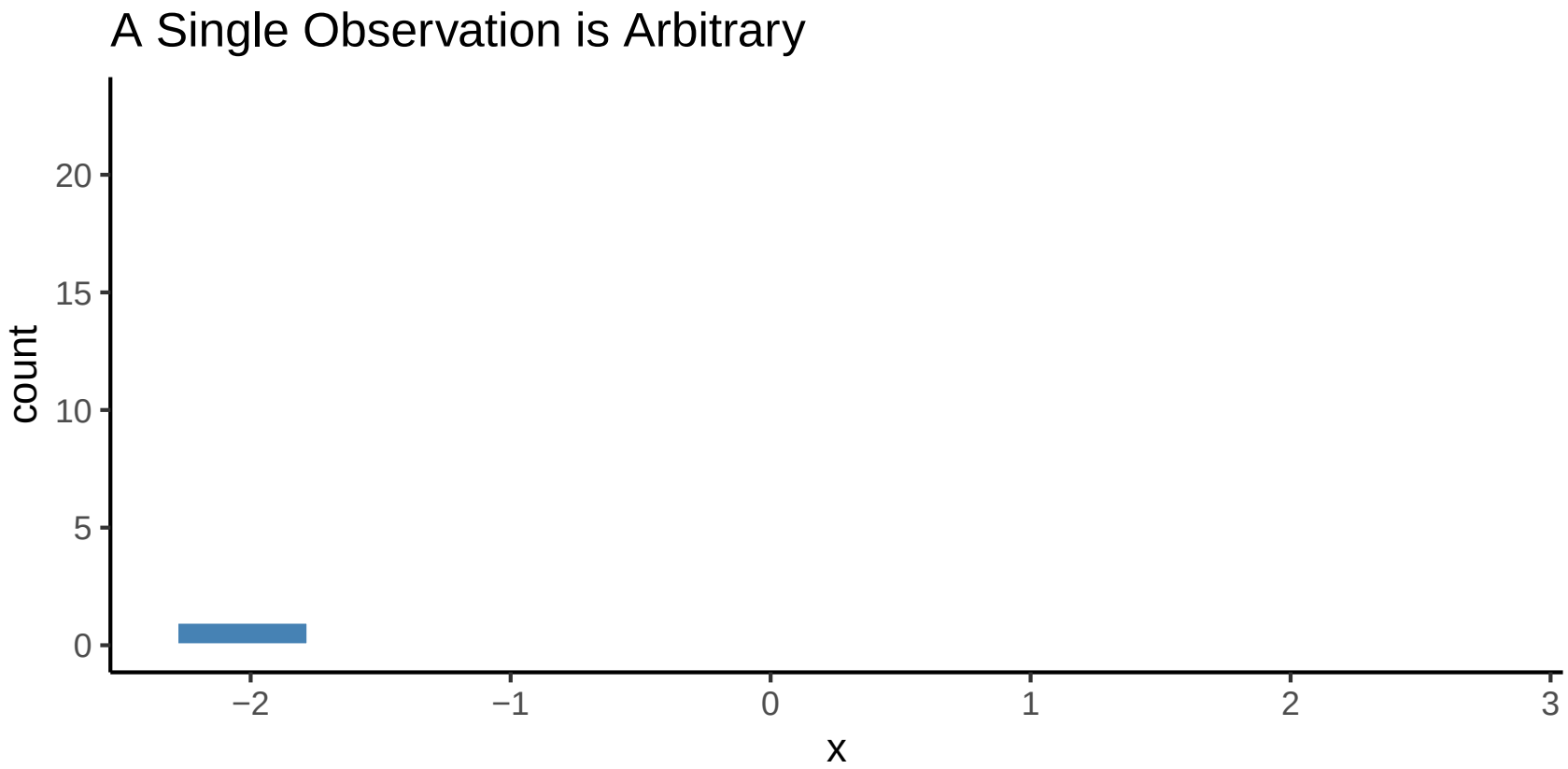
- The distribution of a data set allow us to quantify the shape, center, and spread of the data.

Sampling Overview

- The distribution of a data set allow us to quantify the shape, center, and spread of the data.
- While a single observation in a data set may appear arbitrary, repeated trials show that outcomes indeed follow prescribed patterns.

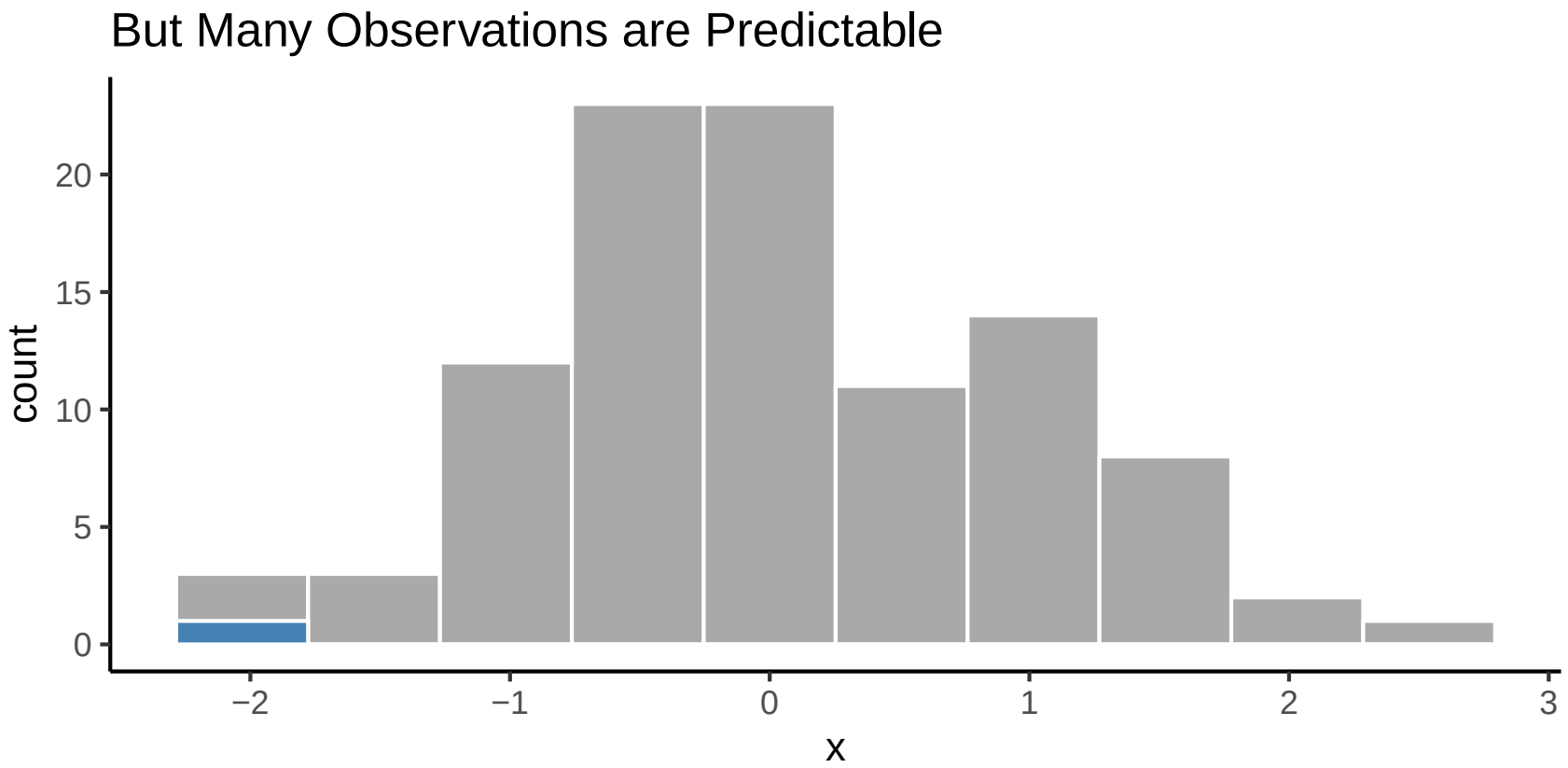
Sampling Overview

- The distribution of a data set allow us to quantify the shape, center, and spread of the data.
- While a single observation in a data set may appear arbitrary, repeated trials show that outcomes indeed follow prescribed patterns.



Sampling Overview

- The distribution of a data set allow us to quantify the shape, center, and spread of the data.
- While a single observation in a data set may appear arbitrary, repeated trials show that outcomes indeed follow prescribed patterns.



Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:

Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:
  - The *proportion* of U.S. voters who plan to vote for a particular presidential candidate.
  - The *mean* life-time earnings for Reed college graduates.

Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:
  - The *proportion* of U.S. voters who plan to vote for a particular presidential candidate.
  - The *mean* life-time earnings for Reed college graduates.
- But it is often prohibitively expensive, impractical, time-consuming, or impossible to perform a census to collect complete information on the population.

Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:
  - The *proportion* of U.S. voters who plan to vote for a particular presidential candidate.
  - The *mean* life-time earnings for Reed college graduates.
- But it is often prohibitively expensive, impractical, time-consuming, or impossible to perform a census to collect complete information on the population.
- So instead, researchers collect a sample and measure the value of a similar **statistic**.

Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:
  - The *proportion* of U.S. voters who plan to vote for a particular presidential candidate.
  - The *mean* life-time earnings for Reed college graduates.
- But it is often prohibitively expensive, impractical, time-consuming, or impossible to perform a census to collect complete information on the population.
- So instead, researchers collect a sample and measure the value of a similar **statistic**.
  - The *proportion* of voters in a sample of size 10 who plan to vote for the candidate.
  - The *mean* life-time earnings for 100 randomly chosen Reed graduates.

Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:
  - The *proportion* of U.S. voters who plan to vote for a particular presidential candidate.
  - The *mean* life-time earnings for Reed college graduates.
- But it is often prohibitively expensive, impractical, time-consuming, or impossible to perform a census to collect complete information on the population.
- So instead, researchers collect a sample and measure the value of a similar **statistic**.
  - The *proportion* of voters in a sample of size 10 who plan to vote for the candidate.
  - The *mean* life-time earnings for 100 randomly chosen Reed graduates.
- Each possible sample has its own corresponding statistic. The collection of all such statistics forms its own data set!

Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:
  - The *proportion* of U.S. voters who plan to vote for a particular presidential candidate.
  - The *mean* life-time earnings for Reed college graduates.
- But it is often prohibitively expensive, impractical, time-consuming, or impossible to perform a census to collect complete information on the population.
- So instead, researchers collect a sample and measure the value of a similar **statistic**.
  - The *proportion* of voters in a sample of size 10 who plan to vote for the candidate.
  - The *mean* life-time earnings for 100 randomly chosen Reed graduates.
- Each possible sample has its own corresponding statistic. The collection of all such statistics forms its own data set!
- As a data set, the statistics themselves have a *mean*, *standard deviation*, and a *5-number summary*.

Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:
  - The *proportion* of U.S. voters who plan to vote for a particular presidential candidate.
  - The *mean* life-time earnings for Reed college graduates.
- But it is often prohibitively expensive, impractical, time-consuming, or impossible to perform a census to collect complete information on the population.
- So instead, researchers collect a sample and measure the value of a similar **statistic**.
  - The *proportion* of voters in a sample of size 10 who plan to vote for the candidate.
  - The *mean* life-time earnings for 100 randomly chosen Reed graduates.
- Each possible sample has its own corresponding statistic. The collection of all such statistics forms its own data set!
- As a data set, the statistics themselves have a *mean*, *standard deviation*, and a *5-number summary*.
  - The mean of the statistic tells us its typical value in a randomly chosen sample.

Sampling Overview

- Researchers wish to investigate the value of a **parameter** in a population. Ex:
  - The *proportion* of U.S. voters who plan to vote for a particular presidential candidate.
  - The *mean* life-time earnings for Reed college graduates.
- But it is often prohibitively expensive, impractical, time-consuming, or impossible to perform a census to collect complete information on the population.
- So instead, researchers collect a sample and measure the value of a similar **statistic**.
  - The *proportion* of voters in a sample of size 10 who plan to vote for the candidate.
  - The *mean* life-time earnings for 100 randomly chosen Reed graduates.
- Each possible sample has its own corresponding statistic. The collection of all such statistics forms its own data set!
- As a data set, the statistics themselves have a *mean*, *standard deviation*, and a *5-number summary*.
  - The mean of the statistic tells us its typical value in a randomly chosen sample.
  - The standard deviation of the statistic tells us how its value fluctuates from sample to sample.

Population vs. Distribution of Sample Statistics

Sampling Activity

Goal: Describe the distribution of a statistic by observing its average value and variability between samples. Compare to the distribution of the variable in the population.

Activity

1. Thoroughly shuffle one of your group's deck of cards.
2. Draw 10 cards from the deck (without replacement) to form a sample.
3. Compute the mean value of your cards (counting Aces as 1 and Faces as 10)
4. Write the value of the mean on a sticky note and add to chalkboard.
5. Repeat steps 1 - 4 an additional four times.

Discussion

Answer the following questions in your group:

- What is the theoretical mean value for the data set of card values?
- How does the distribution of sample means compare to the distribution of card values?
- What is the relationship between the centers of the two distributions?
- Which distribution appears to have more variability?
- How do the shapes of the two distributions compare?
- What does the variability of sample means suggest about the means in repeated samples?