Interpreting Confidence Intervals

LECTURE 24

CSCI 3022

Maribeth Oscamou

Content credit: [Acknowledgments]

Announcements

- **Homework 9** due Thursday
- Nb9 session yesterday (with our TA): video posted on Canvas
- Quiz 8 Friday;
  - Scope: HW 8, nb8, Lessons 19 & 20

Today's Roadmap

CSCI 3022

- Recap of Confidence Intervals
- Assumptions to use Bootstrapping
- Using Confidence Intervals for Hypothesis Testing

Inference: Parameter Estimation

Inference is all about **drawing conclusions** about **population parameters**, given only a **random sample**.

population

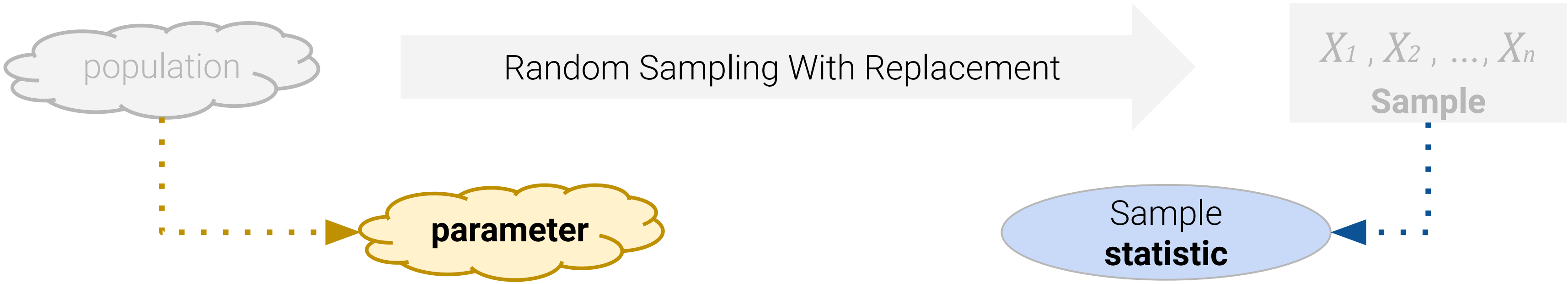
Random Sampling With Replacement

$X_1, X_2, \ldots, X_n$

Sample

[Terminology] Parameters, Statistics, and Estimators

Inference is all about **drawing conclusions** about **population parameters**, given only a **random sample**.



A **parameter** is a numerical function of the **population**.

Population mean

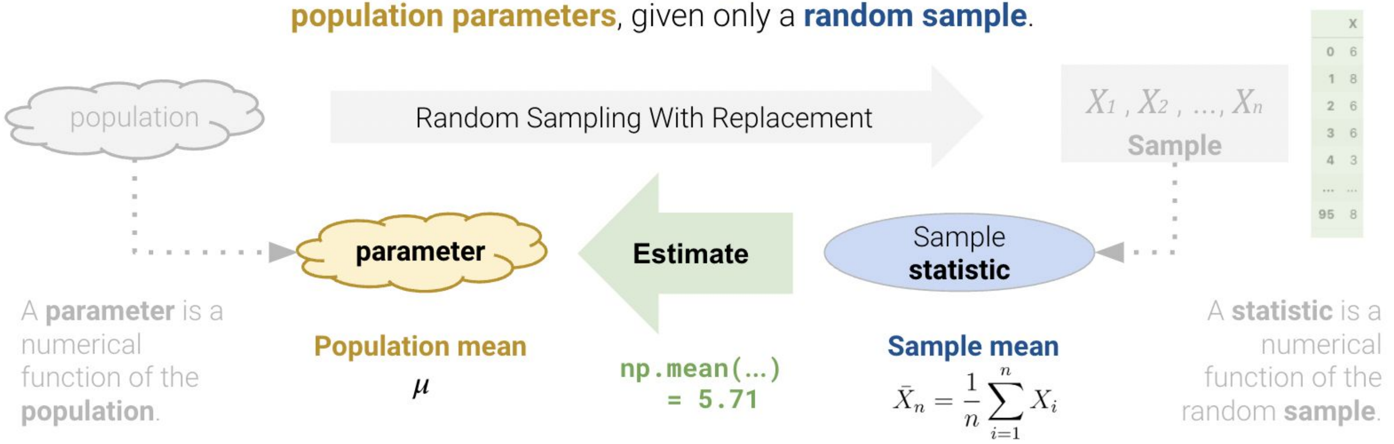
μ

Sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

A **statistic** is a numerical function of the random **sample**.

Inference is all about **drawing conclusions** about **population parameters**, given only a **random sample**.



We can then use the sample statistic as an **estimator** of the true population parameter.

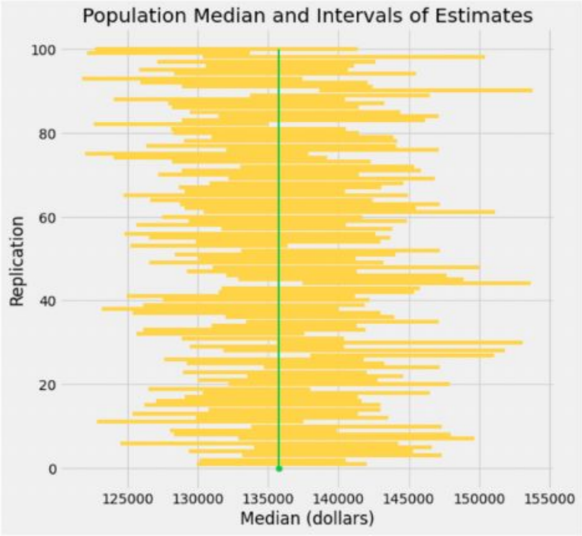
Since our **sample is random**, our statistic (which we use as our estimator) **could have been different.**

Example: When we use the sample mean to estimate the population mean, our estimator is almost always going to be somewhat off.

We want to add ***error bars on our estimate that incorporate the standard error*** (i.e. standard deviation of the sampling distribution). This is what **Confidence Intervals do!**

What is a Confidence Interval?

- A confidence interval is a range of estimates for an unknown population parameter based on a sample statistic
- A confidence interval is computed at a *confidence level (L%)*
- Confidence intervals focus on the **size of an effect** and the **uncertainty around the estimate** rather than just whether an effect exists.



The Meaning of 95% confidence

The **green line** is the parameter value.
It is fixed and unknown.
(For this demo we we had access to the population but you won't in practice.)

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

There are **100 intervals**. We expect **roughly 95** to contain the parameter.

What does a X% CI Mean

This is the source of randomness in the CI calculation.

If we repeatedly follow the same process of

1. **randomly sampling** from the **population** a **sample** of the **same size**
2. and applying the same **CI calculations**

then we expect that **X% of the intervals** will contain the **population parameter**.*

***Fine Print:** Assuming the samples are large enough and the statistic is "well behaved" and we run MANY bootstrap samples…*

Can You Use a CI Like This?

Suppose you calculate an approximate 95% confidence interval for the average age of the mothers in the baby data 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.

A). True   B). False

(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?

Strictly speaking, what is the best interpretation of a 95% confidence interval for the mean?

A) ○ If repeated samples were taken and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the population mean.

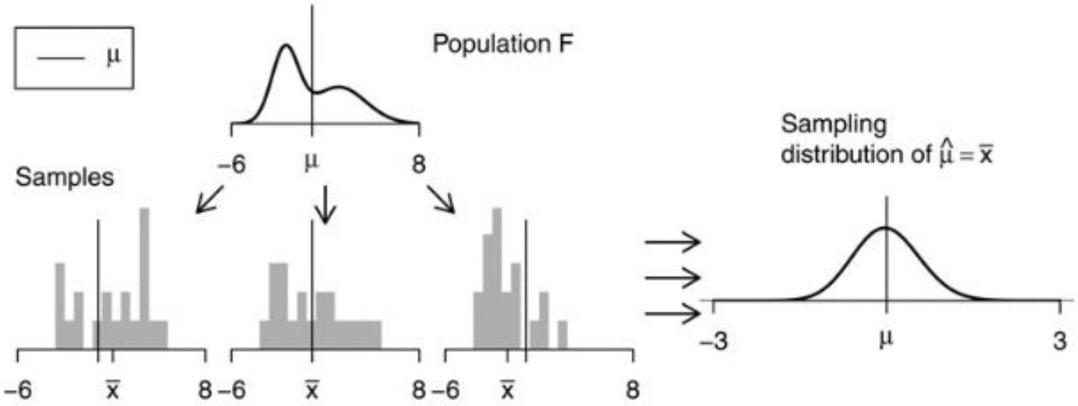
B) ○ A 95% confidence interval has a 0.95 probability of containing the population mean.

C) ○ 95% of the population distribution is contained in the confidence interval.

The Sampling Distribution of a Statistic

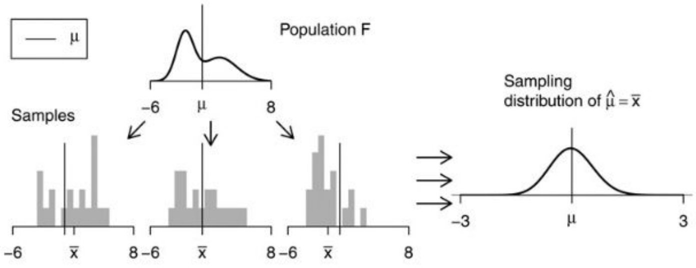
"Ideal world":

- To determine the properties (i.e. shape and standard error) of the sampling distribution of an estimator, we'd need to have access to the population.
- Sampling distributions of a statistic are obtained by drawing repeated samples from the population, computing the statistic of interest for each and collecting (an infinite number of) those statistics as the sampling distribution.



The Sampling Distribution of a Statistic

Ideal world:

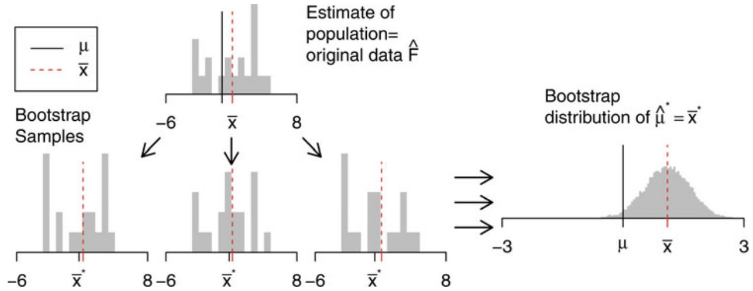


Reality: We don't know the population distribution.

Idea: Treat our random sample as a "population", and resample from it.

Intuition: a random sample resembles the population, so a random resample resembles a random sample.

Bootstrap World:



Bootstrap World: Draw repeated samples from an estimate of the population, computing the statistic of interest for each, and collecting those statistics. The ***bootstrapped distribution is centered at the observed statistic,*** not the actual population parameter.

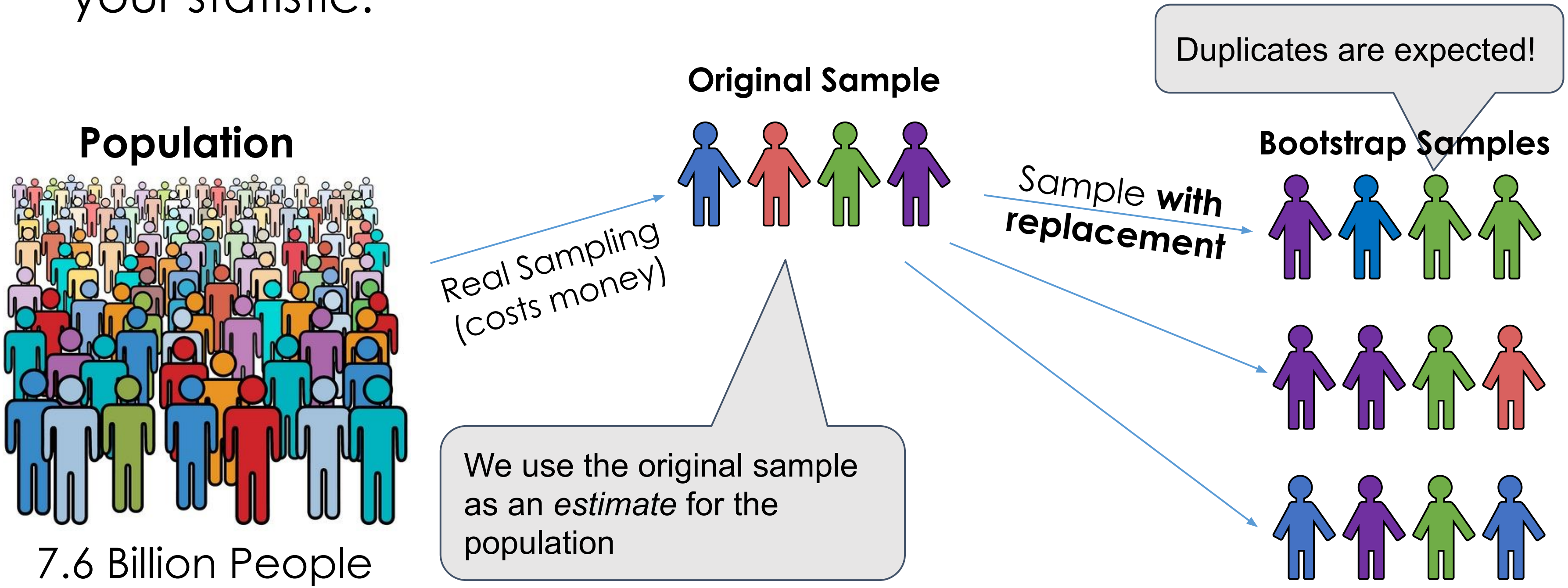
Calculating Confidence Intervals

- Method 1: Bootstrapping
- Method 2: Using the Central Limit Theorem if it applies

(Demo)

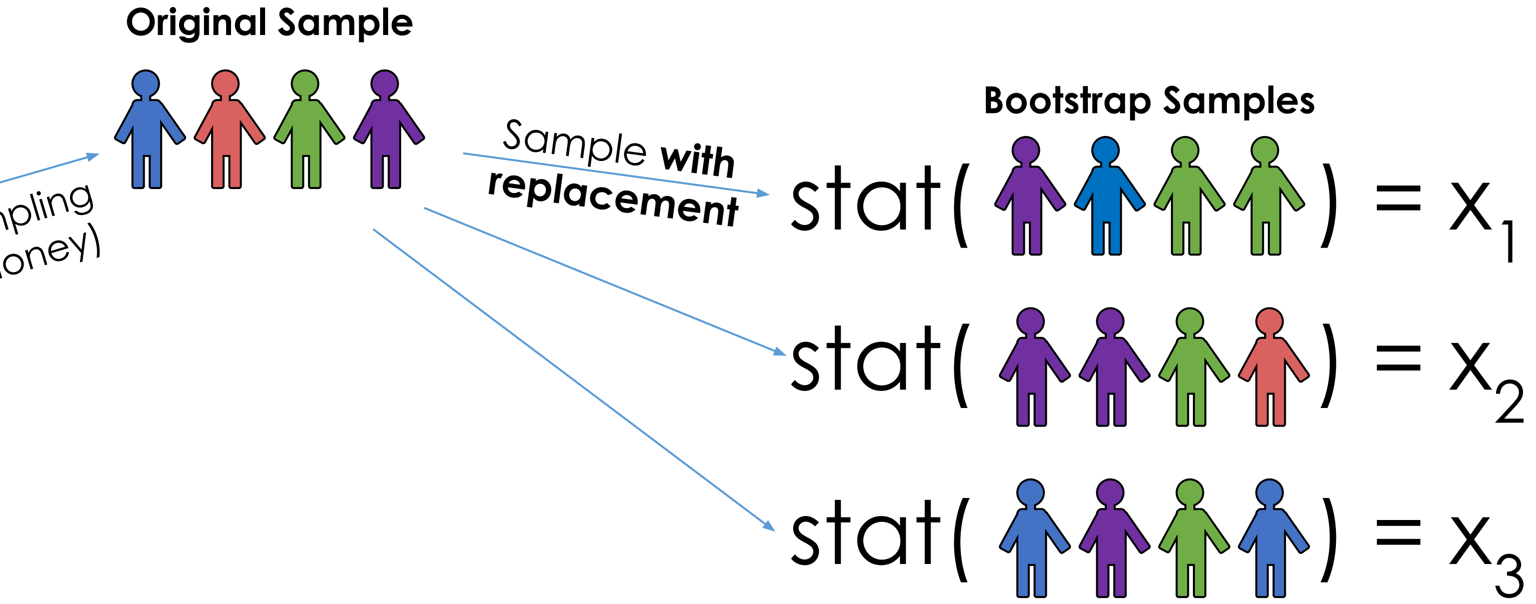
Bootstrap the Distribution of a Statistic

Simulation method to estimate the sample distribution of your statistic.



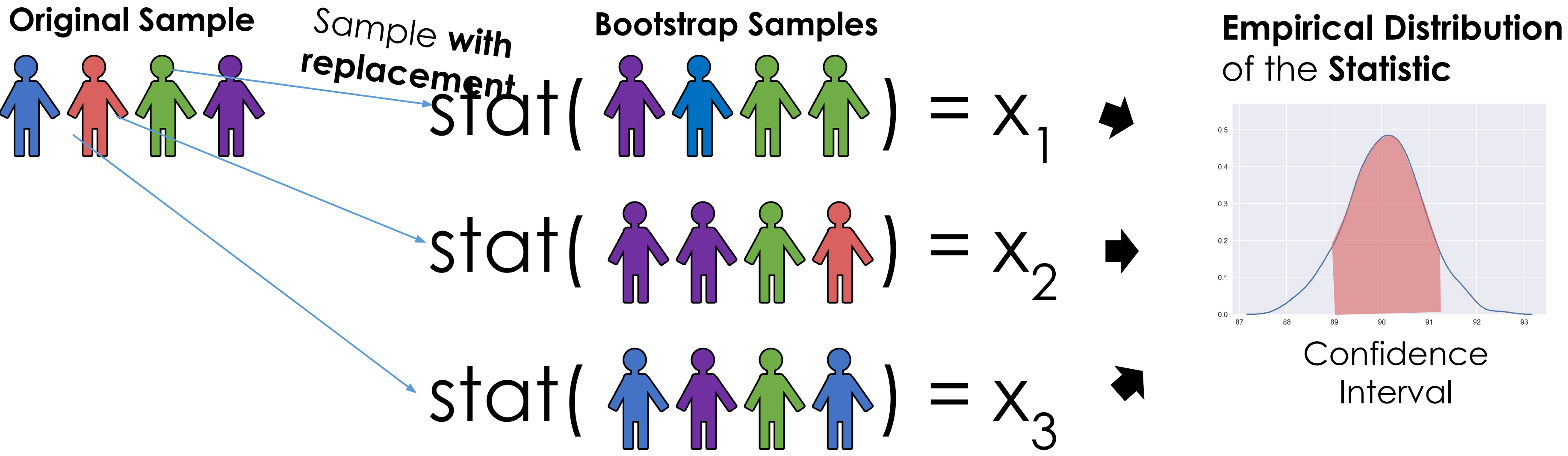
Bootstrap the Distribution of a Statistic

Simulation method to estimate the sample distribution of your statistic.



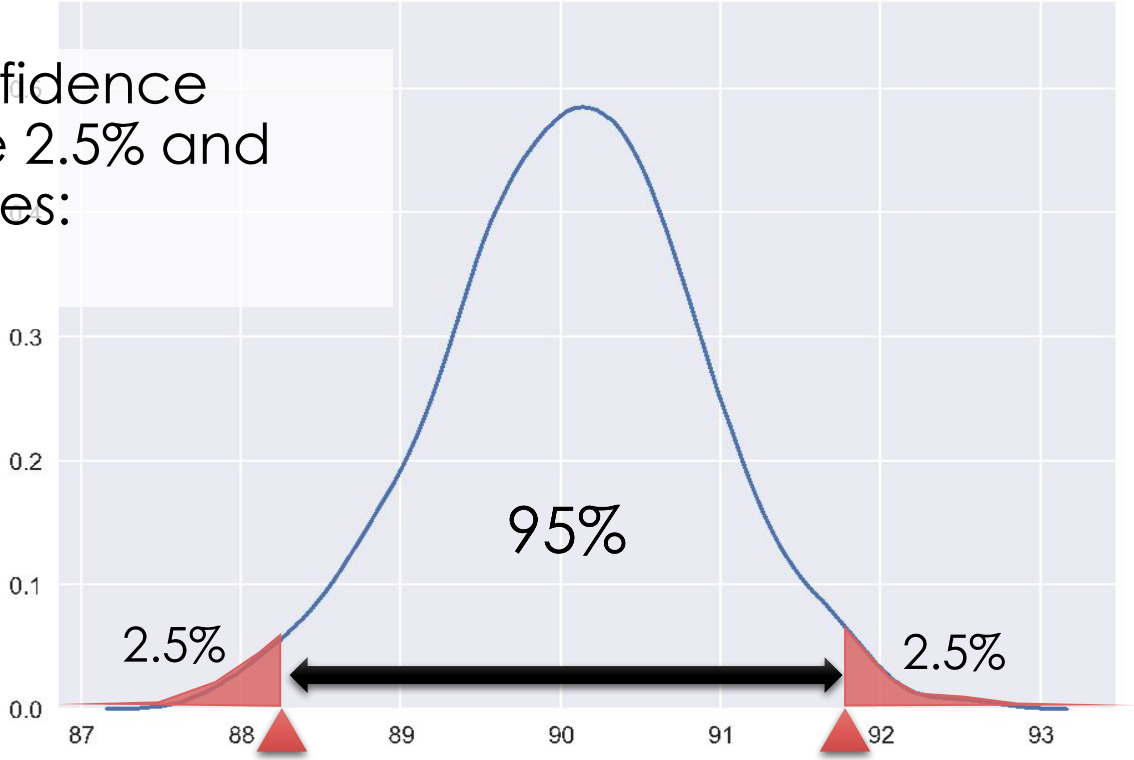
Bootstrap the Distribution of a Statistic

Simulation method to estimate the sample distribution of your statistic.

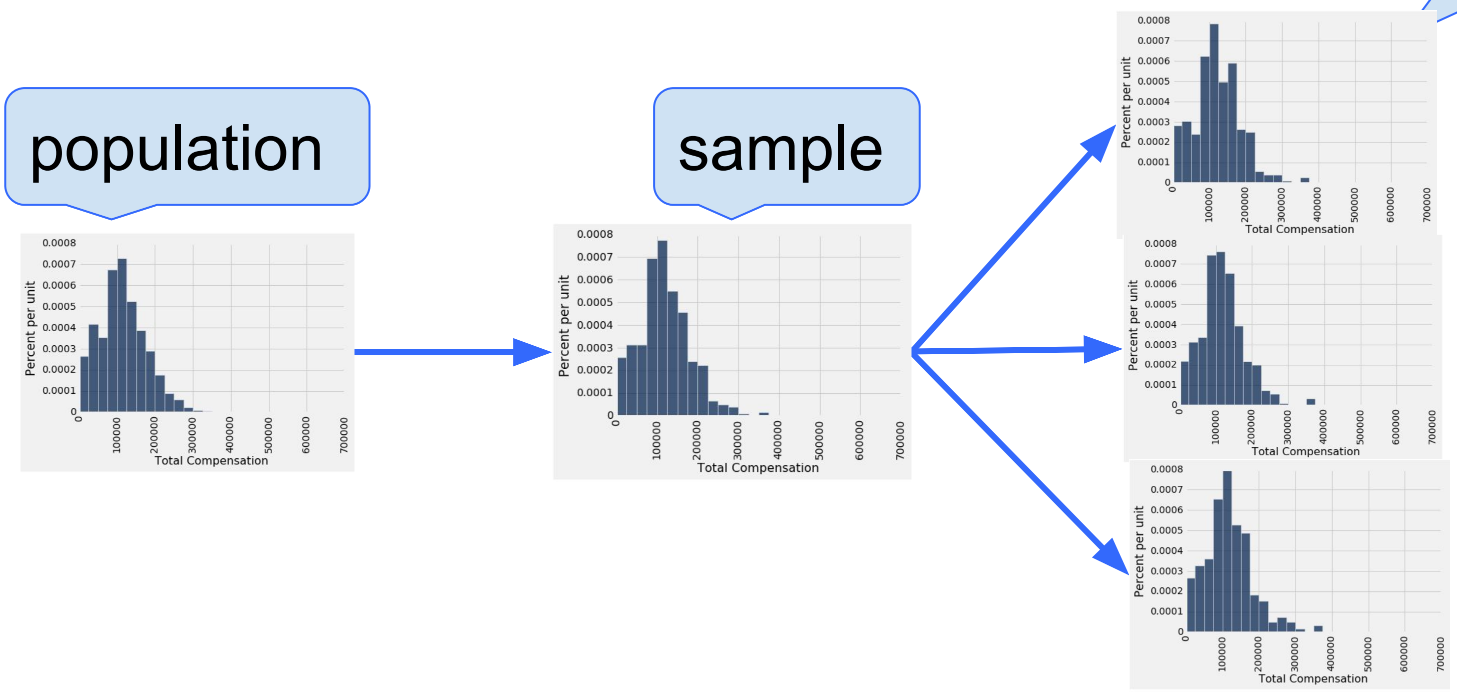


Bootstrap Confidence Interval

Construct a 95% confidence interval by taking the 2.5% and (100 - 2.5)% percentiles:

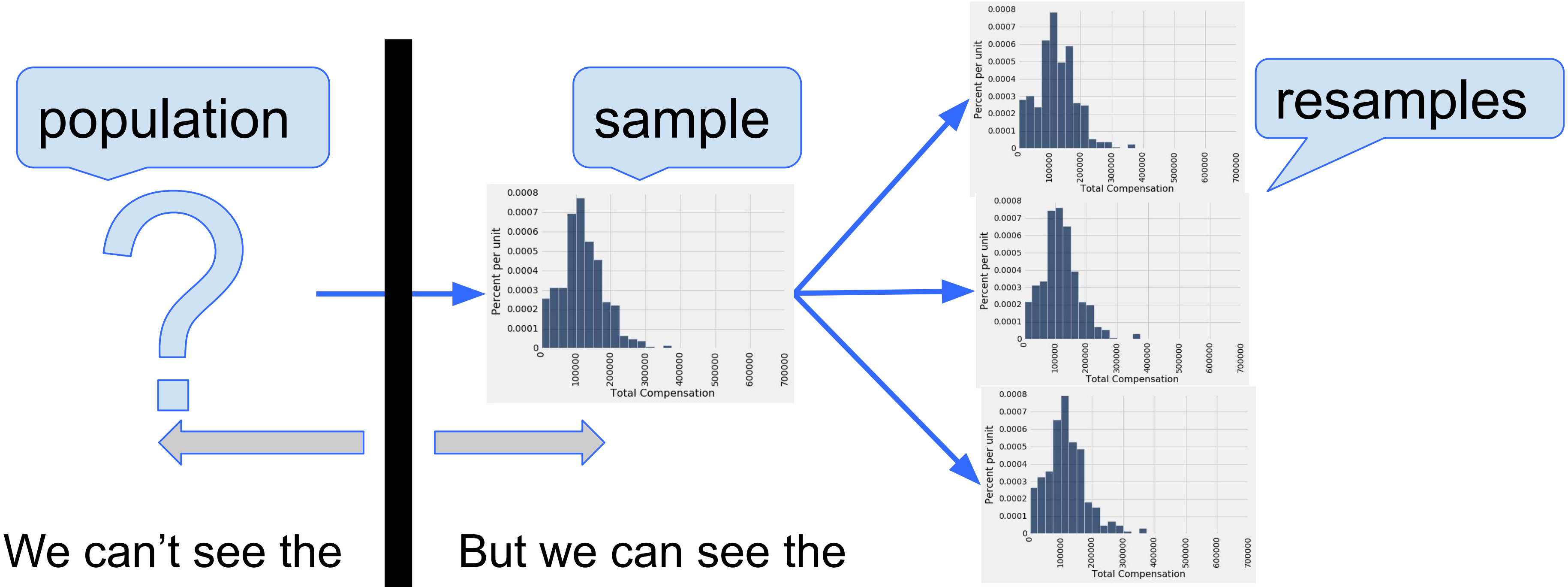


Why the Bootstrap Works



All of these look pretty similar, most likely.

Why We Need the Bootstrap

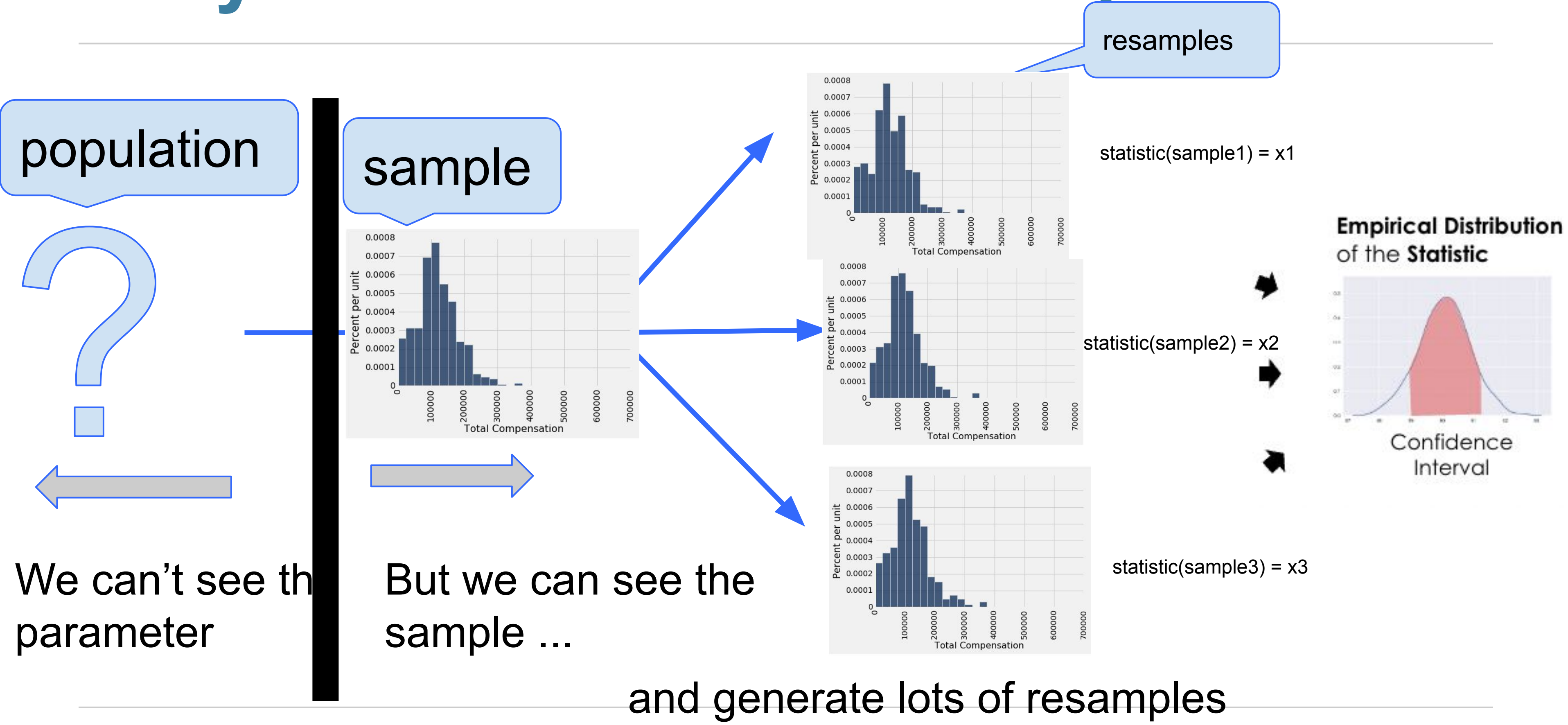


We can't see the parameter

But we can see the sample ...

and generate lots of resamples

Why We Need the Bootstrap



The Bootstrap in words

- From the original sample,
  - draw at random
  - **<u>with replacement</u>**
    - Otherwise you would always get the same sample
  - **Use the same sample size** as the original sample
    - The size of the new sample has to be the same as the original one, so that the standard errors reflect the actual data, rather than a hypothetical larger or smaller dataset.
- For each sample, **compute the statistic**
- Compute **empirical distribution of** the **statistics**

(Demo)

The Bootstrap Principle

<u>Example</u>: Let's write pseudocode for how we would bootstrap a 90% CI for the median

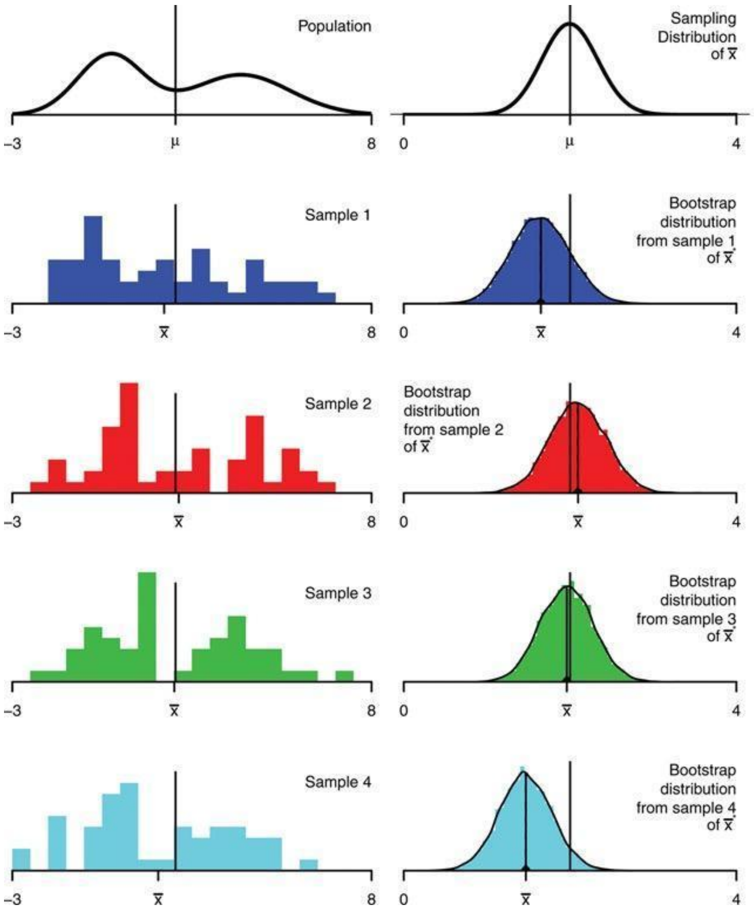
(For accuracy, use 10,000 simulations when bootstrapping)

```
original_sample
new=[ ]
nsamp=10000

for i in range(nsamp):
        resample=np.random.choice(original_sample, size=len(original_sample), replace=True)
        new.append(np.median(resample))

CI =np.percentile(new, [5,95])
```

Use Methods Appropriately



Using this example, discuss the following:

a). How similar/different are the **sample distributions** from the ***population distribution?***

b). How similar/different are the **spreads** of the bootstrapped distributions:

i). From one another?

ii). From the sampling distribution?

c). How similar/different are the **shapes** of the bootstrapped distributions:

i). From one another?

ii). From the sampling distribution?

d). How similar/different is the **mean** of the bootstrapped distributions:

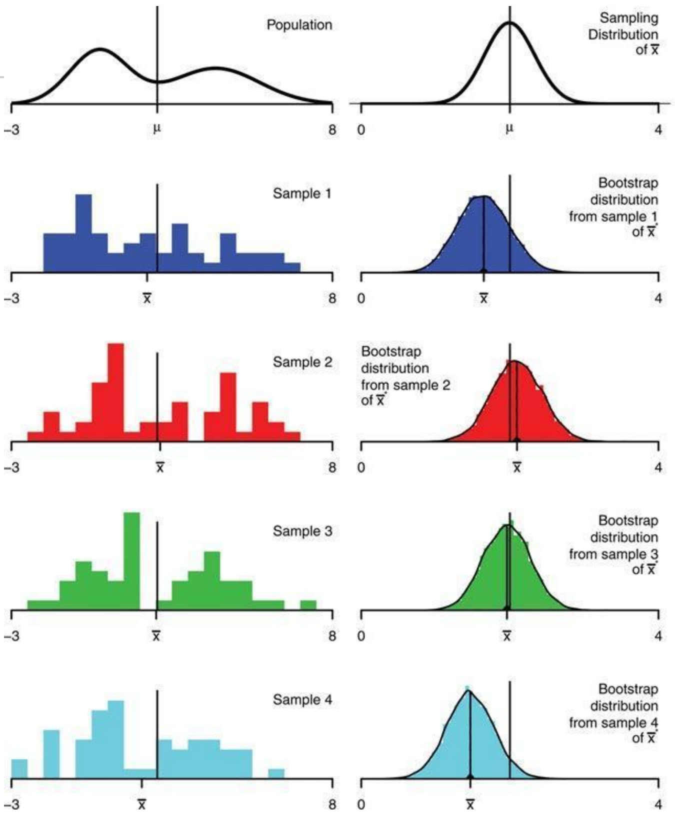
From one another?

From the sampling distribution?

Bootstrap Discussion

- The quality of our bootstrapped distribution depends on the quality of our original sample.
  - ***If our original sample was not representative of the population, bootstrap is next to useless.***

- The **bootstrapped sampling distribution of an estimator** does not exactly match the **sampling distribution of that estimator**.
  - The center and spread are both wrong (but often close).
- **The center** of the bootstrapped distribution is the estimator applied to our original sample.
  - We have no way of recovering the estimator’s true expected value
- **The variance** of the bootstrapped distribution **is often close to the true variance of the estimator.**



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

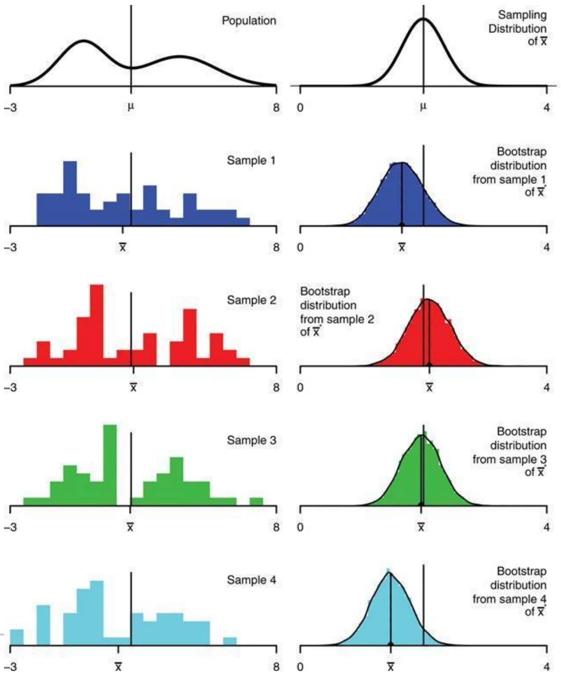
- If the probability distribution ***of your sample statistic*** is not **roughly bell shaped**

- If the original **sample is very small**

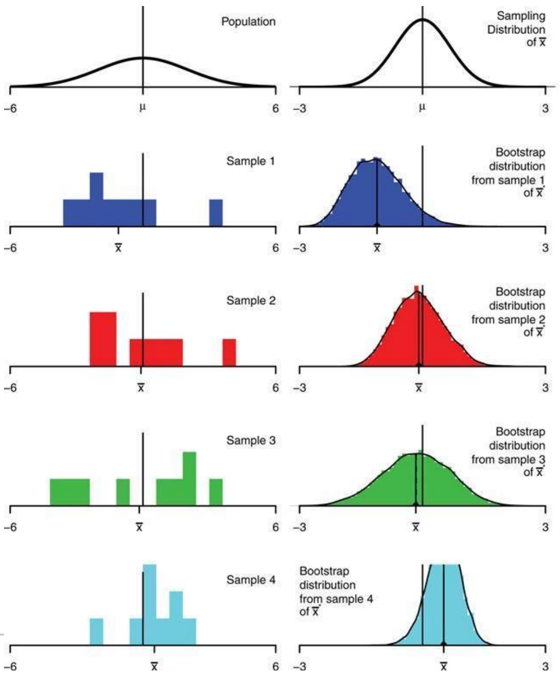
(Demo)

Warning: Bootstrapped Confidence Intervals are not reliable if the original **sample is very small**

Bootstrap for the Mean Sample size=50



Different population; sample size = 9



Notice: **For small sample sizes:**

The spreads and shapes of the bootstrap distributions vary substantially, because the spreads and shapes of the samples vary substantially.

As a result, ***bootstrap confidence interval widths vary substantially*** **and are less reliable.**

Confidence Intervals For Hypothesis 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

(Demo)

CI and Hypothesis Testing

Ex). You're evaluating a training program by comparing the test scores of program participants to those who study on their own. You decide that the difference between these two groups must be *at least 5 points* to represent a practically meaningful effect size. (An effect i.e. difference between these two groups of 4 points or less is too small to care about).

Suppose you conduct a randomized control study and hypothesis test (with the null being no difference) on the training program described above. Suppose the test ***is statistically significan***t (at the 5% significance level) and produces an estimated effect (i.e. a difference between groups of 9 points). This effect looks good because it's greater than our smallest meaningful effect size of 5.

However, this estimate doesn't incorporate the margin of error. The confidence interval (CI) provides that crucial information.

CI and Hypothesis Testing

Ex). You're evaluating a training program by comparing the test scores of program participants to those who study on their own. You decide that the difference between these two groups must be at least five points to represent a practically meaningful effect size. An effect of 4 points or less is too small to care about.

Suppose you conduct a randomized control study and hypothesis test (with the null being no difference) on the training program described above. Suppose the test is statistically significant (at the 5% significance level) and produces an estimated effect of 9. This effect looks good because it's greater than our smallest meaningful effect size of 5. ***However, this estimate doesn't incorporate the margin of error. The confidence interval (CI) provides that crucial information.***

a). Suppose you are told the 95% confidence interval for the difference between groups in example 1 is [3, 15]. (Recall, you decide that the difference between these two groups must be at least 5 points to represent a practically meaningful effect size) What does this Confidence Interval (CI) tell us?

CI and Hypothesis Testing

Ex). You're evaluating a training program by comparing the test scores of program participants to those who study on their own. You decide that the difference between these two groups must be at least five points to represent a practically meaningful effect size. An effect of 4 points or less is too small to care about.

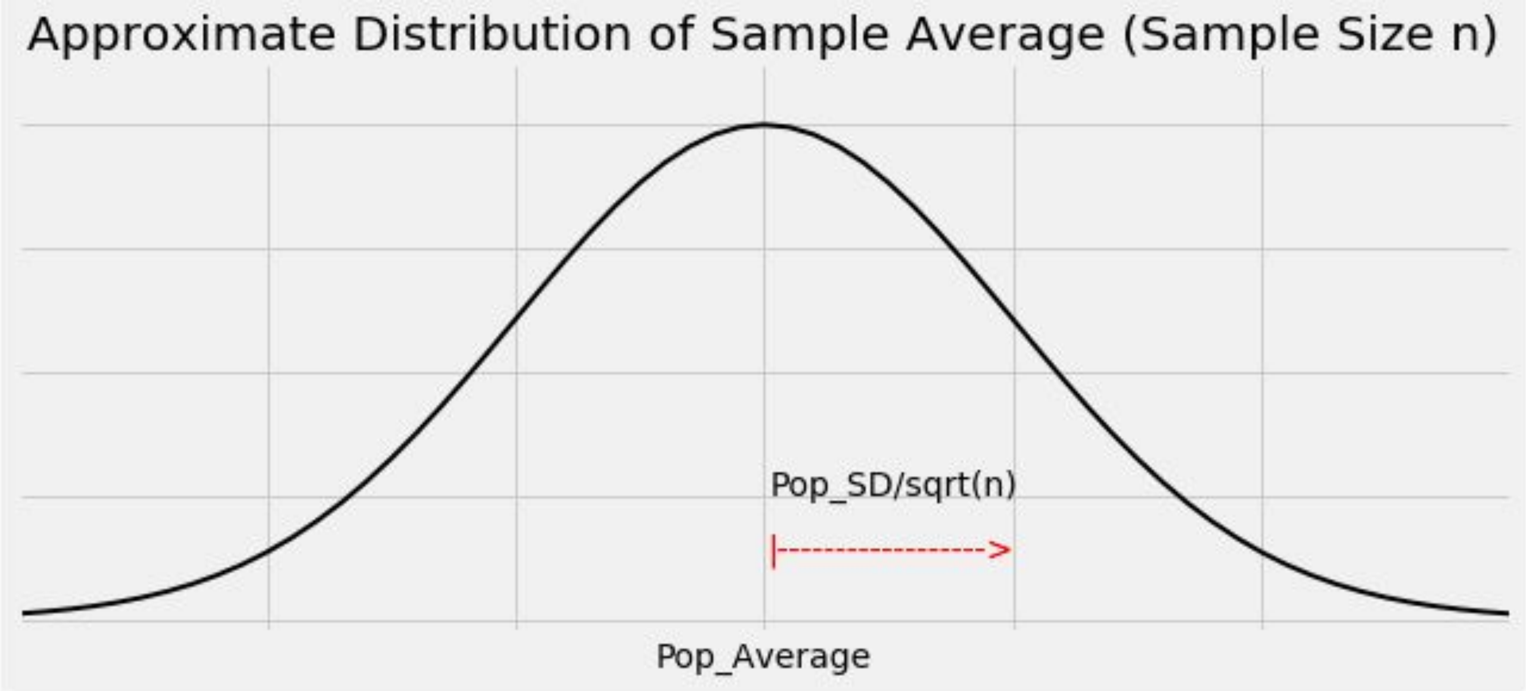
Suppose you conduct a randomized control study and hypothesis test (with the null being no difference) on the training program described above. Suppose the test is statistically significant (at the 5% significance level) and produces an estimated effect of 9. This effect looks good because it's greater than our smallest meaningful effect size of 5. ***However, this estimate doesn't incorporate the margin of error. The confidence interval (CI) provides that crucial information.***

a). Suppose you are told the 95% confidence interval for the difference between groups in example 1 is [3, 15]. (Recall, you decide that the difference between these two groups must be at least five points to represent a practically meaningful effect size). What does this Confidence Interval (CI) tell us?

b). Suppose instead you are told the 95% confidence interval is [7,11]. What does this Confidence Interval (CI) tell us?

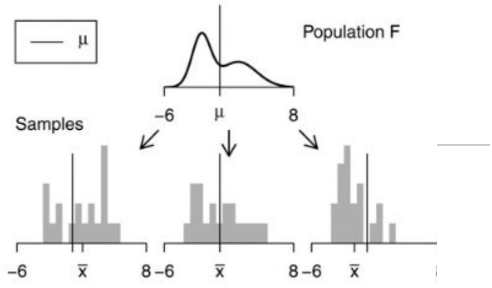
Using the Central Limit Theorem to Calculate Confidence Intervals

Graph of the Distribution

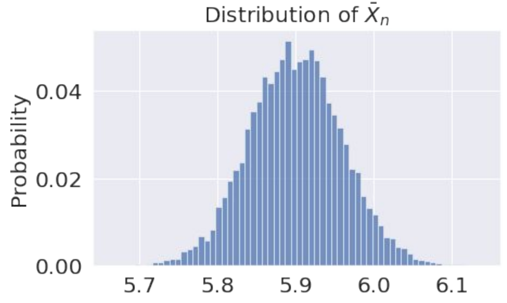


Constructing a CI for the Population Average

- Use sampling distribution of sample average to solve for D, where D is defined as follows:
  - For X% of all samples, if you stand at the **population average** and look a distance of D on both sides, you will find the **sample average**.
- But notice distance is symmetric!
- So if you stand at the ***sample average*** and look a distance of D on both sides, you will capture the **population average** X% of the time.



Ex: Sampling distribution of sample average

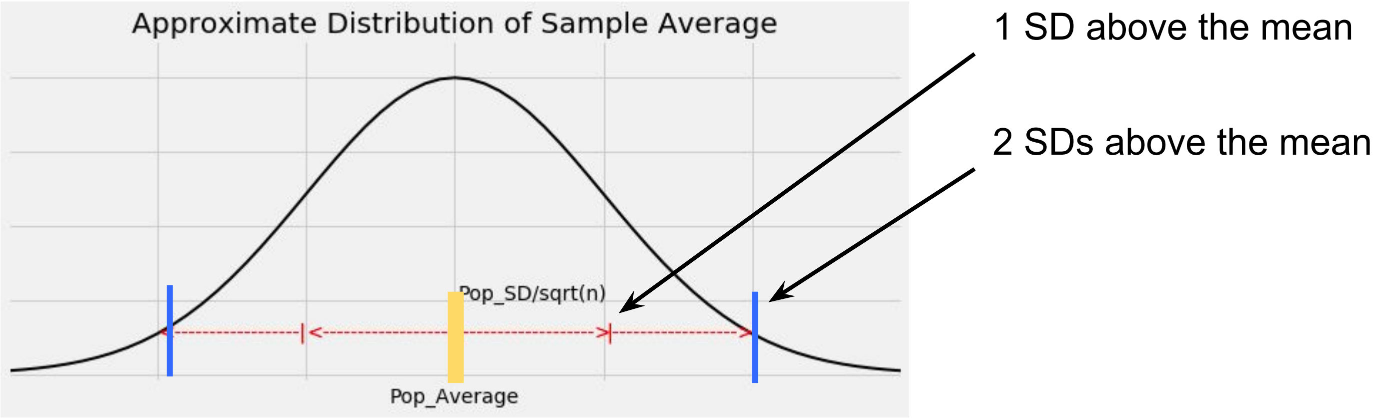


Ex: Suppose pop average = 5.9

You draw one sample with sample ave =5.82

X% Conf Int = [5.82 -D, 5.82+D]

The Key to 95% Confidence



- For about 95% of all samples, the sample average and population average are within **2 SDs** of each other.
- **SD** = SD of sample average

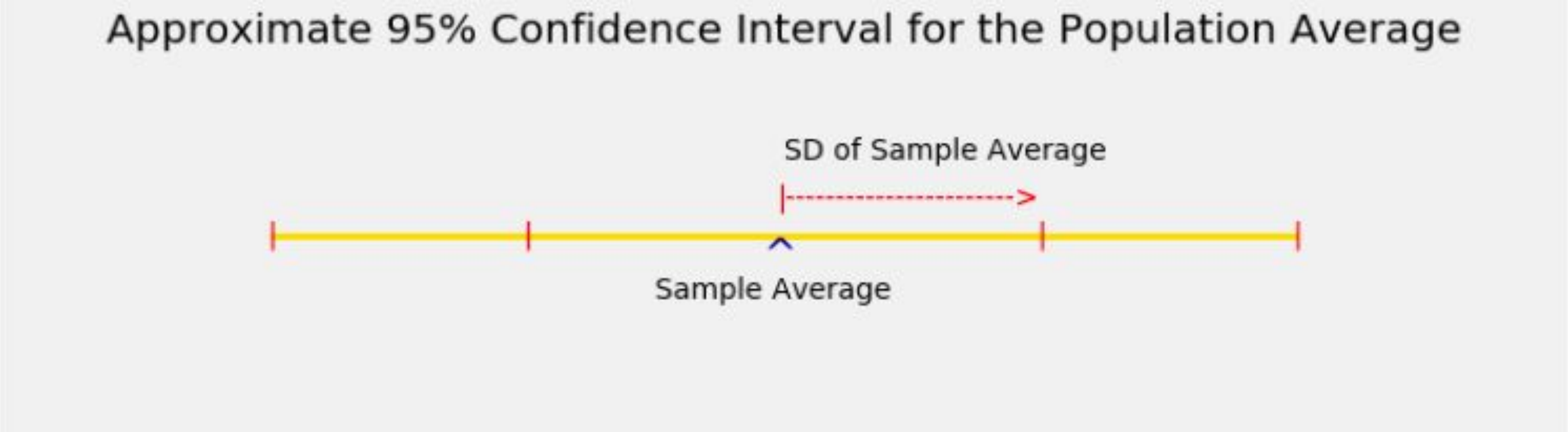
  = (population SD) / $\sqrt{\text{sample size}}$

Constructing the Interval

For 95% of all samples,

- If you stand at the population average and look two **SDs** on both sides, you will find the sample average.
- Distance is symmetric.
- So if you stand at the sample average and look two **SDs** on both sides, you will capture the population average.

The Interval



(Demo)

Summarizing: construction of intervals

- 95% confidence interval for the sample mean
  - Sample_mean +/- 2*SD of the sample mean
- SD of the sample mean
  - (population SD) / $\sqrt{\text{sample size}}$

- But we dont know the population SD
  - We can estimate it using the sample SD
  - Or overestimate it

Question

If we can make 95% confidence interval in this way:

- Sample_mean +/- 2*SD
- Then why do we need to make confidence intervals using bootstraps

This method only works for means and sums (as it is based on CLT) but bootstrap is a much more generalized approach which can work for other statistics like medians as well

Width of the Interval

Total width of a 95% confidence interval for the population average

= 4 * SD of the sample average

= 4 * (population SD) / $\sqrt{\text{sample size}}$

Sample Proportions

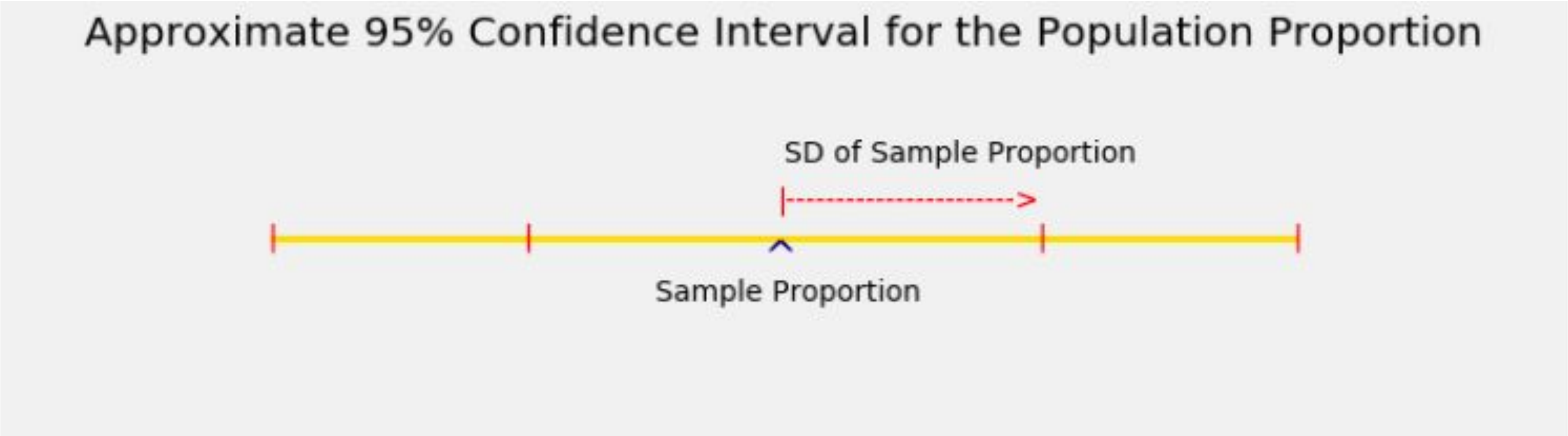
Proportions are Averages

- Data: 0 1 0 0 1 0 1 1 0 0 (10 entries)
- Sum = 4 = number of 1's
- Average = 4/10 = 0.4 = proportion of 1's

If the population consists of 1's and 0's (yes/no answers to a question), then:

- the population average is the proportion of 1's in the population
- the sample average is the proportion of 1's in the sample

Confidence Interval



Controlling the Width

- Total width of an approximate 95% confidence interval for a population proportion

  $= \quad 4 * (\text{SD of 0/1 population}) / \sqrt{\text{sample size}}$

- The narrower the interval, the more precise your estimate.
- Suppose you want the total width of the interval to be no more than 1%. How should you choose the sample size?

The Sample Size for a Given Width

$$0.01 = 4 * (\text{SD of 0/1 population}) / \sqrt{\text{sample size}}$$

- Left side: 1%, the max total width that you'll accept
- Right side: formula for the total width

$$\sqrt{\text{sample size}} = 4 * (\text{SD of 0/1 population}) / 0.01$$

(Demo)

“Worst Case” Population SD

- $\sqrt{\text{sample size}} = 4 * (\text{SD of 0/1 population}) / 0.01$
- SD of 0/1 population is at most 0.5
- $\sqrt{\text{sample size}} \geq 4 * 0.5 / 0.01$
- sample size $\geq$ (4 * 0.5 / 0.01) ** 2 = 40000
- The sample size should be 40,000 or more

Discussion Question

Subscribe

SCIENTIFIC AMERICAN

Cart 0 Sign In | Stay Informed

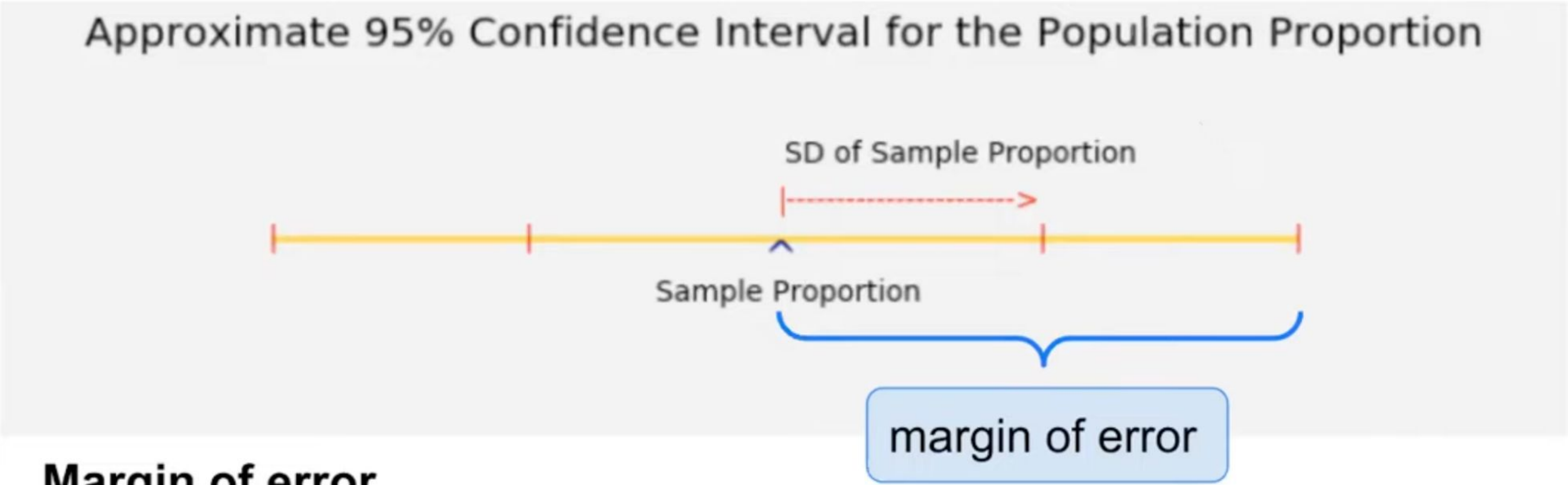
THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS PUBLICATIONS

THE SCIENCES

How can a poll of only 1,004 Americans represent 260 million people with only a 3 percent margin of error?

https://www.scientificamerican.com/article/howcan-a-poll-of-only-100/

Margin of Error in Polls



Margin of error

- Distance from the center to an end
- Half the width of the interval
- 2 * SD of sample proportion

Discussion Question

- 3% margin of error means **width of 6%**

  width = 4 * (0.5) / $\sqrt{1004}$

  width ≈ 0.063, so margin of error ≈ 3.15%

Discussion Question

- A researcher is estimating a population proportion based on a random sample of size 10,000.

Fill in the blank with a decimal:

- With chance at least 95%, the estimate will be correct to within ____________________.

Discussion Question

- With chance at least 95%, the estimate will be correct to within **0.01**.

$$\text{width} = 4 * (0.5) / \sqrt{10000}$$

width = 0.02, so margin of error = 0.01

Discussion Question

- I am going to use a 68% confidence interval to estimate a population proportion.
- I want the total width of my interval to be no more than 2.5%.
- How large must my random sample be?

Discussion Question

- How large must my random sample be?

$$0.025 = \mathbf{2} * (0.5) / \sqrt{\text{sample size}}$$

$$\sqrt{\text{sample size}} = 2 * (0.5) / 0.025$$

sample size = 40**2 = 1600