
DATA 8

Fall 2016

# Lecture 24, October 21

## Using Confidence Intervals

Slides created by Ani Adhikari and John DeNero

# Announcements

- Homework will be assigned today. Due next Wed by 7 pm for an “early” point, or Thurs by 7.
- Homework regrade requests requests are due Friday at 7PM the week that homework grades are released.
- Project 2 will be released next week!
- I’ve posted on Piazza about courses to consider if you are interested in data science.

# 95% Confidence Interval

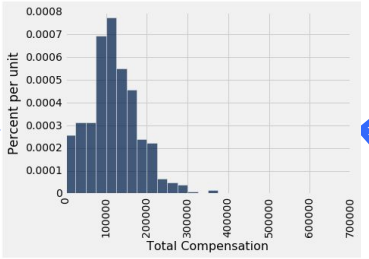
- Interval of **estimates of a parameter**
- Based on random sampling
- 95% is called the **confidence level**
  - Could be any percent between 0 and 100
  - Bigger is better (more confidence in coverage)
- The **confidence is in the process** that generated the interval:
  - It generates a “good” interval about 95% of the time.

# Inference Using the Bootstrap

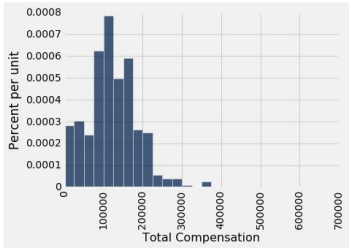
population

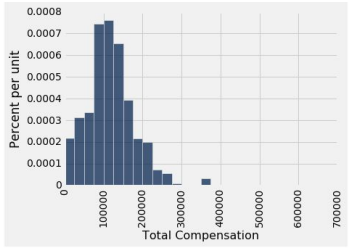
?

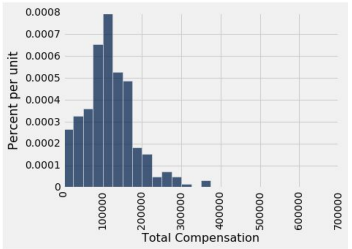
sample




resamples




All of these look pretty similar, most likely.

# Bootstrap Percentile Method

- For constructing a confidence interval for an unknown parameter
- Starting point: one large random sample
- One replication:
  - Bootstrap the sample to get a “resample”
  - Get an estimate based on the resample
- Repeat several thousand times (10,000 recommended)
- For an approximate 80% confidence interval, take the 10th and 90th percentiles of all the bootstrap estimates

(Demo)

# When *Not* to Use The Bootstrap

- If you're trying to estimate very high or very low percentiles, or min and max
- If you're trying to estimate any parameter that's greatly affected by rare elements of the population
- 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

# Can You Use a CI for 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)

# 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