DATA 8
Spring 2022

Lecture 24

Interpreting Confidence

Announcements

Estimation

Inference: Estimation

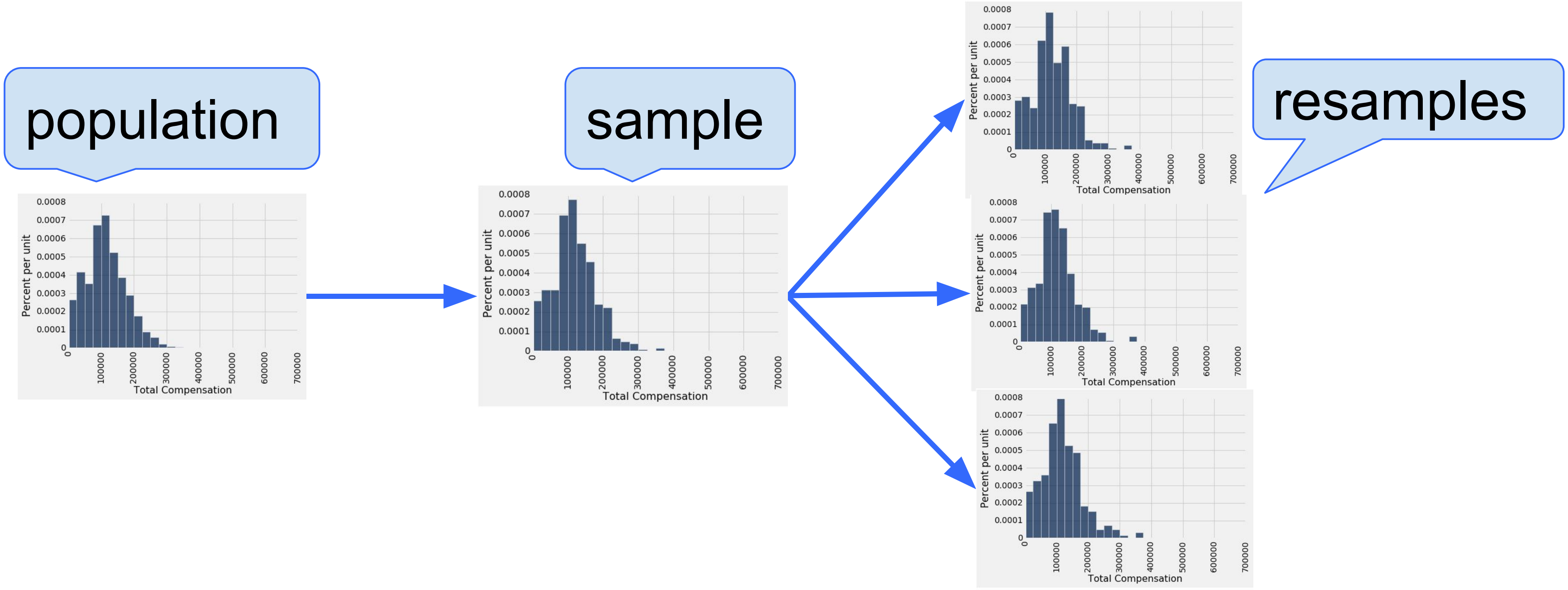
- **Parameter**: Fixed quantity in the population
- How can we figure out the value of an unknown parameter?
- If you don't have a census:
  - Take a random sample from the population
  - Use a statistic as an **estimate** of the parameter
- One sample ➜ One estimate
- But the random sample could have come out differently
- And so the estimate could have been different
- We need to know how different it could have been

Where to Get Another Sample?

- We want to understand variability of our estimate
- We only have the **sample**
- To get many values of the estimate, we need many random samples
- We can’t go back and sample again from the population

The Bootstrap

Why the Bootstrap Works



All of these look pretty similar, most likely.

Key to Resampling

- From the original sample,
  - draw at random
  - with replacement
  - as many values as the original sample contained

- The size of the new sample has to be the same as the original one, so that the two estimates are comparable

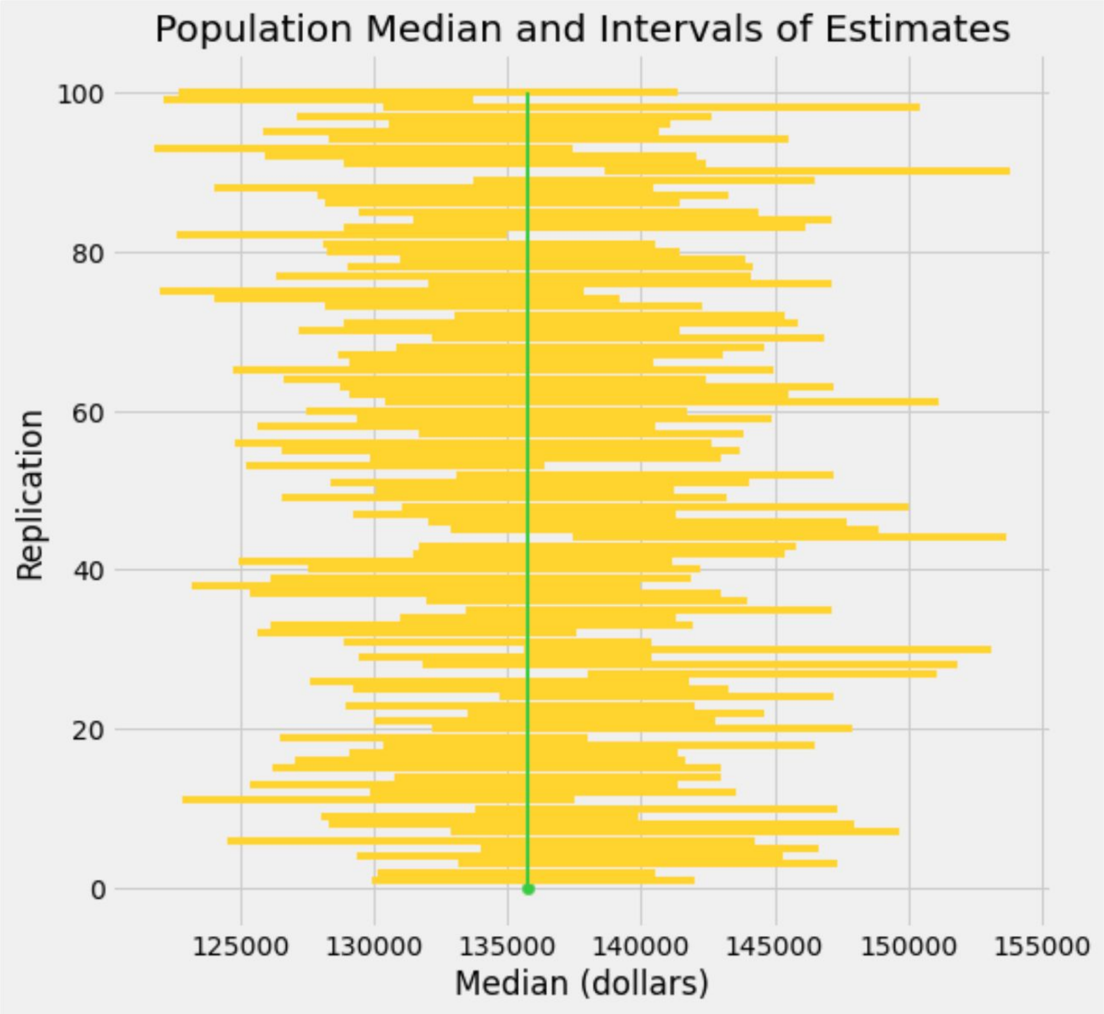
(Demo)

Confidence Intervals

95% Confidence Interval

- Interval of **estimates of a parameter**
- Based on random sampling
- 95% is called the confidence level
  - Could be any percent between 0 and 100
  - Higher level means wider intervals
- A “good” interval is one that contains the parameter
- The **confidence is in the process** that creates the interval:
  - It generates a “good” interval about 95% of the time.

(Demo)



The Meaning of 95% confidence

The green dot is the parameter. It is fixed.

Each yellow line is a 95% confidence interval based on a fresh sample from the population

There are 100 intervals. 95 of them contain the parameter.

Use Methods Appropriately

When *Not* to Use Our Bootstrap Method

- If you're trying to estimate any parameter that's greatly affected by rare elements of the population
  - Very high or very low percentiles, or min and max
- If the probability distribution of your statistic is not roughly bell shaped (the shape of the empirical distribution will be a clue)
- If the original sample is very small

(Demo)

Can You Use a CI Like This?

By our calculation, an approximate 95% confidence interval for the average age of the mothers in the population is (26.9, 27.6) years.

True or False:

- About 95% of the mothers in the population were between 26.9 years and 27.6 years old.

Answer: False. We're estimating that their **average age** is in this interval.

(Demo)

Is This What a CI Means?

An approximate 95% confidence interval for the average age of the mothers in the population is (26.9, 27.6) years.

True or False:

- There is a 0.95 probability that the average age of mothers in the population is in the range 26.9 to 27.6 years.

Answer: False. The parameter is fixed, and the interval (26.9, 27.2) is fixed. The parameter is either in that interval, or not. Once you've picked an interval, there's no probability involved.

95% Confidence

- Interval of estimates of a parameter
- Based on random sampling
- The process results in a random interval
- A “good” interval is one that contains the parameter
- The **confidence is in the process** that creates the interval:
    - It generates a “good” interval with chance 95%

Confidence Intervals For Testing

Using a CI for Testing

- Null hypothesis: **Population average = *x***
- Alternative hypothesis: **Population average ≠ *x***
- Cutoff for p-value: *p*%
- Method:
  - Construct a (100-*p*)% confidence interval for the population average
  - If *x* is not in the interval, reject the null
  - If *x* is in the interval, can’t reject the null