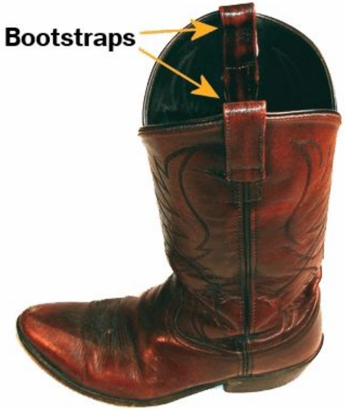
Confidence Intervals and the Bootstrap



LESSON 22

CSCI 3022

Maribeth Oscamou

Content credit: Acknowledgments

Course Logistics: 9th Week At A Glance

Mon 3/11	Tues 3/12	Wed 3/13	Thurs 3/14	Fri 3/15	Sat 3/16
Attend & Participate in Class	(Optional): Attend Notebook Discussion with our TA (5-6pm Zoom)	Attend & Participate in Class	SNOW DAY ~~HW 8 Due 11:59pm~~	SNOW DAY ~~Attend & Participate in Class~~ ~~Quiz 6: L17-L19, TA discussion nb 8, HW 7~~ ~~HW 8 Due 11:59pm~~ HW 9 released	HW 8 Due 11:59pm

Today's Roadmap

CSCI 3022

- Parameter Estimation
- Confidence Intervals (CI)
  - Bootstrapping
  - Interpretation
  - Caveats
- Relationship between CI and Hypothesis Tests

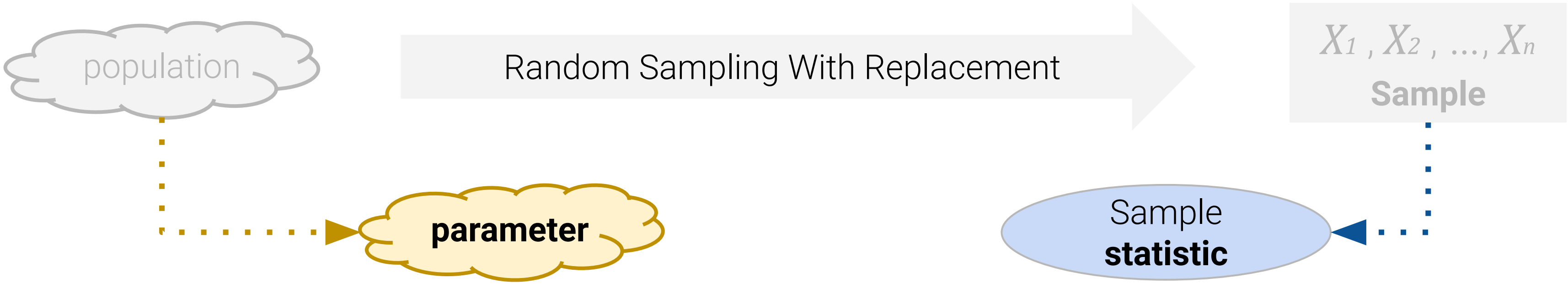
Inference: Parameter Estimation

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**.

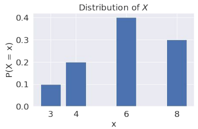
We can then use the sample statistic as an **estimator** of the true population parameter.

- Example: average height of CU undergraduates.
- We typically can collect a **single sample**. It has just one average.
- Since that sample was random, it *could have* come out differently.

Data Generation Process: Estimating a Value

One View: Randomly draw a random sample, then compute the statistic for that sample.

population



$X_1, X_2, \ldots, X_n$

Sample

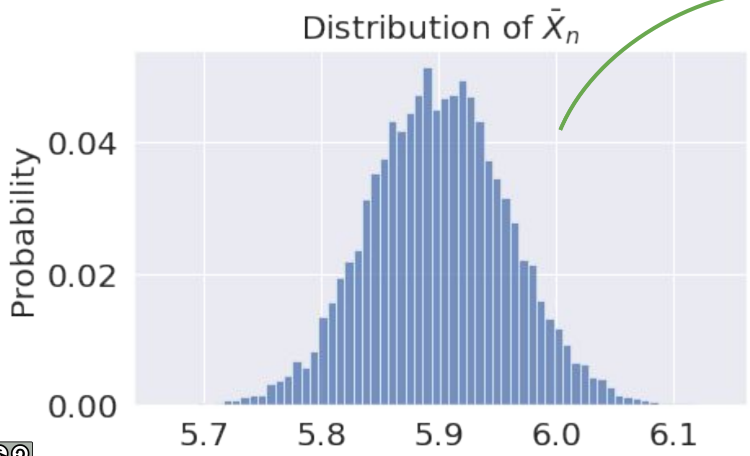
	x
0	6
1	8
2	6
3	6
4	3
...	...
95	8

Sample mean

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

`np.mean(...) = 5.82`

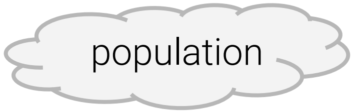
Another View: Randomly draw from the distribution of the statistic (generated from all possible samples).

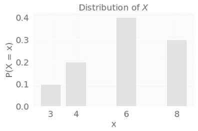


5.82

We often use the 2nd view because we want to interpret the estimator directly, and not the random sample.

If We Drew a Different Sample, We'd Get A Different Estimator

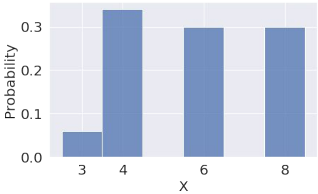




$X_1, X_2, \ldots, X_n$

Sample

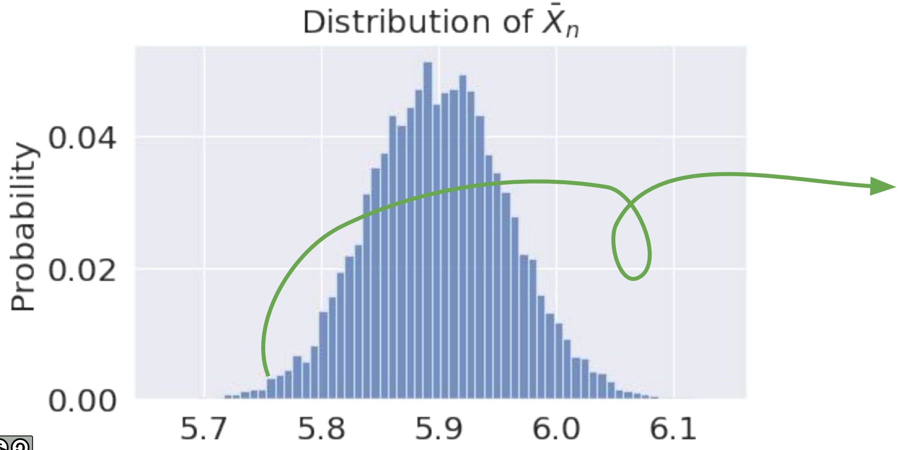
	x
0	6
1	6
2	6
3	4
4	6
...	...
95	8



Sample mean

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

`np.mean(...) = 5.74`



5.74

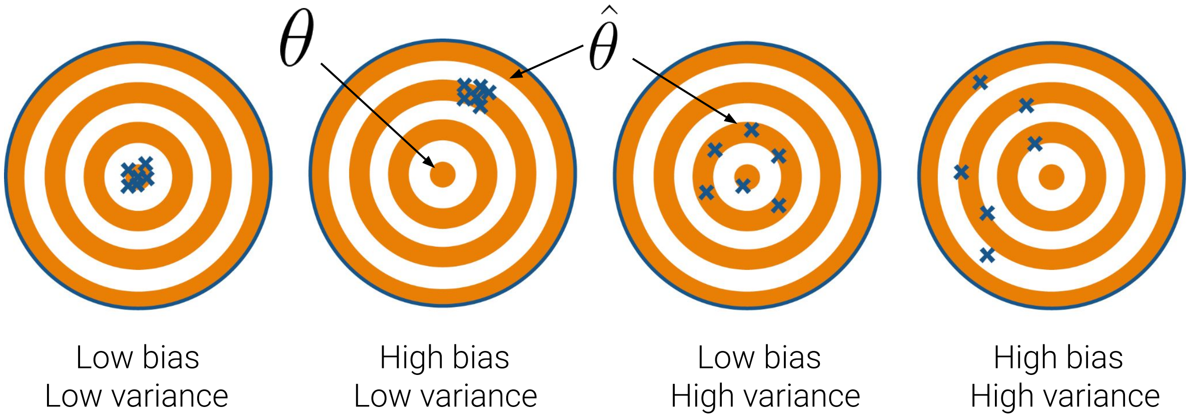
The value of our estimator is a function of the random sample. The estimator is therefore also random.

Performance of an Estimator

Suppose we want to estimate a population parameter θ using an estimator $\hat{\theta}$.

How good is the estimator? Questions we might ask:

- Do we get the right answer on average? **(Bias)**
    - Definition: An estimator is **unbiased** if it produces parameter estimates that are on average correct. Mathematically: E [estimator] = parameter
- How variable is the answer? **(Variance)**



Instead of a single **estimator** we want **an interval around our estimator** ***that incorporates information about the standard error (i.e. standard deviation) of the estimator.***

This is what **Confidence Intervals do!**

Confidence Intervals

CSCI 3022

- Parameter Estimation
- **Confidence Intervals (CI)**
  - Bootstrapping
  - Interpretation
  - Caveats
- Relationship between CI and Hypothesis Tests

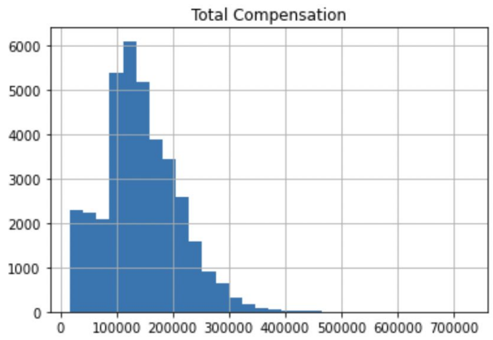
What is a Confidence Interval?

- 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 designated *confidence level (L%)*
    - Can be any percent between 0 and 100
- The **confidence is in the process** that creates the interval:
    - It generates a “good” interval about L% of the time.
    - Unfortunately no way to tell if a particular confidence interval contains the true value of the population parameter.

Demo: Lec22 NB

CSCI 3022

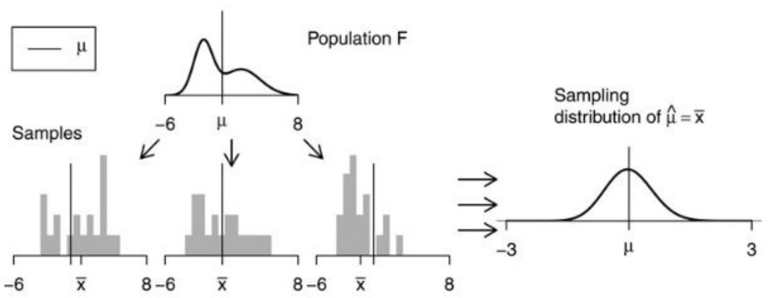


Salary data from city employees in San Francisco

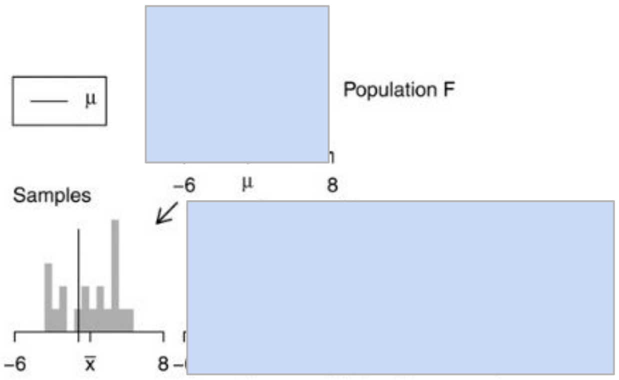
Confidence Intervals: Ideal World vs Reality

- We want to understand **variability** of our **estimate.**
  - **Need sampling distribution of sample statistic to do this.**
- Given the population, we could simulate:
- But we don't know the population, we only have ONE sample!
  - To get many values of the estimate, we needed **many random samples**
  - Can't go back and sample repeatedly from the population:
    - No time, no money
- Are we stuck?

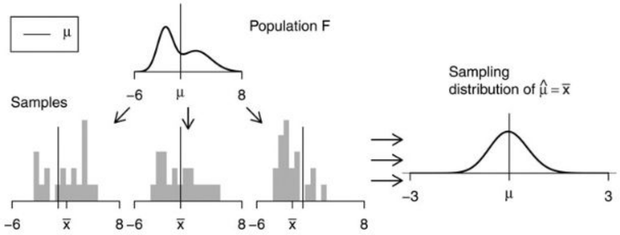
"Ideal world":



"Reality":

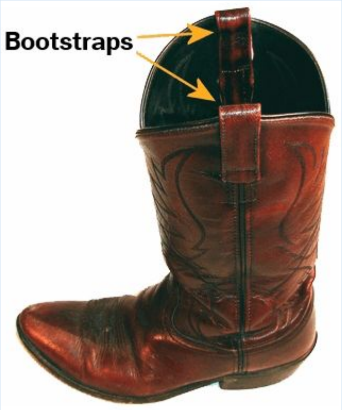


Methods for Calculating Confidence Intervals



We need a technique for estimating the sampling distribution of an estimator.

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



"Pull Yourself Up By Your Bootstraps"

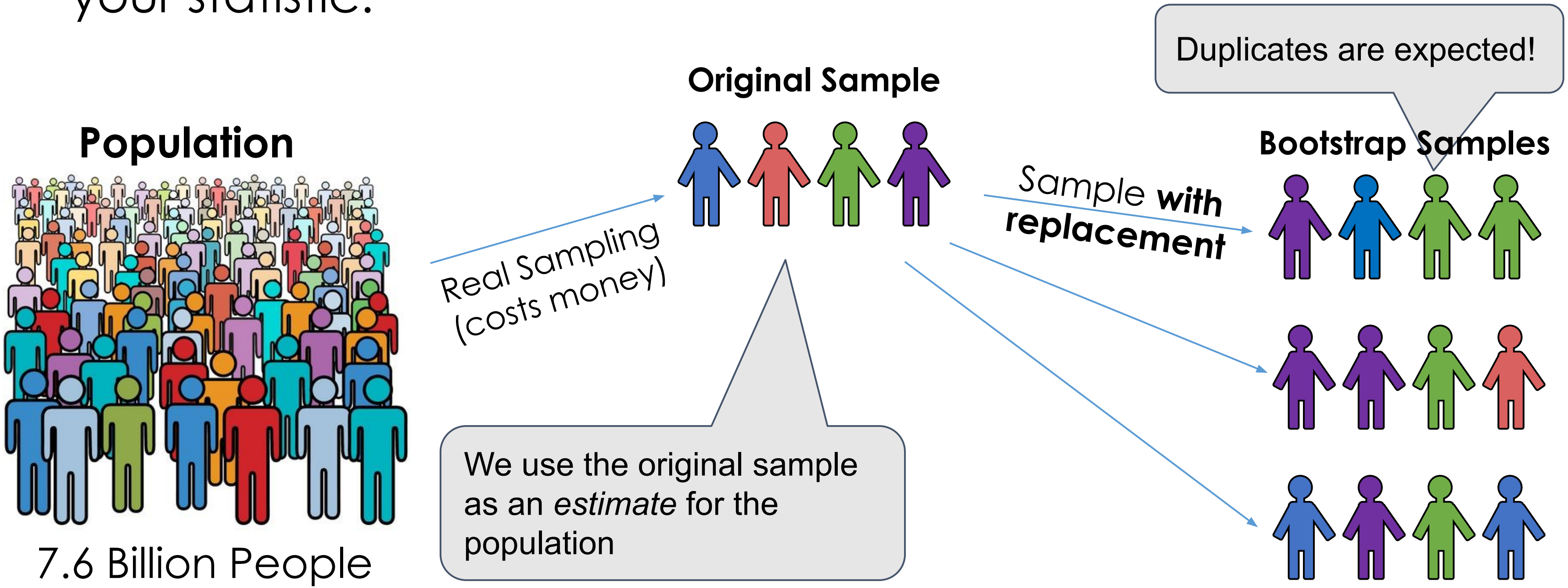
Bootstrapping

CSCI 3022

- Parameter Estimation
- Confidence Intervals (CI)
  - **Bootstrapping:**
  - Interpretation
  - Caveats
- Relationship between CI and Hypothesis Tests

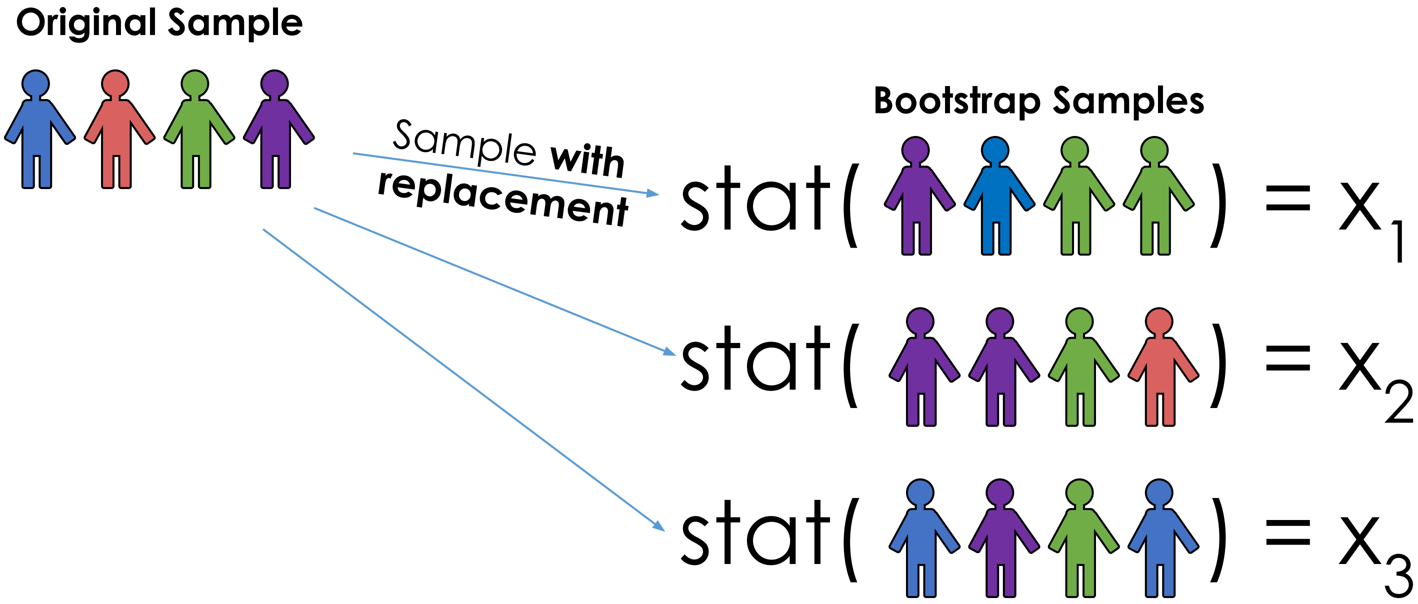
Bootstrap the Distribution of a Statistic

Simulation method to estimate the sample distribution of your statistic.



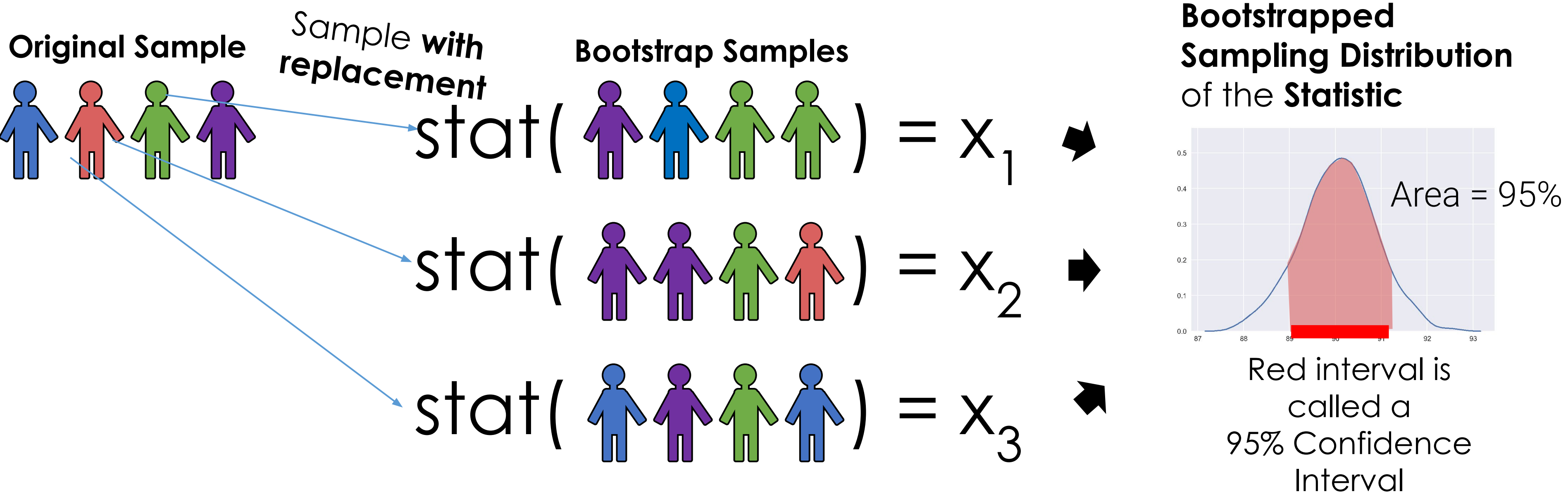
Bootstrap the Distribution of a Statistic

Simulation method to estimate the sample distribution of your statistic.



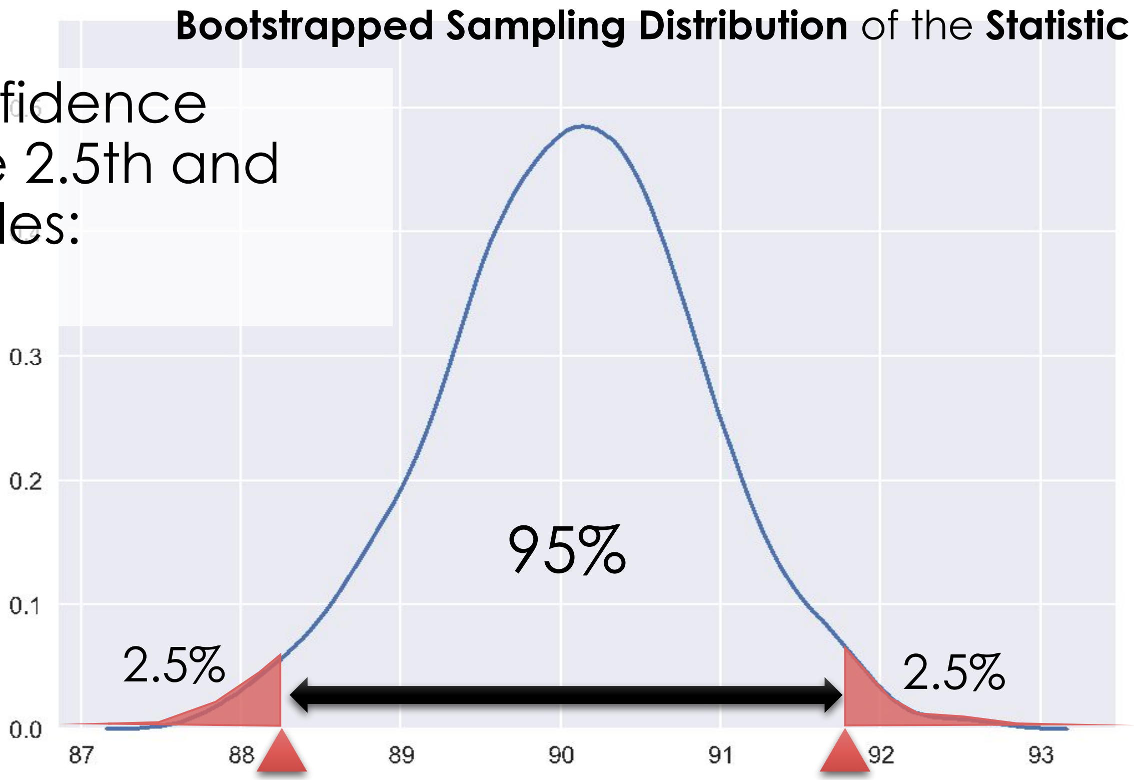
Bootstrap the Sampling Distribution of a Statistic

Simulation method to estimate the sample distribution of your statistic.



Bootstrap Percentile Method for Calculating a Confidence Interval

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



NB 22 DEMO

Lesson 22 Part 2: Remote, Asynchronous, Make-Up Lesson for Snow Day 3/15

This content IS in the scope of HW, quizzes and exams in this class

Per the CU Boulder policy:

colorado.edu/academicaffairs/sites/default/files/attached-files/academic_instruction_guidance_edition_4.2_final_01.12.24.pdf

Academic Instruction Guidance Edition 4.2 FINAL 01.12.24.docx   10 / 17   100%

Do's and don'ts for instruction modes and makeup class sessions in cases of emergency

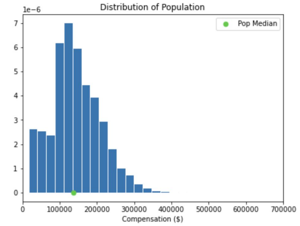
If you cannot hold your in-person class on campus due to a weather or non-weather emergency, you may not simply change your class to remote instruction mode and require your students to attend. Students may not receive notice in time, and/or they may not be able to access the class remotely because they do not have access to the necessary technology.

If your class is canceled due to a weather or non-weather emergency, you may schedule an in-person or remote makeup session, but you may not require that students attend that session if it is not during your regular class time. However, you may also record a makeup session for students to access on their own schedule. If your makeup session covers material that will be on an exam, you should record the lecture and make it very clear to your students (put it in writing) that they are responsible for that material.

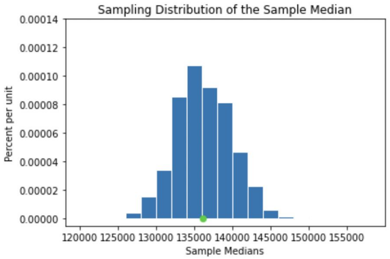
Bootstrapped Percentile Method for Calculating a Confidence Interval

What we wish we could see

population

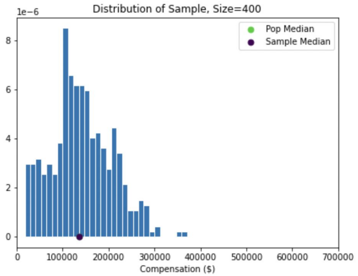


Sampling distribution of statistic

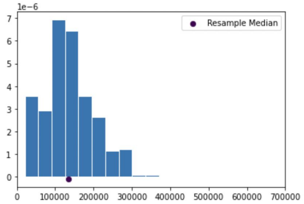


What we get to see

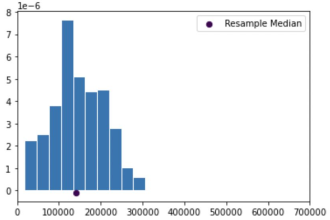
sample

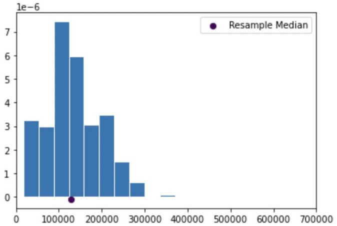


Bootstrapped Resampling

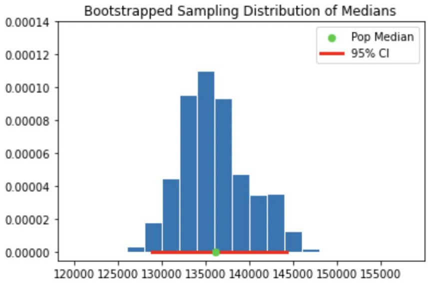



```
resample=orig_sample.sample(frac=1,replace=True)
```



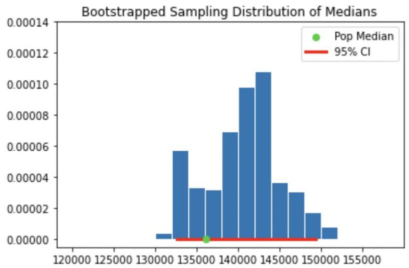


Bootstrapped Sampling distribution of statistic



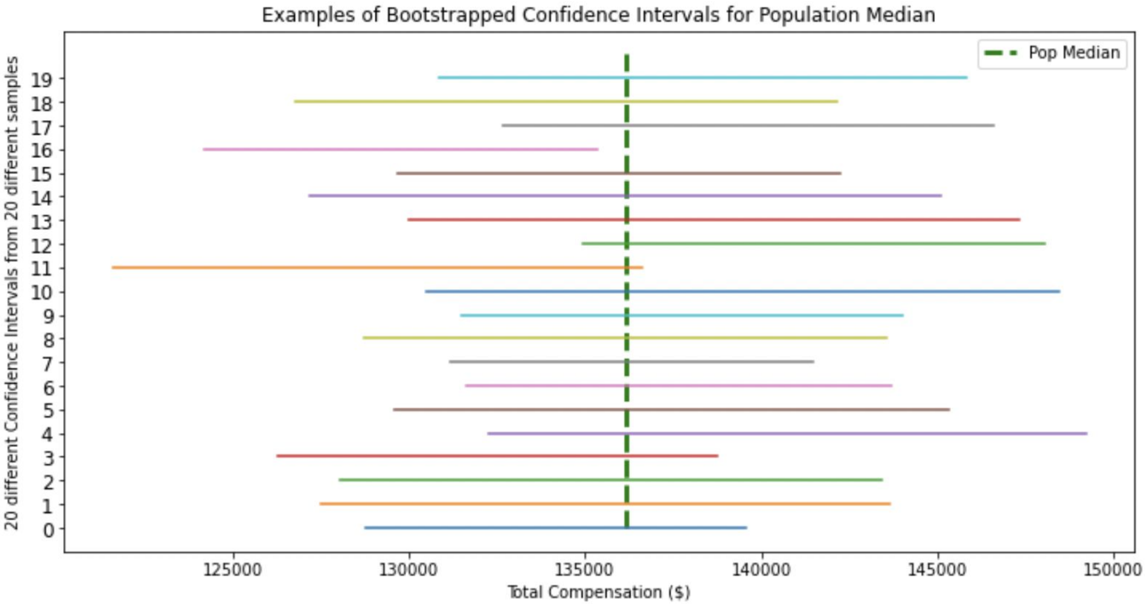
Construct the red 95% confidence interval by taking the 2.5% and (100 - 2.5)% percentiles

Meaning of 95% Confidence Interval



The 95% Confidence is in the **PROCESS used to create the interval, not** in any one particular interval.

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



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

There are **20 intervals**.
We expect **roughly 95% of them (19)** to contain the parameter.

True or False

Suppose we take a single sample of size 400 and construct a 95% Confidence Interval for the population median and then a 90% CI for the population median.

True or False:

The 95% CI will be narrower than the 90% CI.

A). True

B). False

Changing the Confidence Level:

Fill in the code below to bootstrap a 90% CI for the population median using the percentile method. Assume that the sample data is in an np.array called `original_sample`

```
new=[ ]
nsamp=10000

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

CI =np.percentile(          )
```

Changing the Confidence Level:

Fill in the code below to bootstrap a 90% CI for the population median using the percentile method. Assume that the sample data is in an np.array called `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])
```

Summary in Words: Creating Confidence Intervals Using Bootstrapping and Percentiles

- 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 variability estimates are comparable
- For each sample, compute the statistic you are using to estimate the parameter
- Repeat process above 10,000 times (to ensure empirical sampling distribution is close to the theoretical sampling distribution of the statistic)
- Calculate distribution of bootstrapped sample statistics
- Construct a L% confidence interval by taking the **(100-L)/2 and 100 - (100-L)/2** percentiles

Recap: What does a L% Bootstrapped Percentile 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 **L% 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…*

Ex 2:

Given the birth data from our previous lessons, use bootstrapping to calculate a 95% CI for the **average age of the mothers in the population.**

What we wish we could see

population

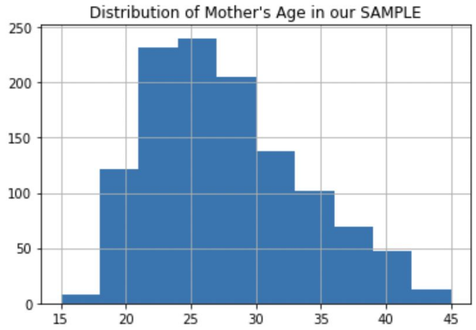
??

Sampling distribution of sample mean

??

What we get to see

SAMPLE



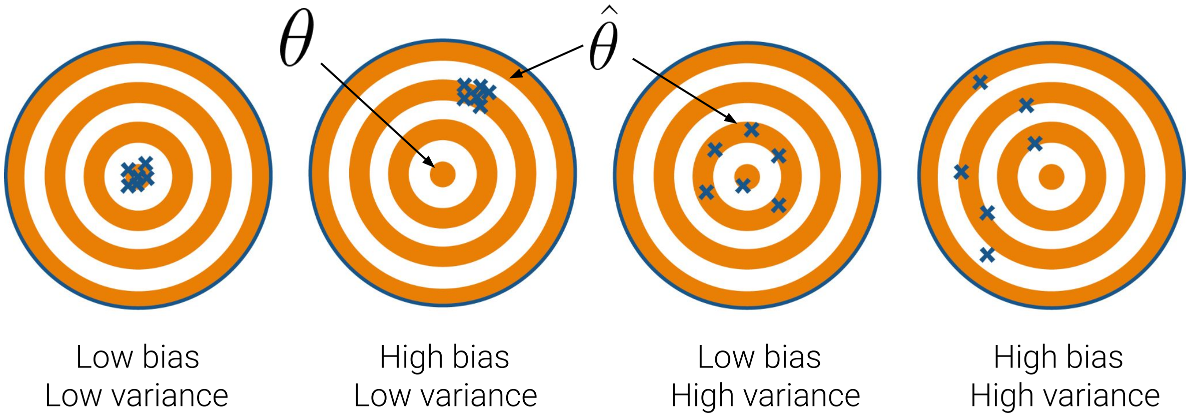
DEMO

Revisiting: Performance of an Estimator

Suppose we want to estimate a population parameter θ using an estimator $\hat{\theta}$.

How good is the estimator? Questions we might ask:

- Do we get the right answer on average? **(Bias)**
  - Definition: An estimator is **unbiased** if it produces parameter estimates that are on average correct. Mathematically: E [estimator] = parameter
- How variable is the answer? **(Variance)**



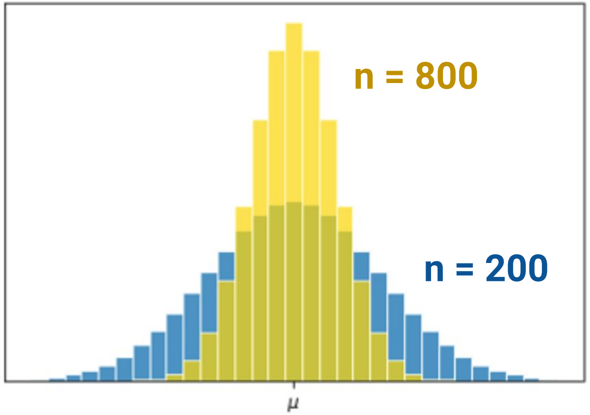
Instead of a single **estimator** we want **an interval around our estimator** ***that incorporates information about the standard error (i.e. standard deviation) of the estimator.***

This is what **Confidence Intervals do!**

Using the Sample Mean to Estimate the Population Mean

Our goal is often to **estimate** some characteristic of a population.

- Example: average height of undergraduates.
- We typically can collect a **single sample**. It has just one average.
- Since that sample was random, it *could have* come out differently.



We should consider the **average value and spread** of all possible sample means, and how it scales with the sample size n.

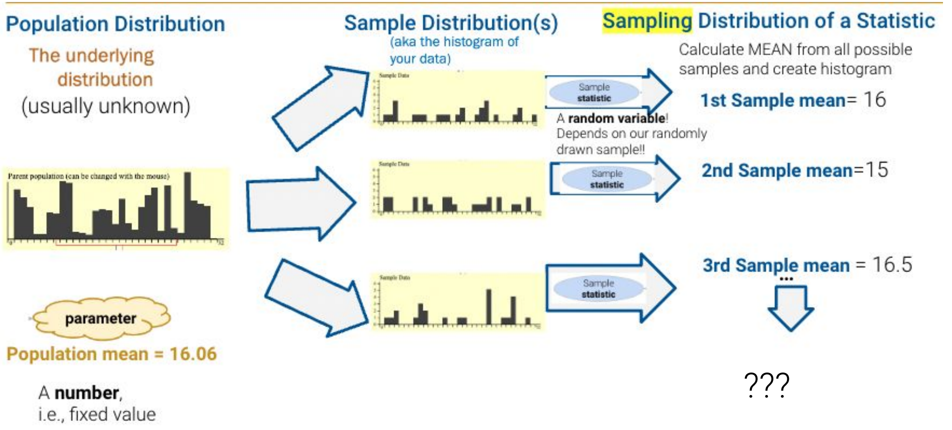
$$\mathbb{E}[\bar{X}_n] = \qquad\qquad \mathrm{SD}(\bar{X}_n) =$$

Recall:Sampling Distribution of the SAMPLE MEAN

Consider an IID sample $X_1, X_2, \ldots, X_n$ drawn from a numerical population with **mean μ and SD σ**.

Define the sample mean: $$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Thus the sample mean is a RANDOM VARIABLE



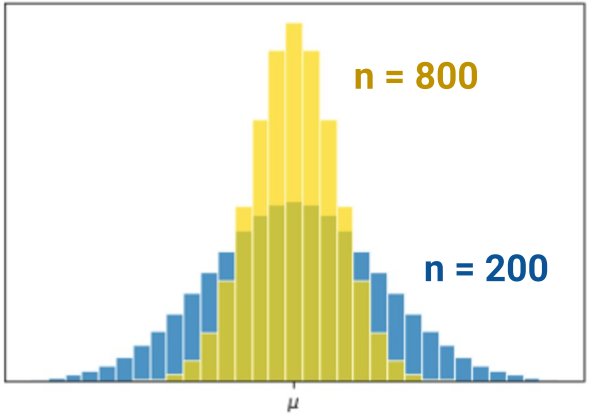
1.
2. What is the expected value of the sample mean? $$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}(n\mu) = \boxed{\mu}$$
3. What is the variance of the sample mean? $$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\underbrace{\left(\sum_{i=1}^{n}\mathrm{Var}(X_i)\right)}_{\text{IID} \rightarrow \mathrm{Cov}(X_i, X_j) = 0} = \frac{1}{n^2}\left(n\sigma^2\right) = \boxed{\frac{\sigma^2}{n}}$$
4. What is the **standard error** of the sample mean? $$\mathrm{SD}(\bar{X}_n) = \boxed{\frac{\sigma}{\sqrt{n}}}$$

Definition: The **standard error** of a statistic is the standard deviation of the sampling distribution of that statistic.

Using the Sample Mean to Estimate the Population Mean

Our goal is often to **estimate** some characteristic of a population.

- Example: average height of undergraduates.
- We typically can collect a **single sample**. It has just one average.
- Since that sample was random, it *could have* come out differently.



We should consider the **average value and spread** of all possible sample means, and what this means for how big n should be.

$$\mathbb{E}[\bar{X}_n] = \mu \qquad \mathrm{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

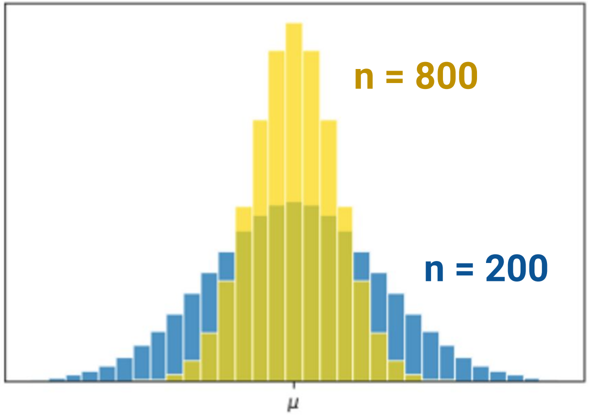
For every sample size, the expected value of the sample mean is the population mean.

We call the **sample mean** an **unbiased estimator** of the population mean.

Using the Sample Mean to Estimate the Population Mean

Our goal is often to **estimate** some characteristic of a population.

- Example: average height of Cal undergraduates.
- We typically can collect a **single sample**. It has just one average.
- Since that sample was random, it *could have* come out differently.



We should consider the **average value and spread** of all possible sample means, and what this means for how big n should be.

$$\mathbb{E}[\bar{X}_n] = \mu$$

For every sample size, the expected value of the sample mean is the population mean.

We call the sample mean an **unbiased estimator** of the population mean.

$$\mathrm{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

(Also called Standard Error)

If you increase the sample size by a factor, the SD decreases by the square root of the factor.

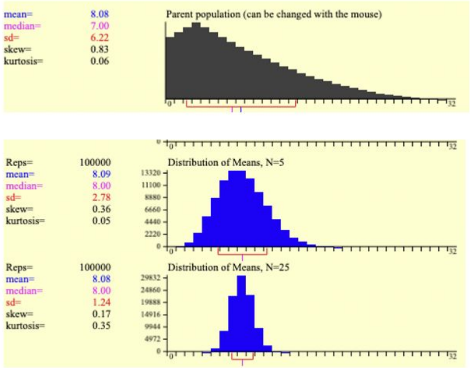
The sample mean is more likely to be close to the population mean if we have a larger sample size.

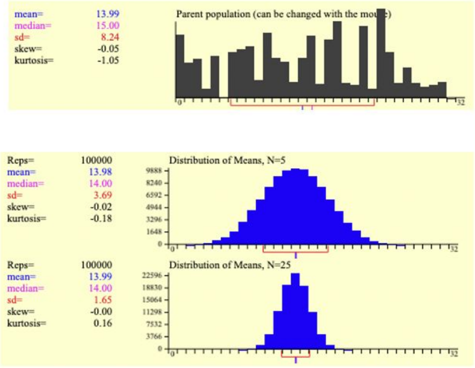
Review: Unbiased Estimators

For every sample size, the expected value of the sample mean is the population mean. $$\mathbb{E}[\bar{X}_n] = \mu$$

We call the sample mean an
unbiased estimator of the population mean.

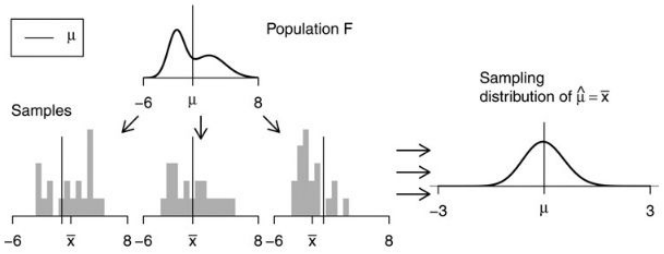
An **unbiased estimator** means that if we were to repeat this sampling process many times, the expected value of our estimates should be equal to the true values we are trying to estimate.





Recap: Confidence Intervals: Ideal World vs Bootstrap World

- We want to understand **variability** of our **estimate.**
  - Need sampling distribution of sample statistic to do this.
- Given the population, we could simulate:



Reality: We don't know the population distribution.

Method 1: Bootstrapping

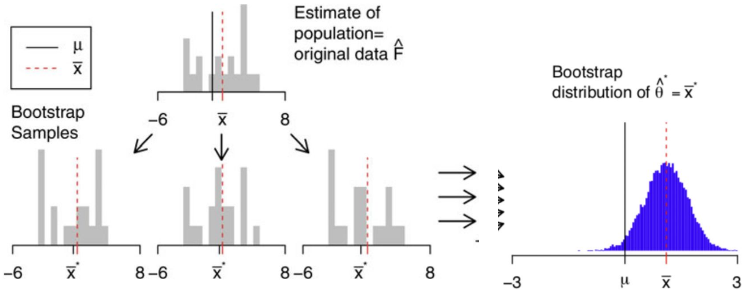
- Treat our random sample as a "population", and resample from it **with replacement** computing the statistic of interest for each resample
- Create distribution of bootstrapped statistics
- Use the middle X% of this distribution to calculate the X% Confidence Interval for the population parameter

Note: The ***bootstrapped distribution is NOT centered at*** the actual population parameter.

Method 2: Use Central Limit Theorem (if it applies)

Bootstrap World:

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



Interpretation

CSCI 3022

- Parameter Estimation
- Confidence Intervals (CI)
    - Bootstrapping
    - **Interpretation**
    - Caveats
- Relationship between CI and Hypothesis Tests

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.

Caveats

CSCI 3022

- Parameter Estimation
- **Confidence Intervals (CI)**
    - Bootstrapping
    - Interpretation
    - **Caveats**
- Relationship between CI and Hypothesis Tests

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.*

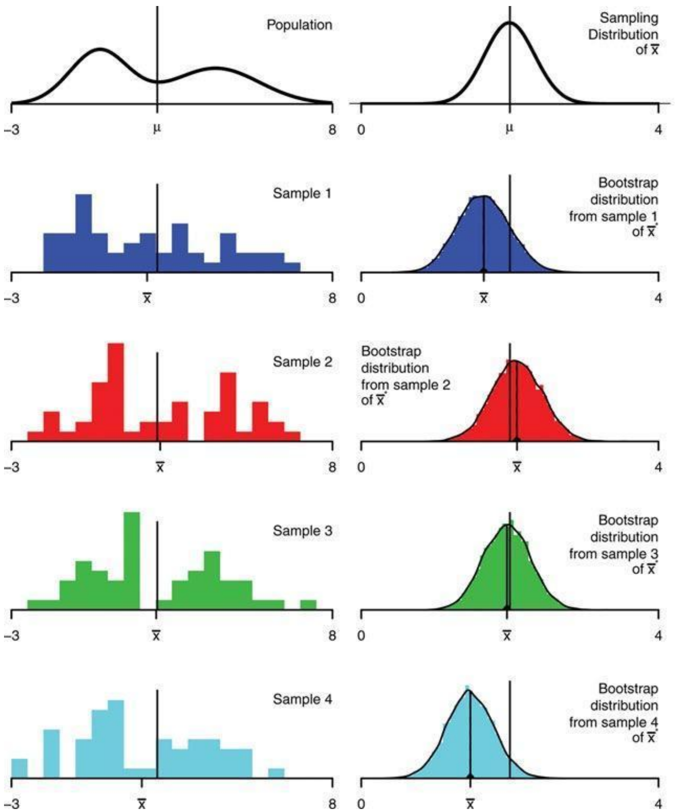
- The bootstrap does not replace or add to the original data. We use the bootstrap distribution as a way to estimate the variation in a statistic based on the original data.

- **BOOTSTRAP DISTRIBUTIONS AND SAMPLING DISTRIBUTIONS**

  For most statistics, bootstrap distributions approximate the shape, spread, and bias of the actual sampling distribution.

  Bootstrap distributions differ from the actual sampling distributions in the location of their centers.
  The bootstrapped distribution is centered at the original statistic value (plus any bias) rather than the parameter value (plus any bias).

Bootstrap for the Mean
Sample size=50

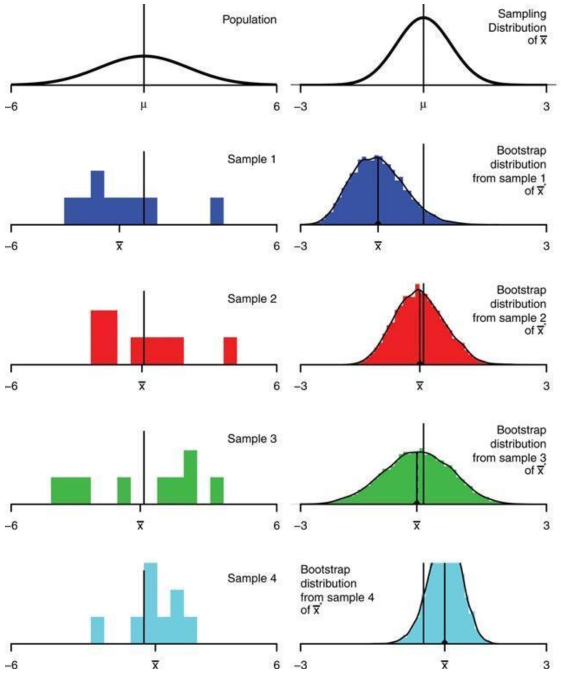


When *Not* to Use Our Bootstrap Method

- The bootstrap percentile method works well for estimating the population median or mean based on a large random sample. However, it has limitations, as do all methods of estimation. For example, it is ***not expected to do well*** in the following situations.
  - The goal is to estimate the minimum or maximum value in the population, or a very low or very high percentile, or parameters that are greatly influenced by rare elements of the population.
  - The probability distribution of the statistic is not roughly bell shaped.
  - The original sample is very small, say less than 10 or 15.

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

Bootstrap for mean; 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

CSCI 3022

- Parameter Estimation
- Confidence Intervals (CI)
  - Bootstrapping
  - Interpretation
  - Caveats
- **Relationship between CI and Hypothesis Tests**

Using a CI for Hypothesis Testing

- Confidence intervals focus on the **size of an effect** and the **uncertainty around the estimate** rather than just whether an observed test statistic is statistically significant.
- We can use a CI for testing the following 2 types of hypothesis tests:

- **Null hypothesis**: Population average = x
- **Alternative hypothesis**: Population average ≠ x
- Significance Level (i.e. Cutoff for **p-value**): a%
- Method:
  - Construct a (100-a)% confidence interval for the population average
  - If x is not in the interval, reject the null
  - If x is in the interval, fail to reject the null

- **Null hypothesis**: Population proportion = p
- **Alternative hypothesis**: Population proportion ≠ p
- Significance Level (i.e. Cutoff for **p-value**): a%
- Method:
  - Construct a (100-a)% confidence interval for the population proportion
  - If p is not in the interval, reject the null
  - If p is in the interval, fail to reject the null

Ex 2 cont'd: Using a Confidence Interval to test a Hypothesis

Consider the following hypothesis test:

Null: The average age of mothers in the population is 25 years; the random sample average is different due to chance.

Alternative: The average age of the mothers in the population is **not** 25 years.

Suppose you use the 5% cutoff for the p-value.

In the previous example we found the 95% confidence interval for the average age of mothers in the population is (26.9, 27.6).
Based on this information, what should you conclude for your hypothesis test and why?

Recap: What does a L% 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 **L% 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…*

Appendix

CSCI 3022

Using Bootstrapping to Check for Bias

BIAS

A statistic used to estimate a parameter is **biased** when its sampling distribution is not centered at the true value of the parameter. The bias of a statistic is the mean of the sampling distribution minus the parameter.

The bootstrap method allows us to check for bias by seeing whether the bootstrap distribution of a statistic is centered at the statistic of the original random sample. The bootstrap estimate of bias is the mean of the bootstrap distribution minus the statistic for the original data.