Confidence Intervals: Bootstrap Distribution

Quan Zhang, Ph.D.

Slides adopted from Dr. Kari Lock Morgan

Common Misinterpretations

- "A 95% confidence interval contains 95% of the data in the population"
- "I am 95% sure that the mean of a sample will fall within a 95% confidence interval for the mean"
- "The probability that the population parameter is in this particular 95% confidence interval is 0.95"

Confidence interval for frequentists

- In frequentist (non-Bayesian) statistics, a parameter is an unknown constant value without any randomness

- The correct interpretation of a 95% confidence interval of some parameter:
  You can be 95% certain that this range of values contains the true parameter

Standar error

The ***standard error*** of a statistic, SE, is an estimated standard deviation of an estimator (a parameter estimation).

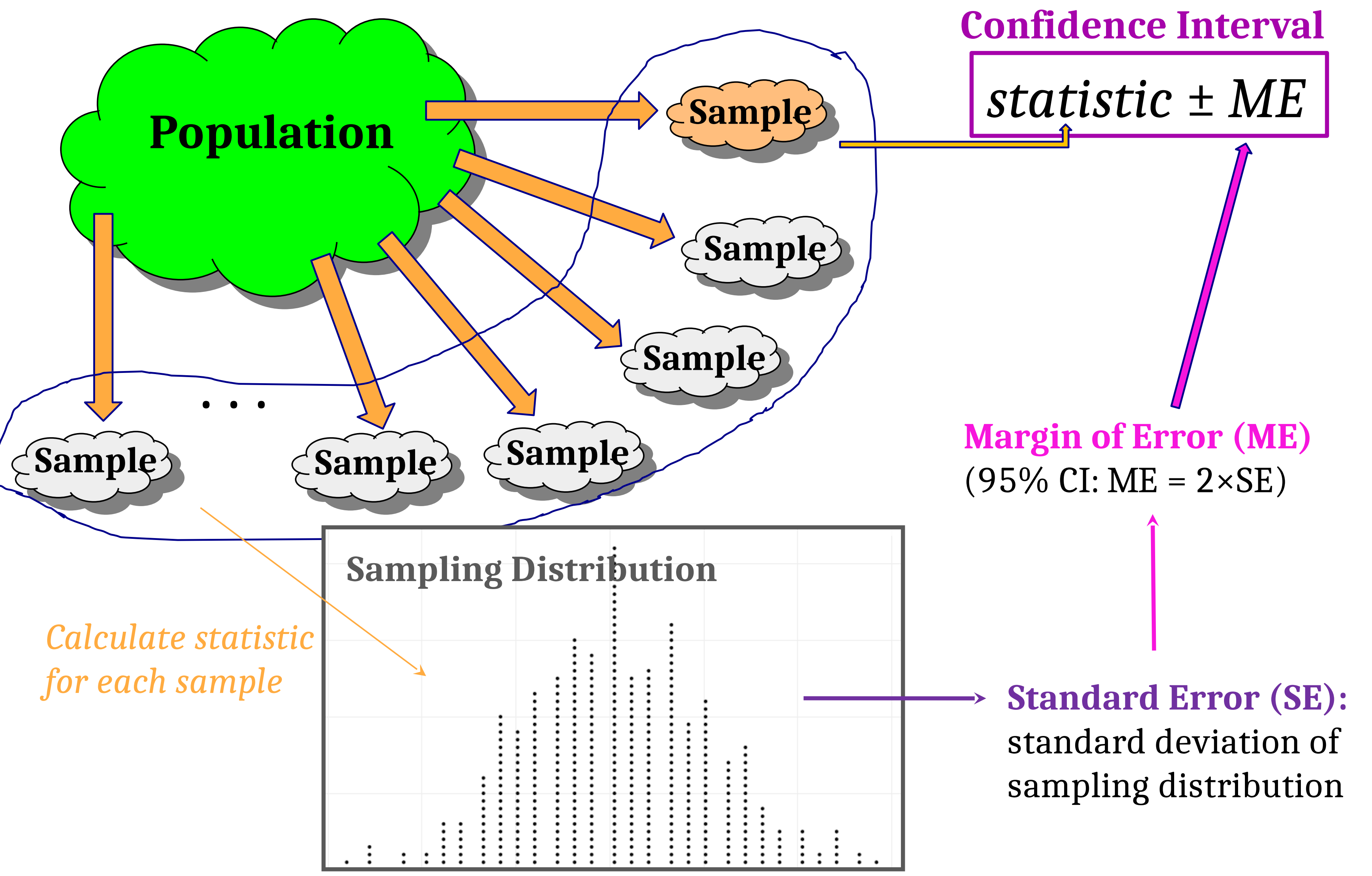
The parameter is estimated by some sample statistic. So it is also a random variable

Point estimations

Population with unknown paramters ->
A random sample ->
Parameter estimation by a function of the sampled data
(e.g., adult height)

One point estimation from one sample

Confidence Intervals



Summary

- To create a plausible range of values for a parameter:
  - Take many random samples from the population, and compute the sample statistic for each sample
  - Compute the standard error as the standard deviation of all these statistics
  - Use statistic ± 2×SE

- One small problem…

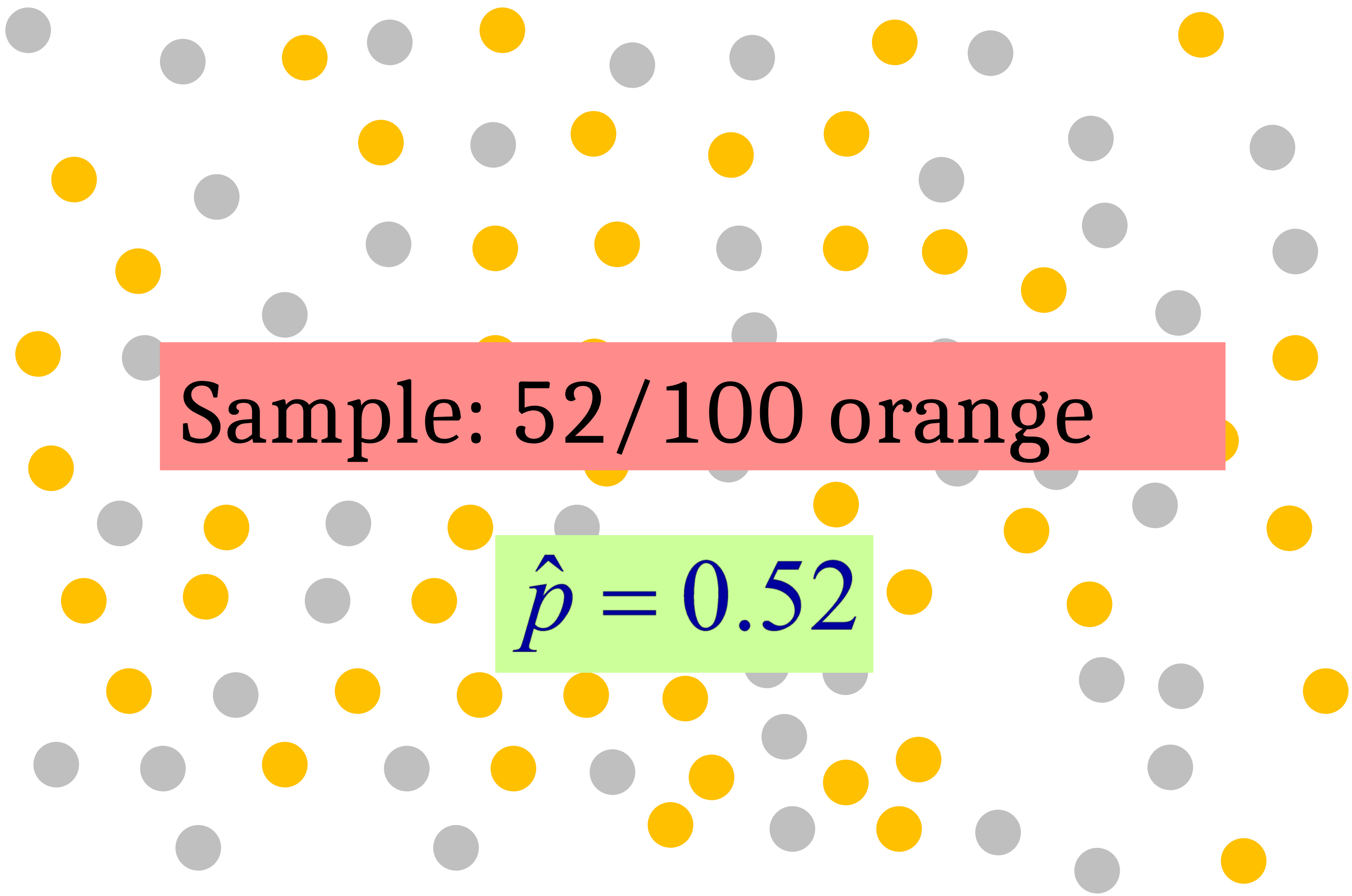
Reality

… WE ONLY HAVE ONE SAMPLE!!!!

- How do we know how much sample statistics vary, if we only have one sample?!?

BOOTSTRAP!

A sample of binary variables



Sample: 52/100 orange

$\hat{p} = 0.52$

Where might the "true" p be?

“Population”

- Imagine the “population” is many, many copies of the original sample

- (What do you have to assume?)

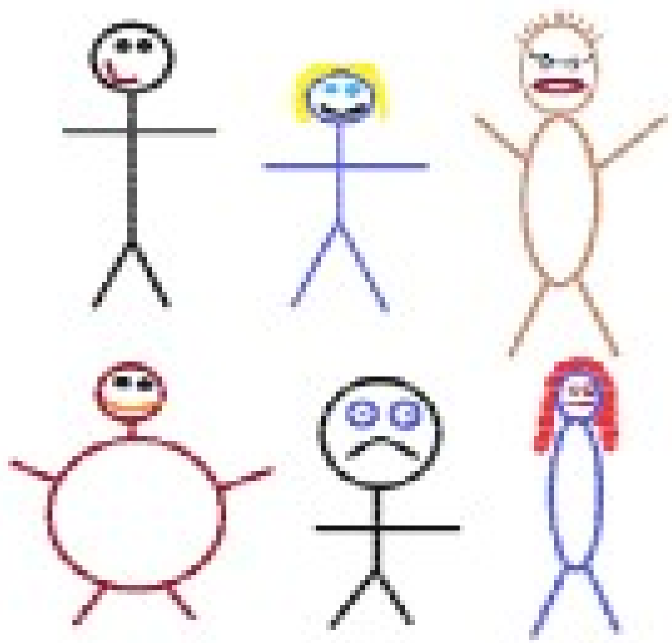
Augment the sample to a Population”

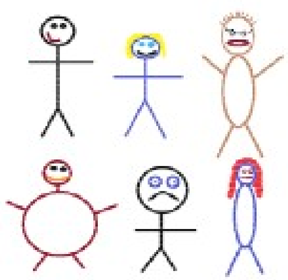
Sample repeatedly from this “population”

Sampling with Replacement

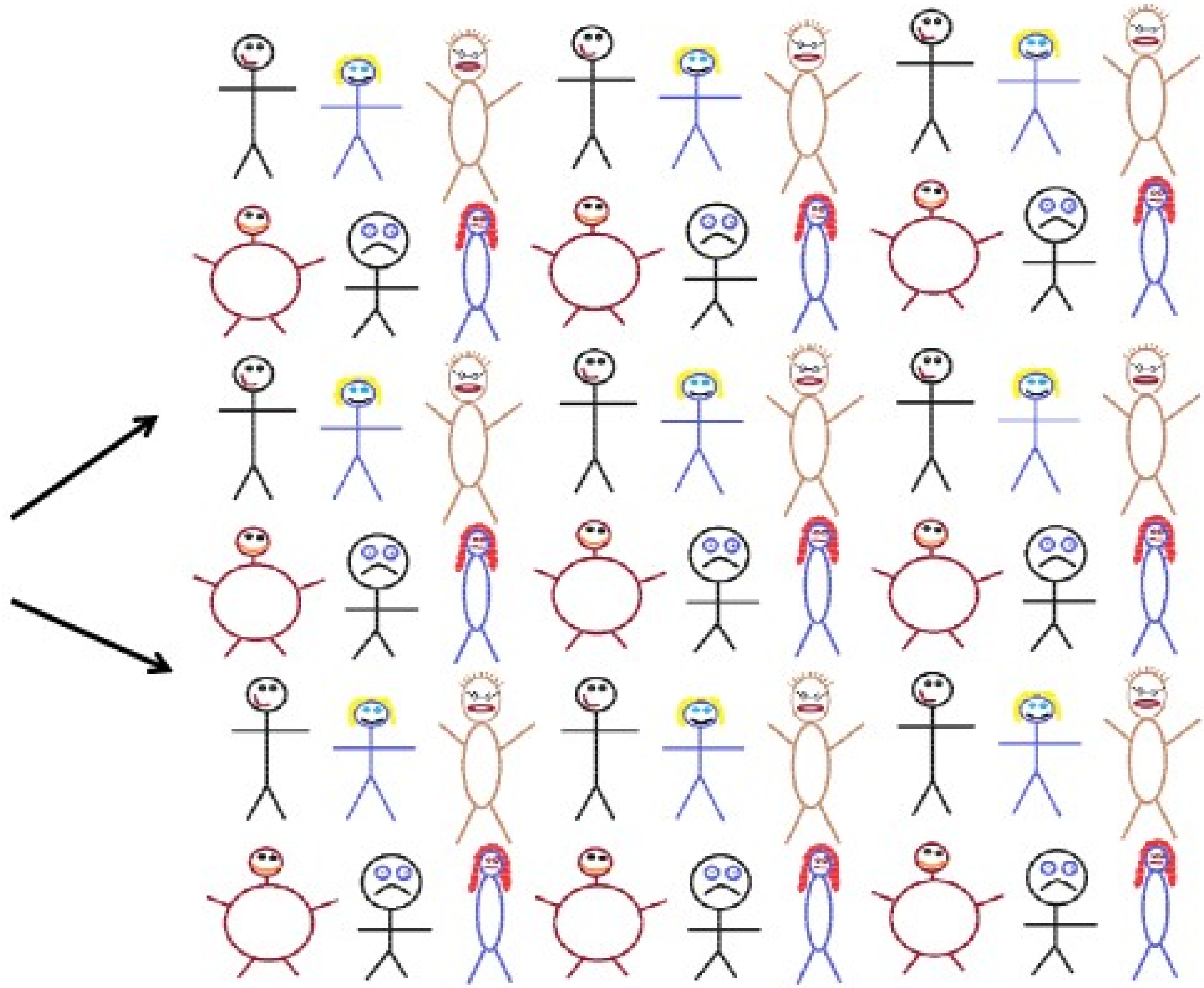
- To simulate a sampling distribution, we can just take repeated random samples from this "population" made up of many copies of the sample
- In practice, we can't actually make infinite copies of the sample…
- … but we can do this by sampling *with replacement* from the sample we have (each unit can be selected more than once)

Suppose we have a random sample of 6 people:



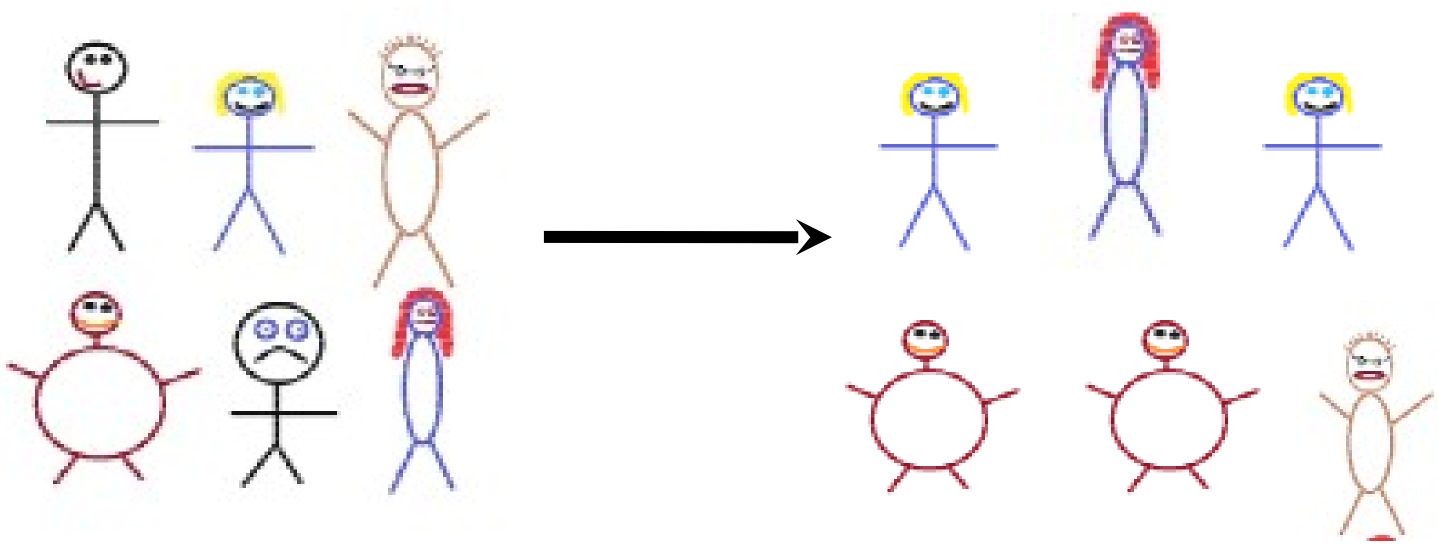


Original Sample



A simulated “population” to sample from

Bootstrap Sample: Sample with replacement from the original sample, using the same sample size.



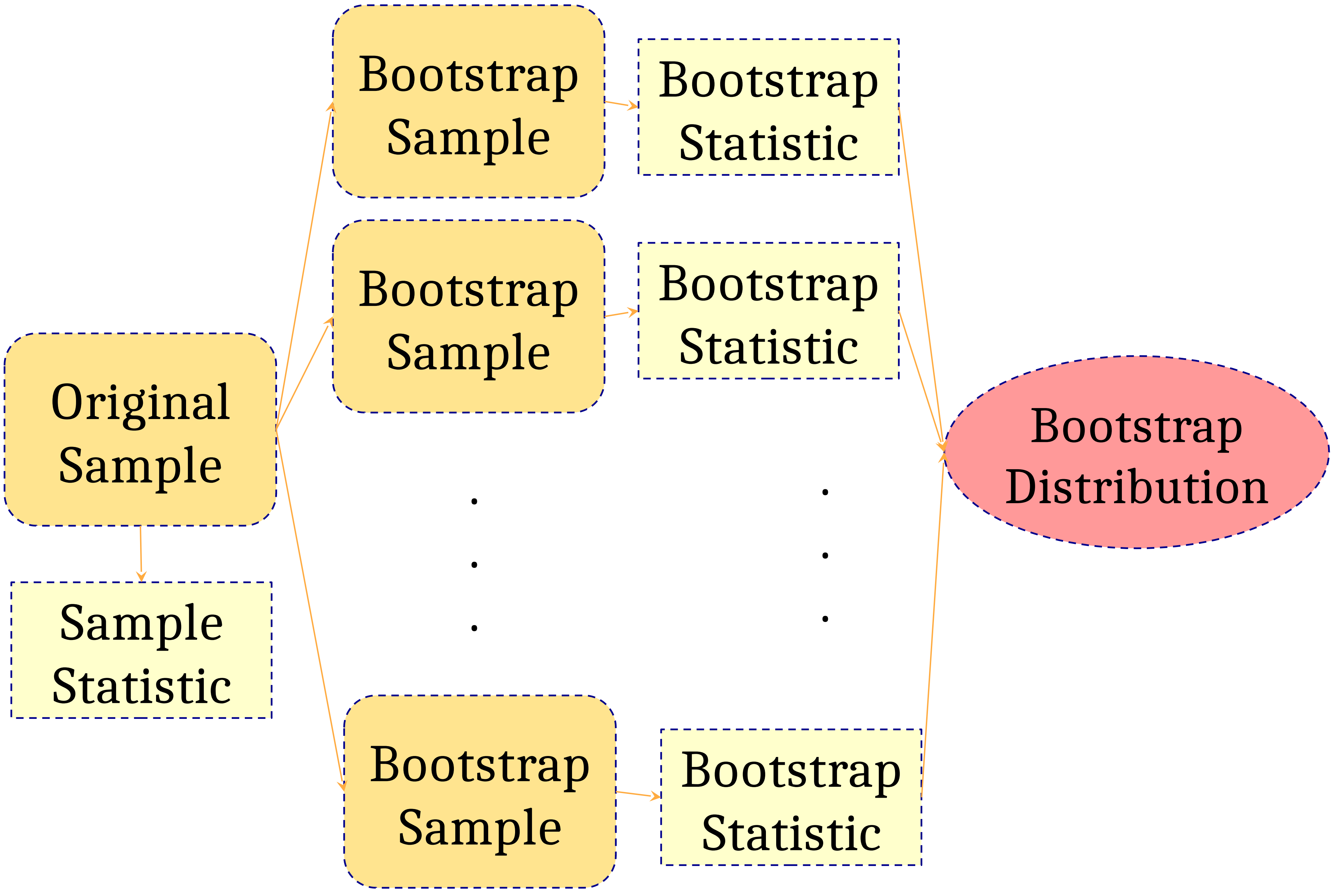
Original Sample

Bootstrap Sample

A ***bootstrap sample*** is a random sample taken with replacement from the original sample, of the same size as the original sample

A ***bootstrap statistic*** is the statistic computed on a bootstrap sample

A ***bootstrap distribution*** is the distribution of many bootstrap statistics


Original Sample
Sample Statistic
Bootstrap Sample
Bootstrap Statistic
Bootstrap Sample
Bootstrap Statistic
Bootstrap Sample
Bootstrap Statistic
Bootstrap Distribution

Bootstrap Sample

Your original sample has data values

18, 19, 19, 20, 21

Is the following a possible bootstrap sample?

18, 19, 20, 21, 22

a) Yes
b) No

22 is not a value from the original sample

Bootstrap Sample

Your original sample has data values

18, 19, 19, 20, 21

Is the following a possible bootstrap sample?

18, 19, 20, 21

a) Yes
b) No

Bootstrap samples must be the same size as the original sample

Bootstrap Sample

Your original sample has data values

18, 19, 19, 20, 21

Is the following a possible bootstrap sample?

18, 18, 19, 20, 21

Same size, could be gotten by sampling with replacement

a) Yes
b) No

Bootstrap Sample

You have a sample of size n = 50. You sample with replacement 1000 times to get 1000 bootstrap samples.

What is the sample size of each bootstrap sample?

Bootstrap samples are the same size as the original sample

(a) 50

(b) 1000

Bootstrap Distribution

You have a sample of size $n = 50$. You sample with replacement 1000 times to get 1000 bootstrap samples.

How many bootstrap statistics will you have?

(a) 50
(b) 1000

One bootstrap statistic for each bootstrap sample

Why “bootstrap”?

“Pull yourself up by your bootstraps”

- Lift yourself in the air simply by pulling up on the laces of your boots
- Metaphor for accomplishing an “impossible” task without any outside help

Golden Rule of Bootstrapping

Bootstrap statistics are to the original sample statistic

as

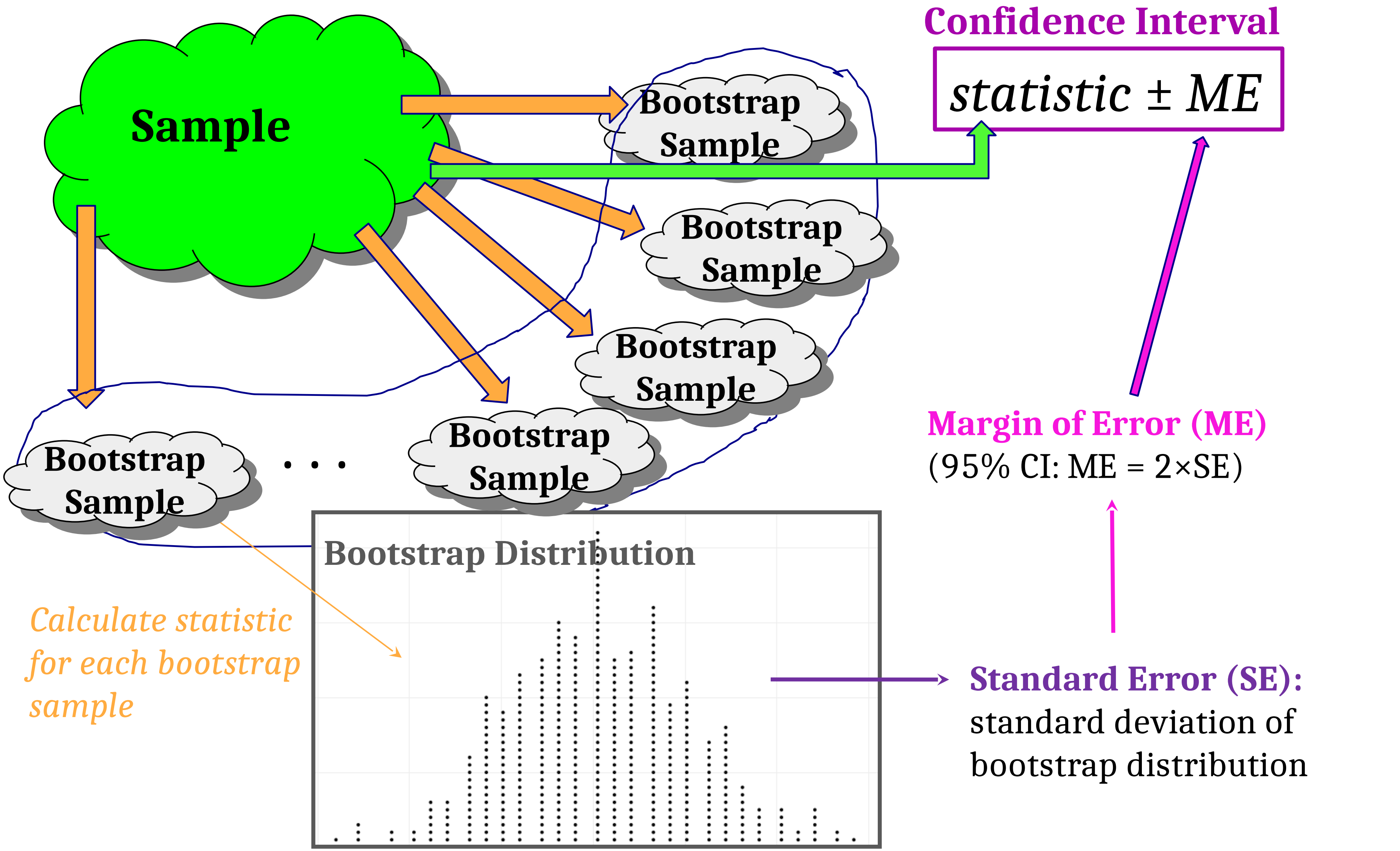
the original sample statistic is to the population parameter

Standard Error

- The variability of the bootstrap statistics is similar to the variability of the sample statistics

- The standard error of a statistic can be estimated using the standard deviation of the bootstrap distribution!

Confidence Intervals



What about Other Parameters or functions of parameters?

Estimate the standard error and/or a confidence interval for...

- proportion (p)
- difference in means $(\mu_1 - \mu_2)$
- difference in proportions $(p_1 - p_2)$
- standard deviation (σ)
- correlation (ρ)
- ...

Generate samples with replacement
Calculate sample statistic
Repeat...

The Magic of Bootstrapping

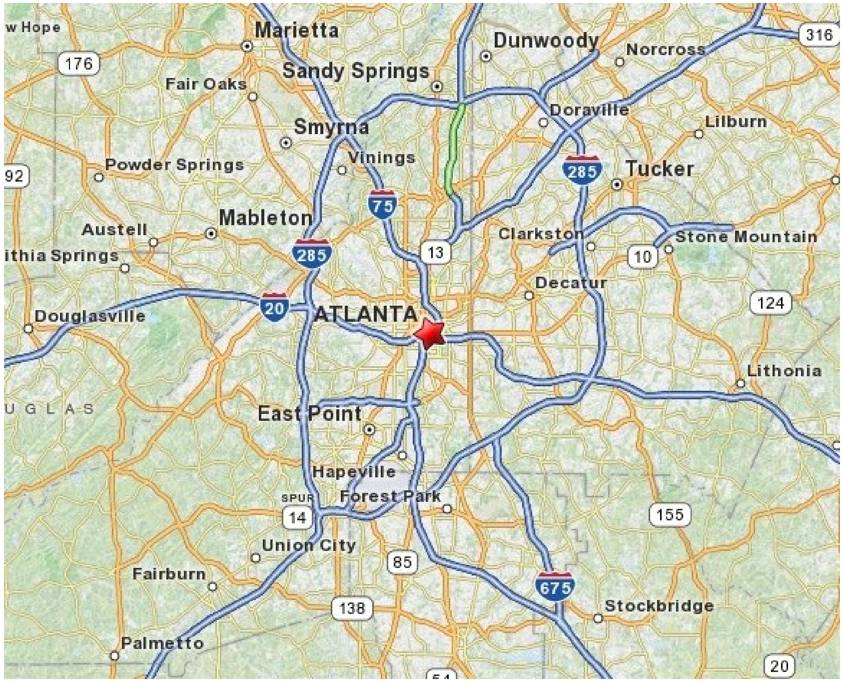
- We can use bootstrapping to assess the uncertainty surrounding ANY sample statistic!

- If we have sample data, we can use bootstrapping to create a 95% confidence interval for any parameter! (well, almost…)

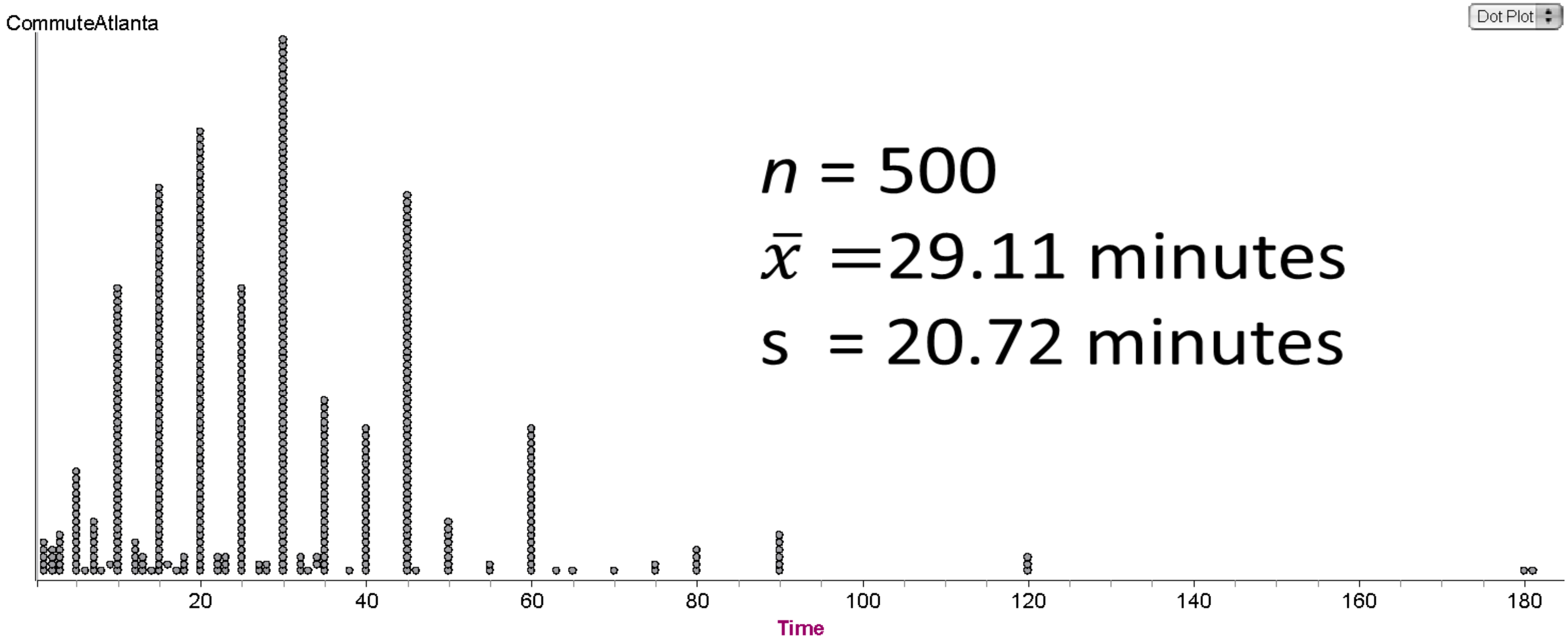
Atlanta Commutes

What's the mean commute time for workers in metropolitan Atlanta?

Data: The American Housing Survey (AHS) collected data from Atlanta in 2004



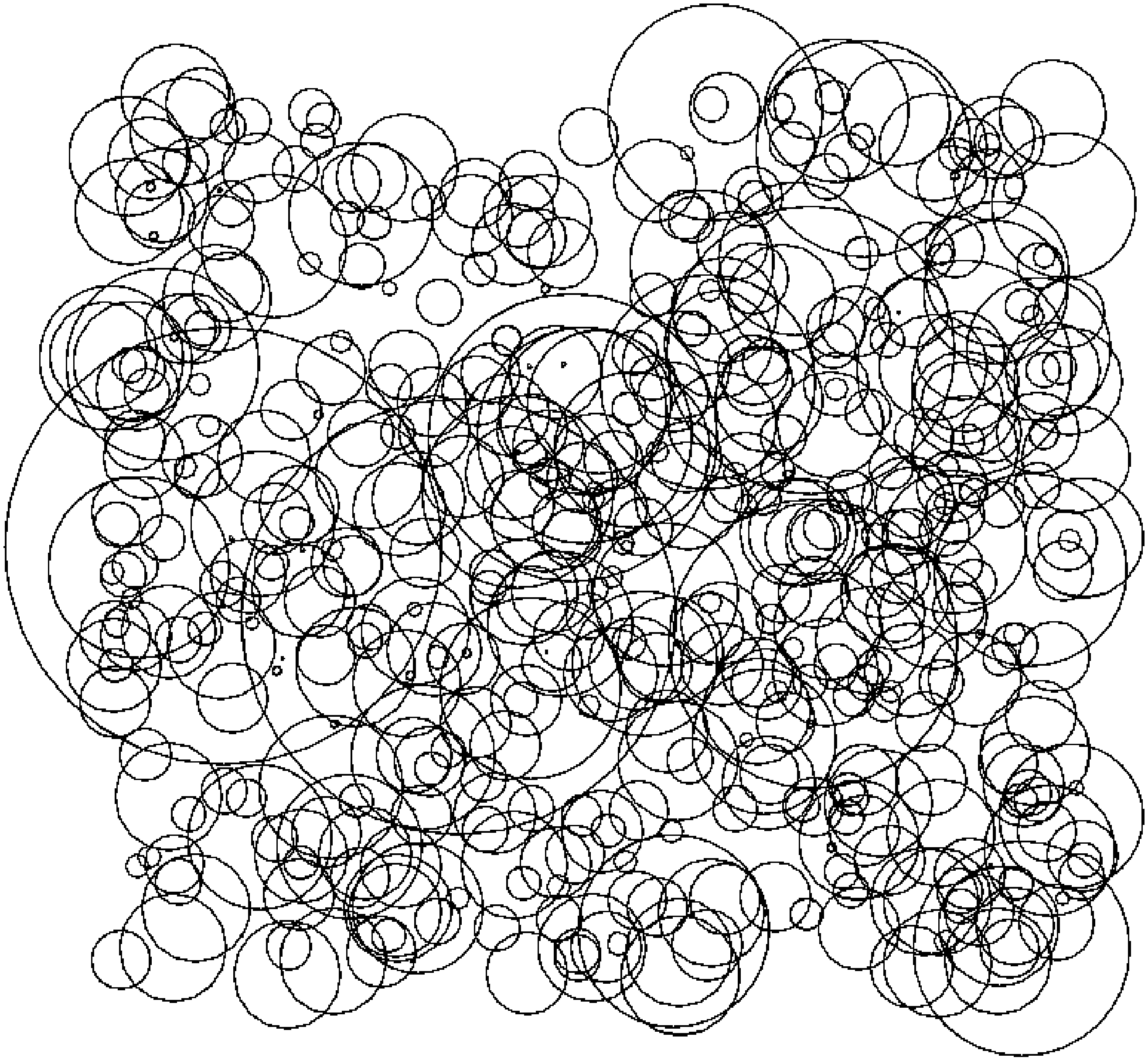
Random Sample of 500 Commutes



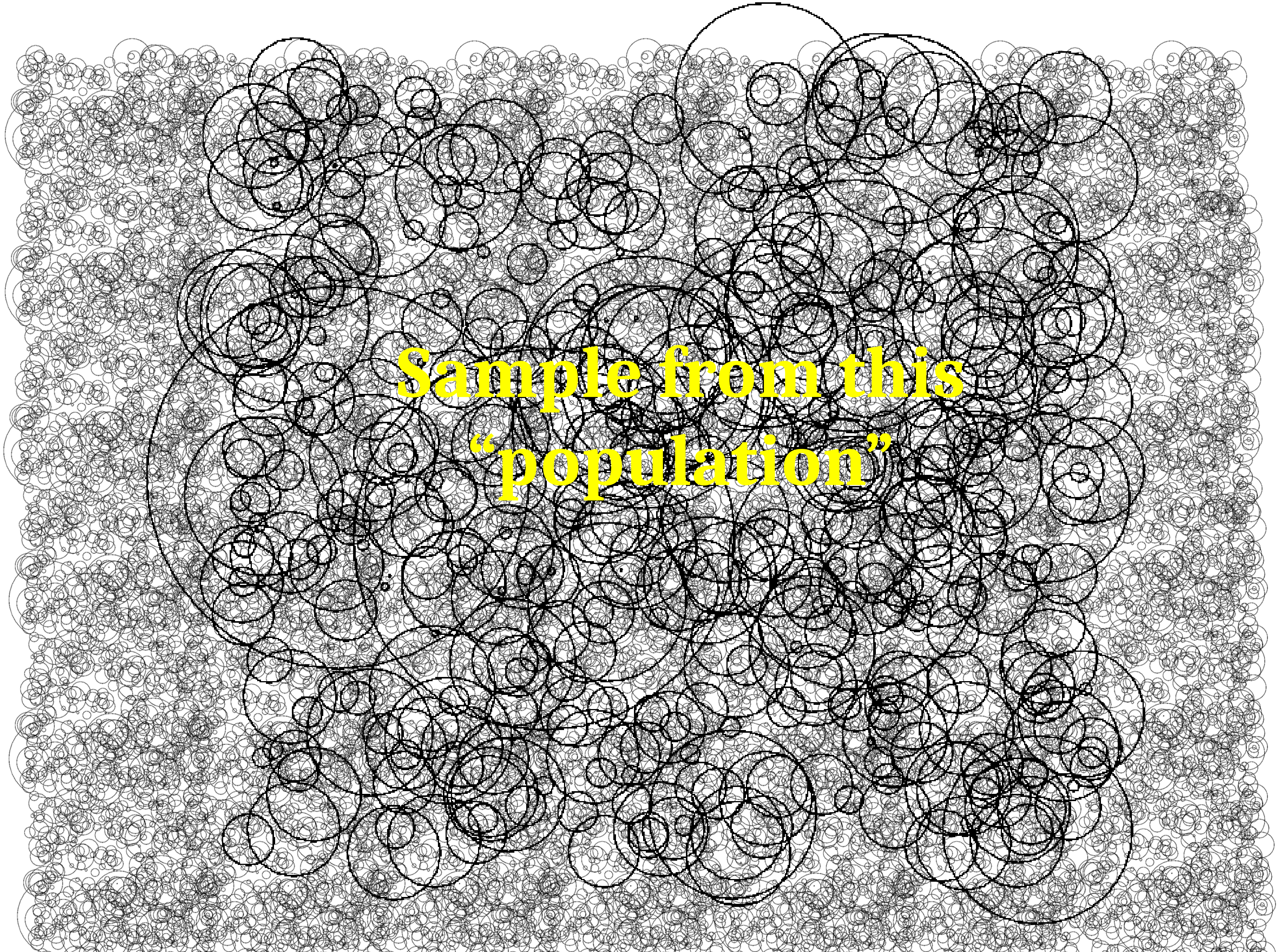
Where might the “true” μ be?

WE CAN BOOTSTRAP TO FIND OUT!!!

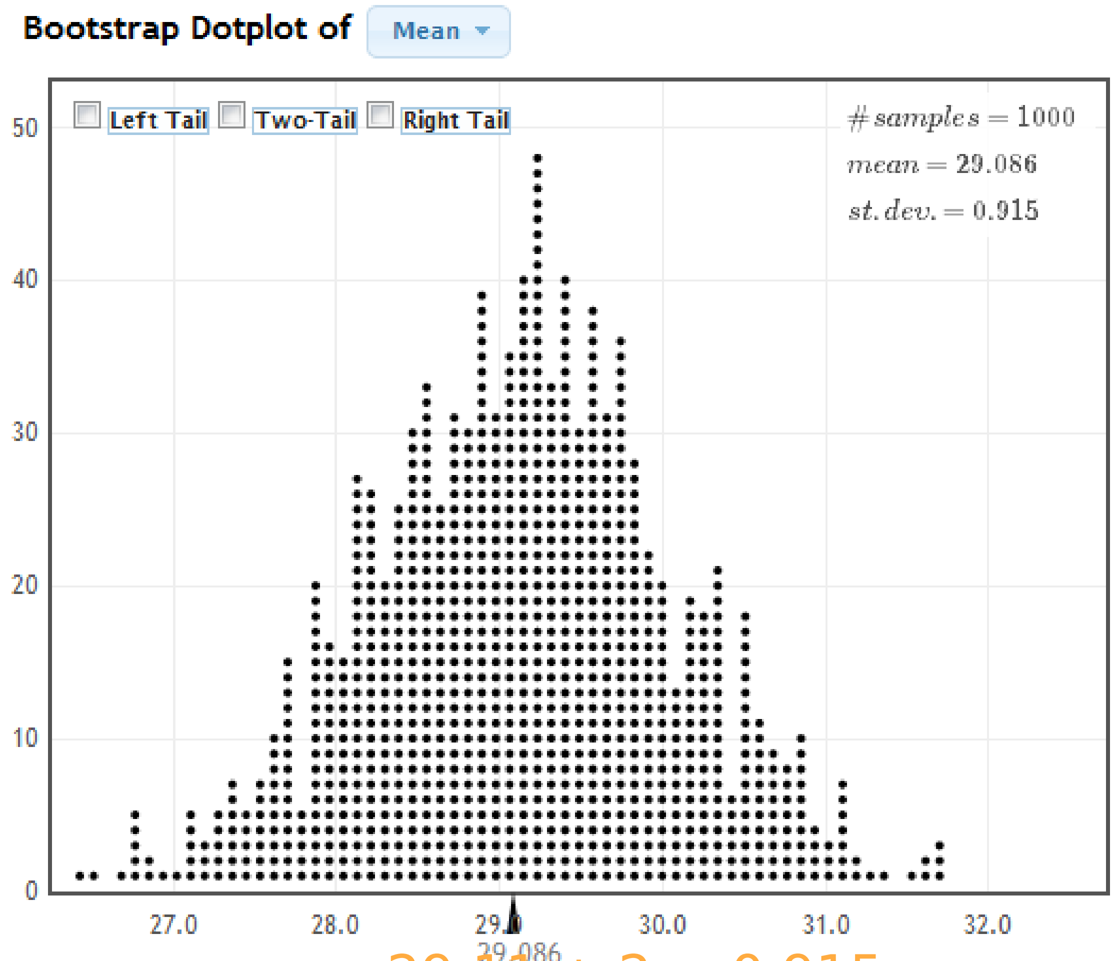
Original Sample

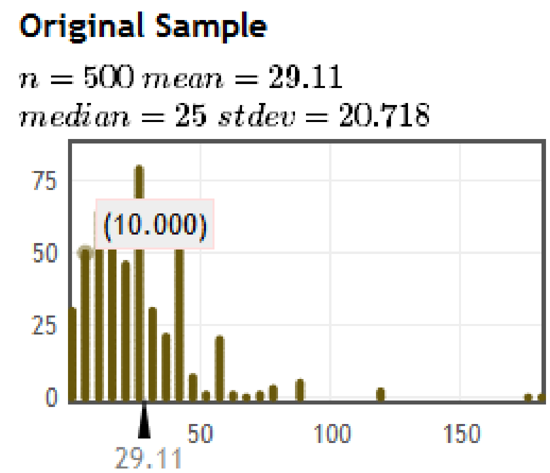


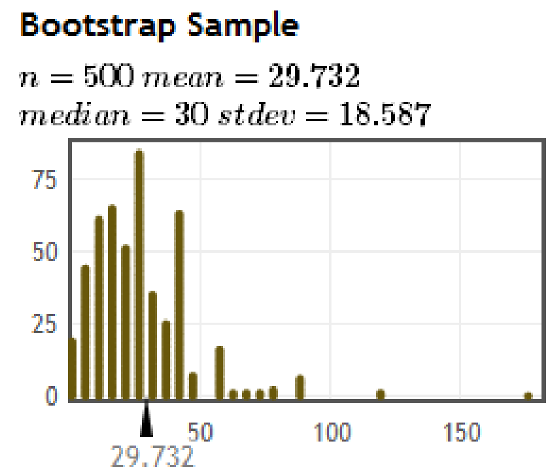
“Population” = many copies of sample



Atlanta Commutes





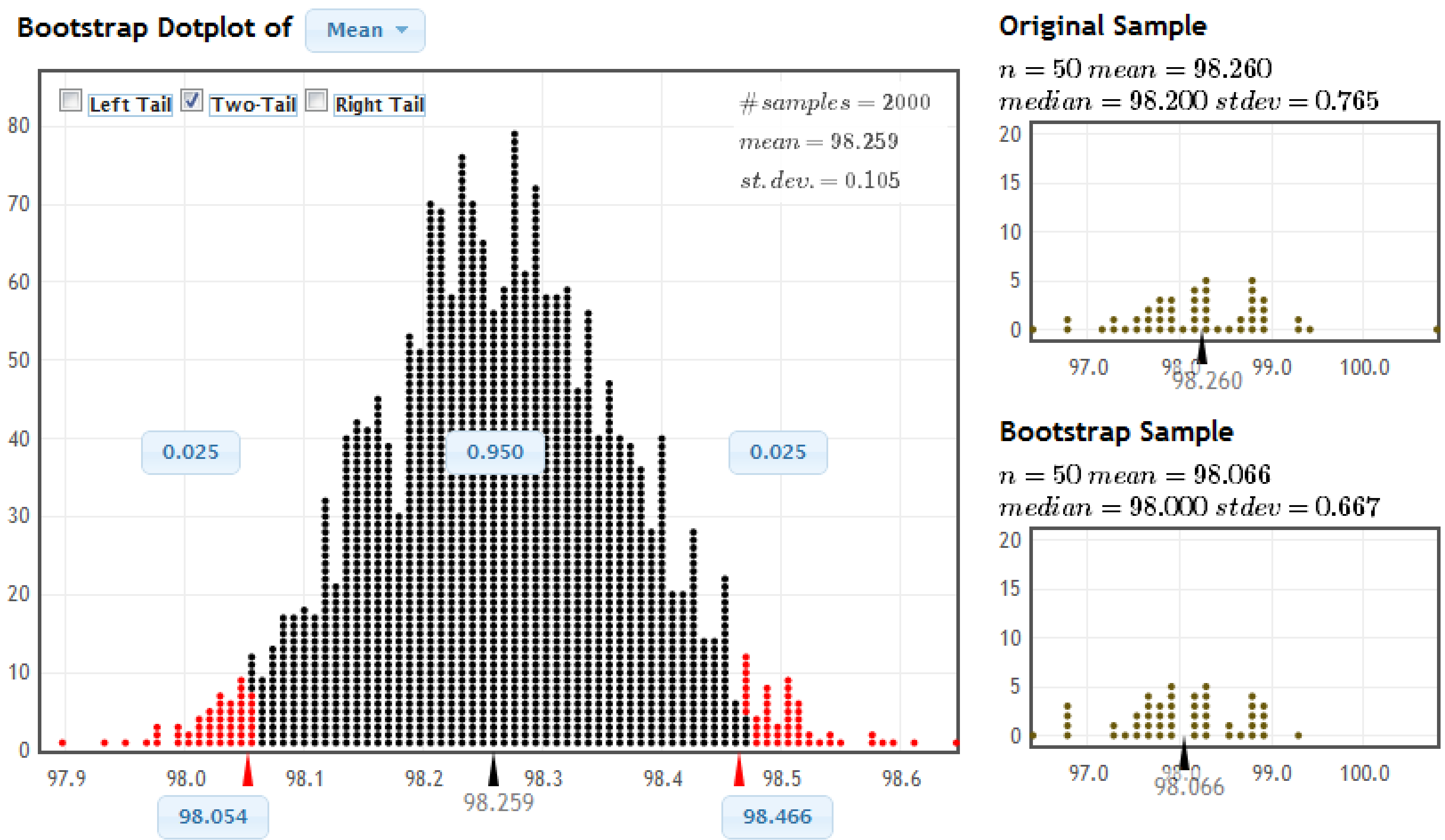


$29.11 \pm 2 \times 0.915$

95% confidence interval for the average commute time for Atlantans:

(a) (28.2, 30.0) (b) (27.3, 30.9) (c) 26.6, 31.8 (d) No idea

Two Methods for 95%



statistic ± 2 × SE :

$$98.26 \pm 2 \times 0.105 = (98.05, 98.47)$$

Percentile Method :

$$(98.05, 98.47)$$

Two Methods for SE

- For a symmetric, bell-shaped bootstrap distribution, using either the standard error method or the percentile method will given similar 95% confidence intervals

- If the bootstrap distribution is not bell-shaped or if a level of confidence other than 95% is desired, use the percentile method

Criteria for Bootstrap CI

Using the percentile method for a confidence interval bootstrapping for a confidence interval works for any statistic, as long as the bootstrap distribution is

1. Approximately symmetric
2. Approximately continuous

Using the standard error method also requires

3. Approximately bell-shaped

Always look at the bootstrap distribution to make sure these are true!

Number of Bootstrap Samples

- The more bootstrap samples you use, the more precise your answer will be.
- In real data analysis, you probably want to take 10,000 or even 100,000 bootstrap samples
- Start from a small number of bootstrap samples
- In R, use the function “sample(..., replace = T)”

Summary

1. Confidence intervals can be created using the standard error or the percentiles of a bootstrap distribution for any parameter, as long as the bootstrap distribution is approximately symmetric and continuous
2. Bootstrap may not be appropriate if

- distributions that do not have finite moments
- small sample sizes
- estimating extreme values from the distribution
- ...