



190F
Fall 2018

Foundations of Data Science

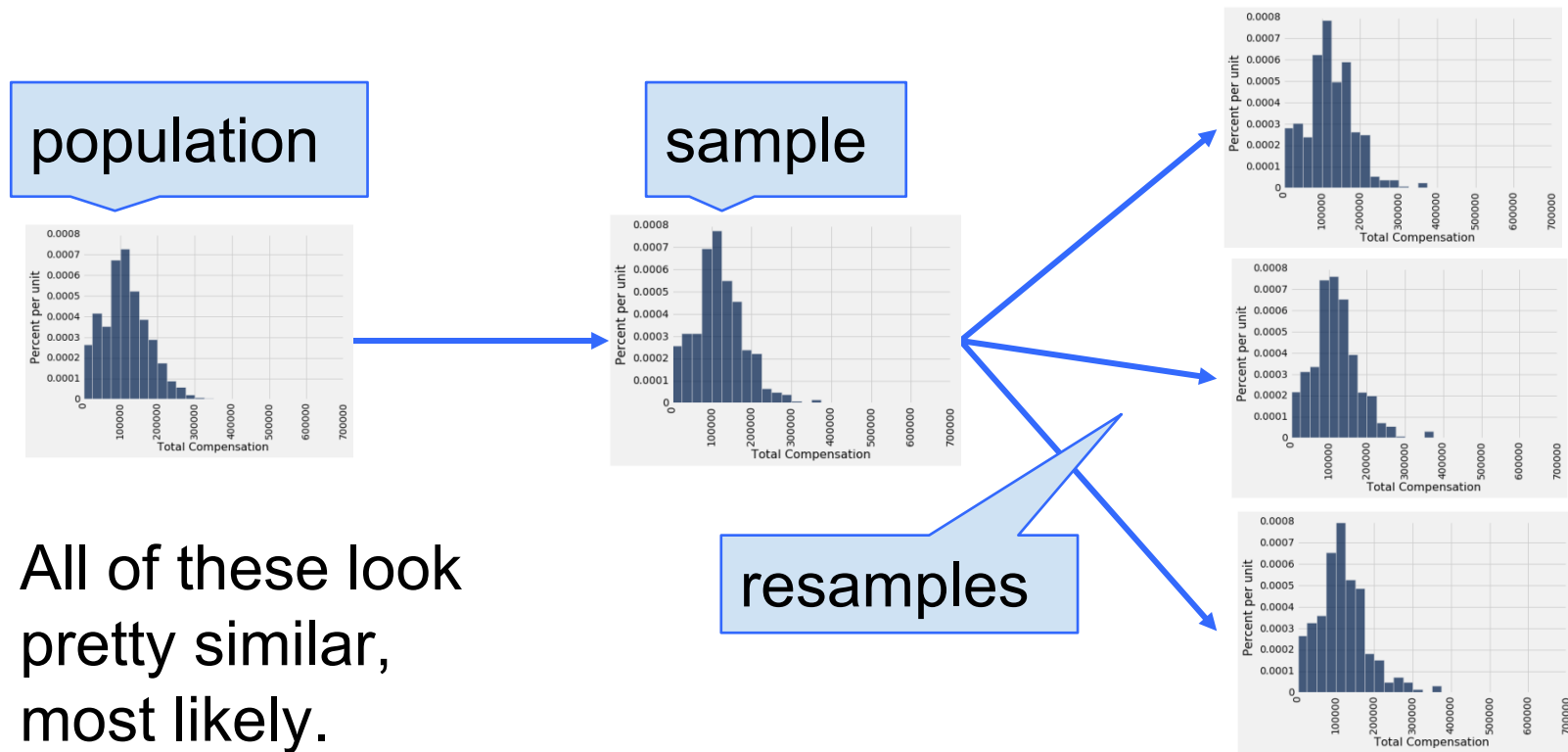
Lecture 20

Interpreting Confidence

Announcements

The Bootstrap

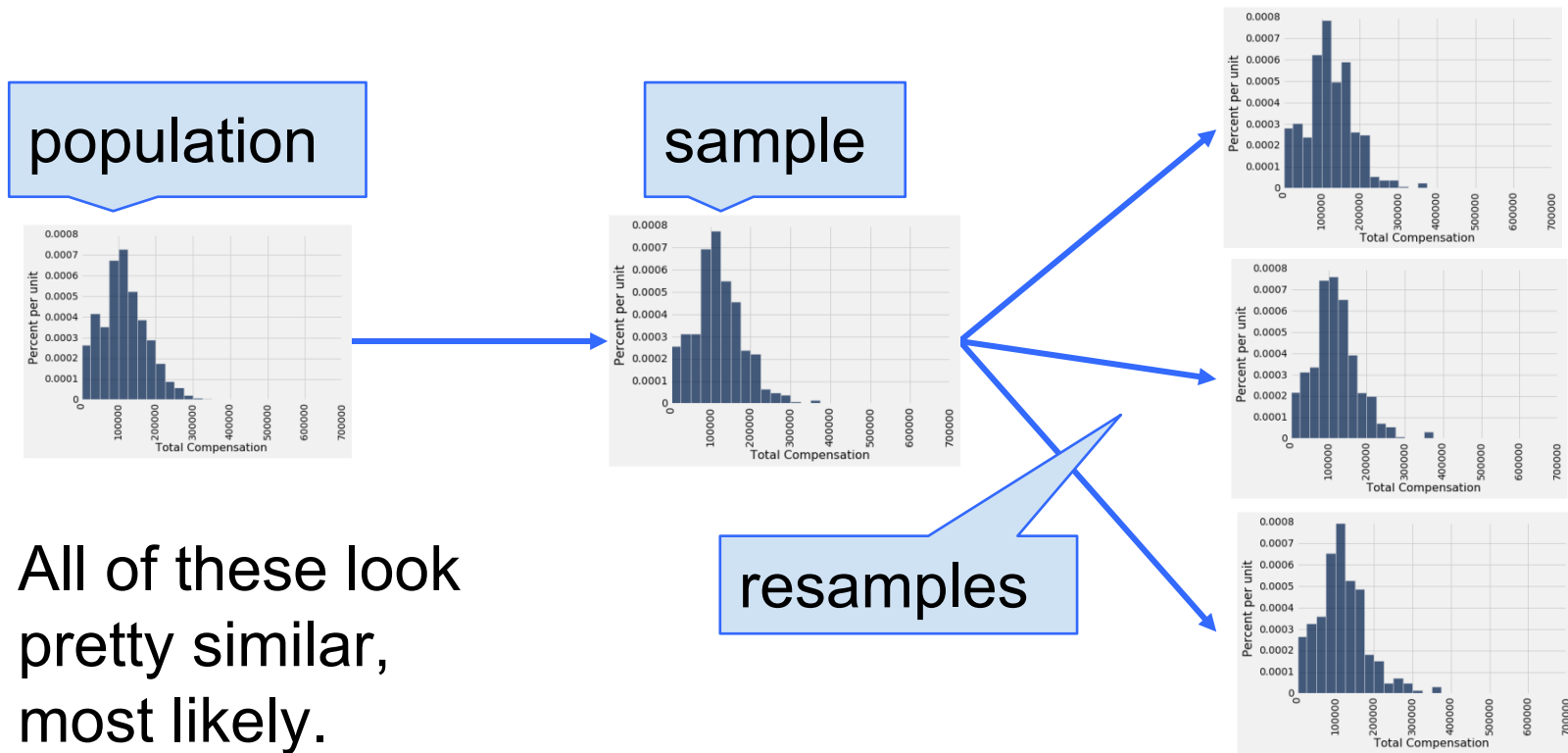
Why the Bootstrap Works



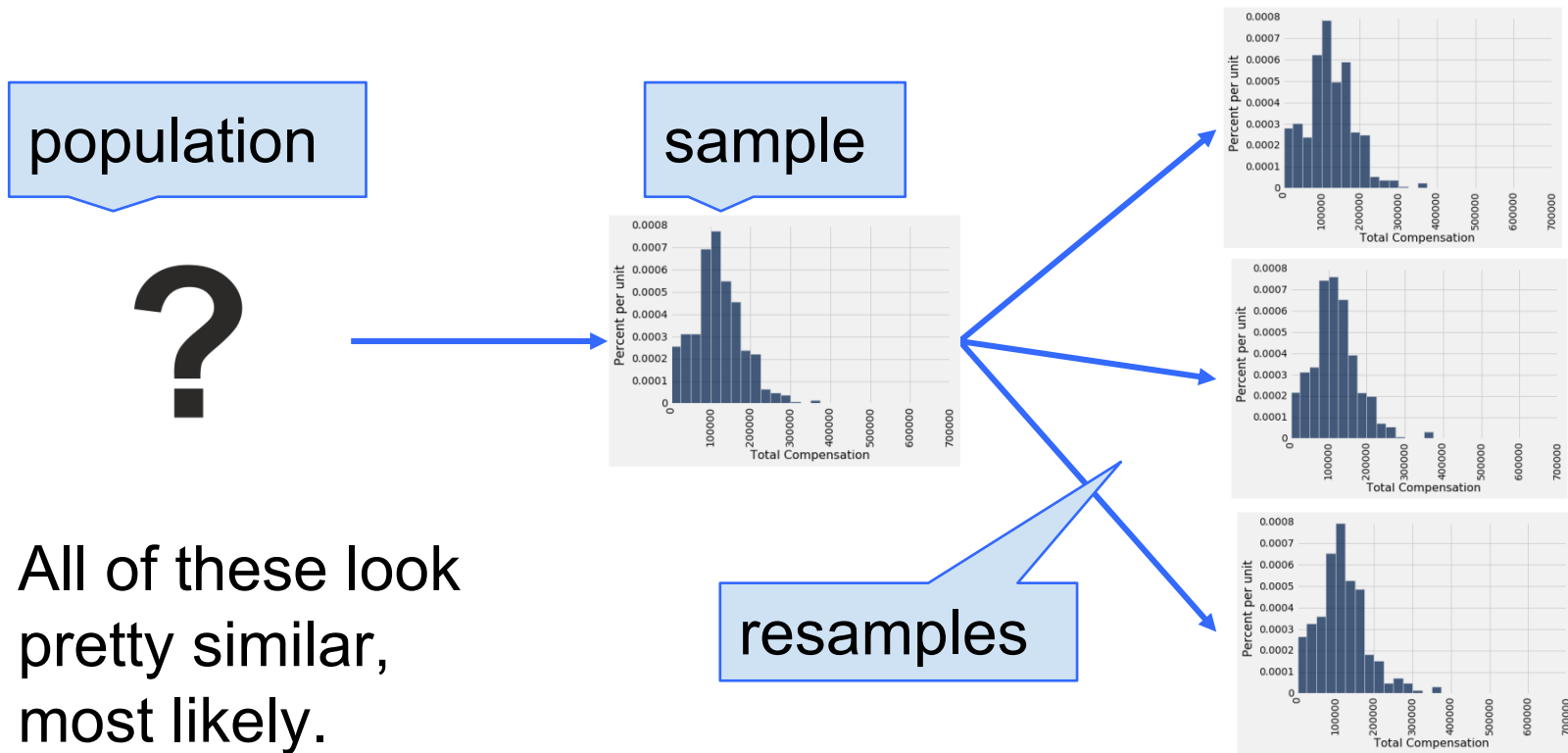
Key to Resampling

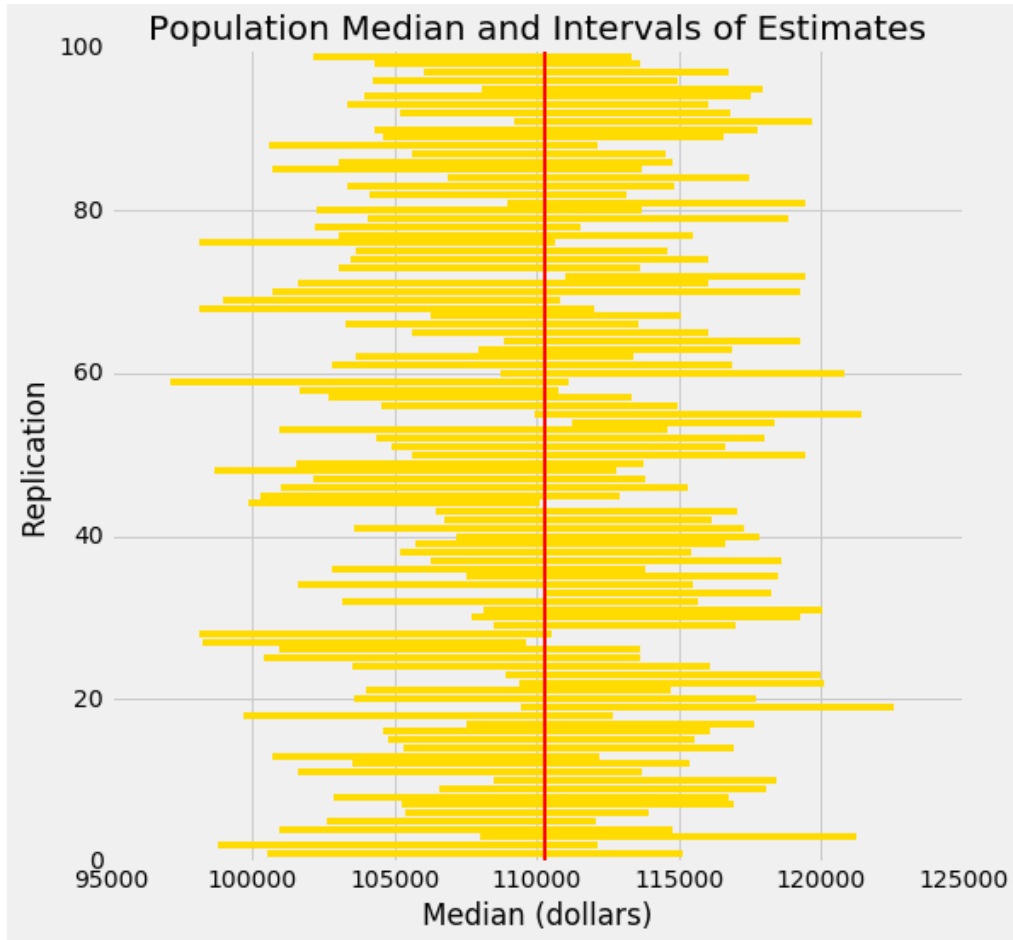
- From the original sample,
 - draw at random
 - with replacement
 - as many values as the original sample contained
 - The size of the new sample has to be the same as the original one, so that the two estimates are comparable
-

Why the Bootstrap Works



Inference Using the Bootstrap





Each line here is a confidence interval from a fresh sample from the population

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 means wider intervals
- The **confidence is in the process** that generated the interval:
 - It generates a “good” interval about 95% of the time.

(Demo)

Use Methods Appropriately

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 95% of the time.

(Demo)

Is This What a CI Means?

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:

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

Answer: False. It's not a probability. It's either true or false that the average age of mothers is in the range 26.9..27.6

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

(Demo)

Confidence Interval Tests

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 means wider intervals
- The **confidence is in the process** that generated the interval:
 - It generates a “good” interval about 95% of the time.

(Demo)

Using a CI for Testing

- Null hypothesis: **Population mean = x**
 - Alternative hypothesis: **Population mean $\neq x$**
 - 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
-