Bootstrap Intervals

Dr. Kari Lock Morgan

PSU 016

11/12/14

p-value

- How extreme would *your observed statistic* be, under the null hypothesis?
- Many of you calculated your p-value without ever using your observed statistic.

Sleep versus Caffeine

- Students were given words to memorize, then randomly assigned to take either a 90 min nap, or a caffeine pill. 2 ½ hours later, they were tested on their recall ability.
- Explanatory variable: sleep or caffeine
- Response variable: number of words recalled
- Is sleep better than caffeine for memory?

Mednick, Cai, Kanady, and Drummond (2008). "Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory," *Behavioral Brain Research,* 193, 79-86.

IMPORTANT POINTS

- Sample statistics ***vary*** from sample to sample. (they will not match the parameter exactly)

- ***KEY QUESTION***: For a given sample statistic, what are plausible values for the population parameter? How much uncertainty surrounds the sample statistic?

- ***KEY ANSWER***: It depends on how much the statistic varies from sample to sample!

Reese's Pieces

- What proportion of Reese's pieces are orange?
- Take a random sample of 10 Reese's pieces. What is your sample proportion?
- Come to the board to make a class dotplot
- You just made a sampling distribution!

Sampling Distribution

A ***sampling distribution*** is the distribution of sample statistics computed for different samples of the same size from the same population.

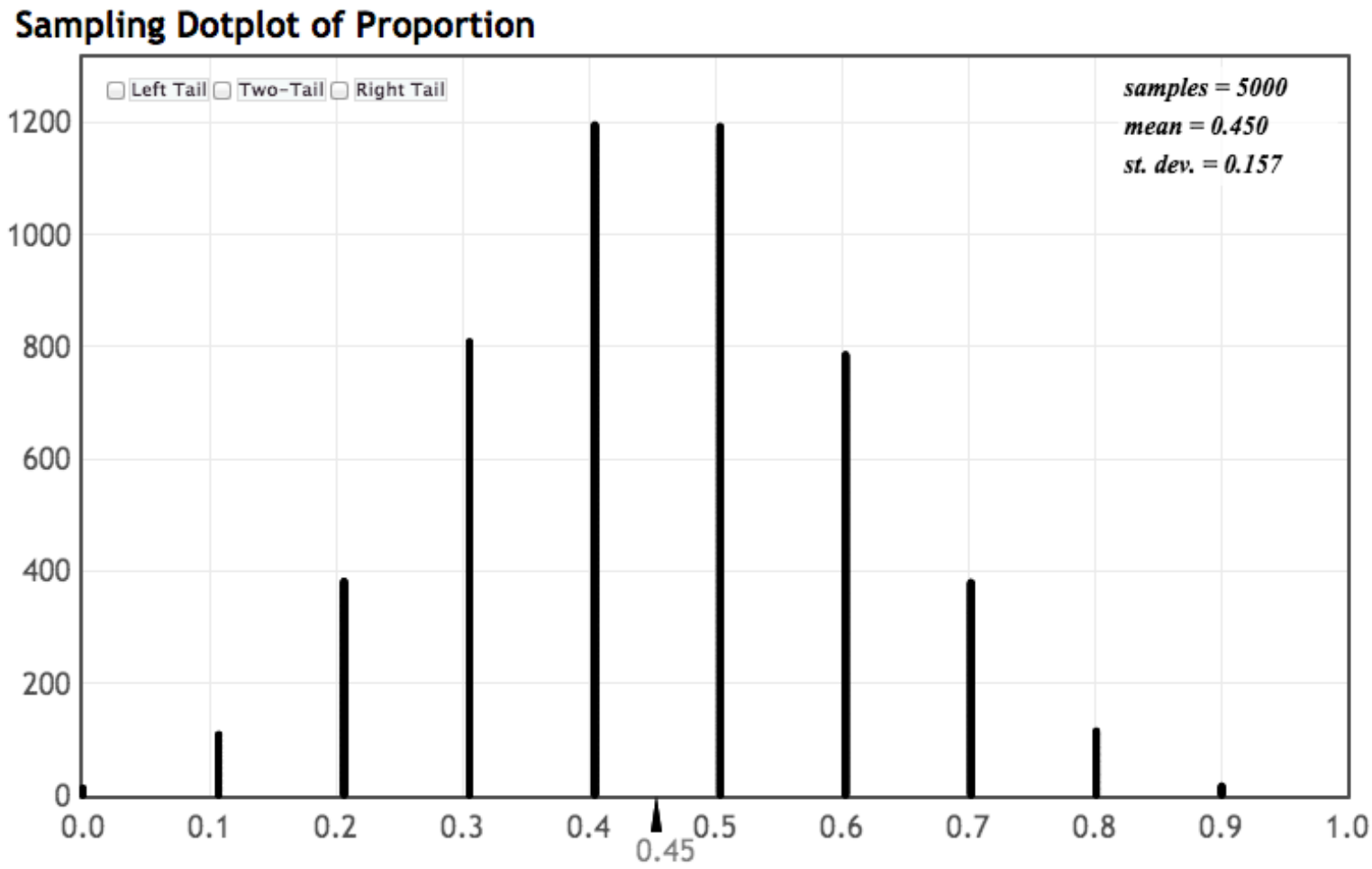
- A sampling distribution shows us how the sample statistic varies from sample to sample

Lots of simulations!

- We need many more simulations!

www.lock5stat.com/statkey

Reese's Pieces

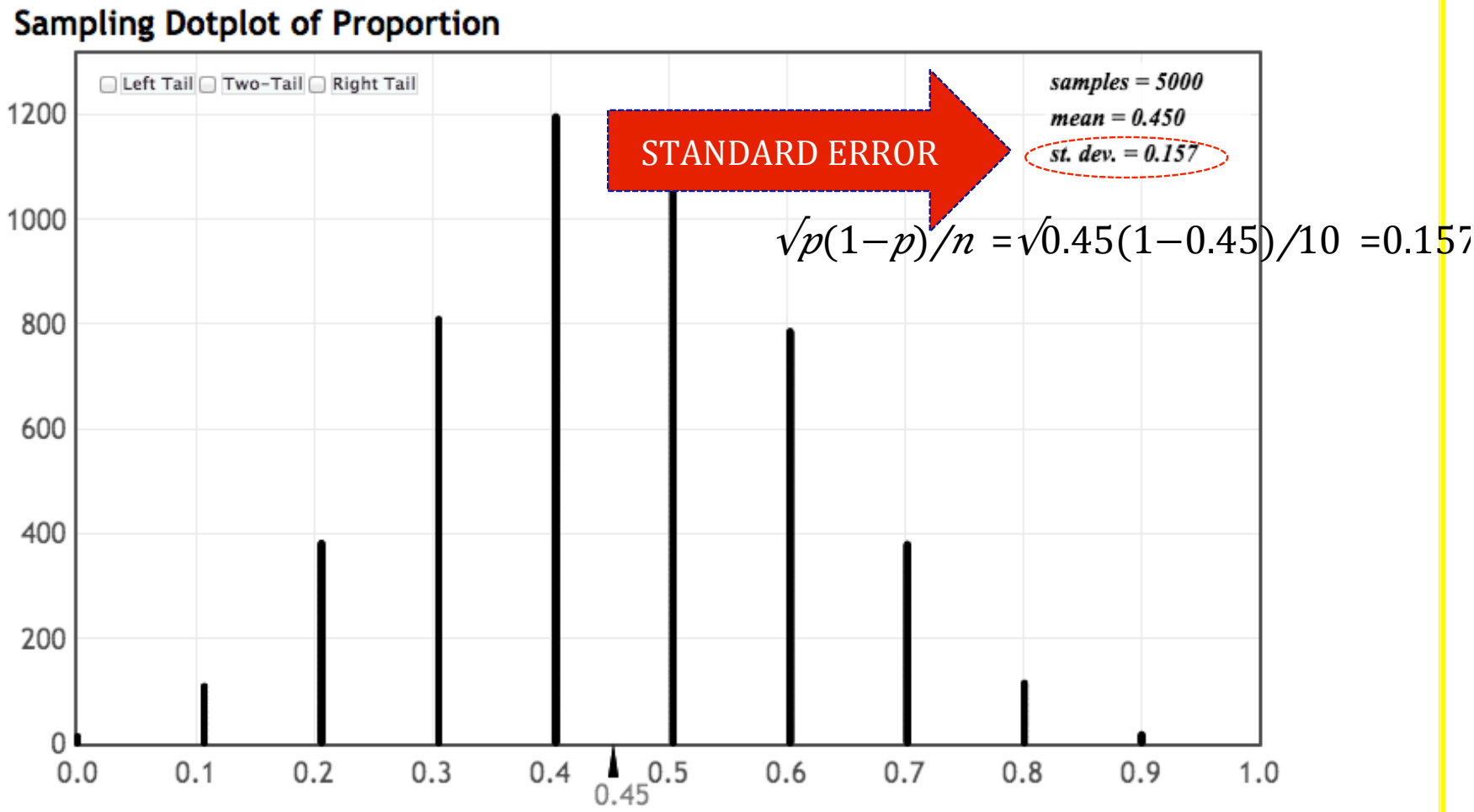


Standard Error

- The **variability** of the sample statistic (how much it varies from sample to sample) is so important it gets it's own name…

> The ***standard error*** of a statistic, SE, is the standard deviation of the sample statistic

Reese's Pieces



$$\sqrt{p(1-p)/n} = \sqrt{0.45(1-0.45)/10} = 0.157$$

95% Confidence Interval

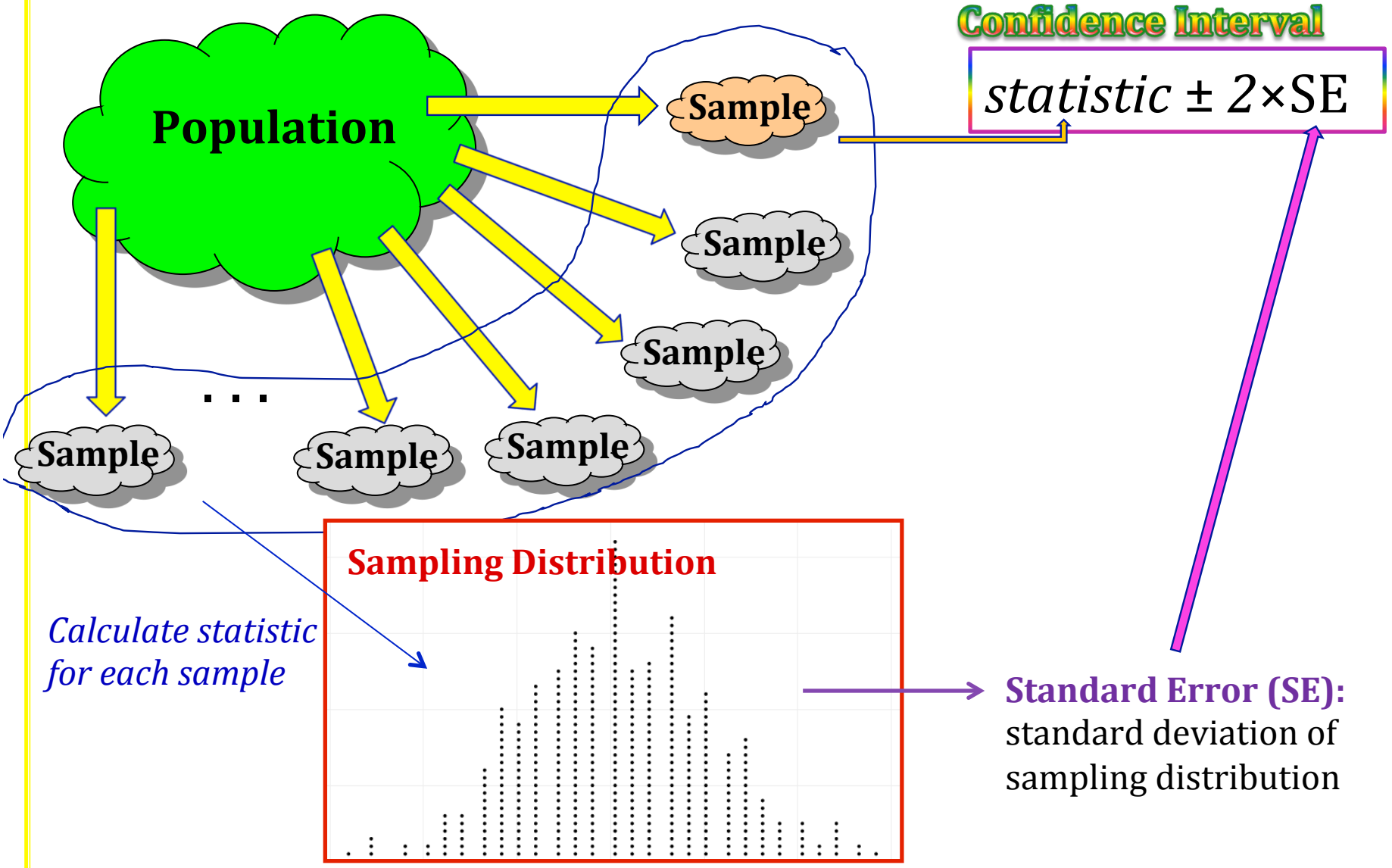
If the sampling distribution is relatively symmetric and bell-shaped, a 95% confidence interval can be estimated using

statistic ± 2 × SE

Reese's Pieces

- Use our estimated SE and your p to create a 95% confidence interval based on your data.
- Did your 95% interval include the true $p = 0.45$?
- What proportion of intervals will contain the true p?

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?!?

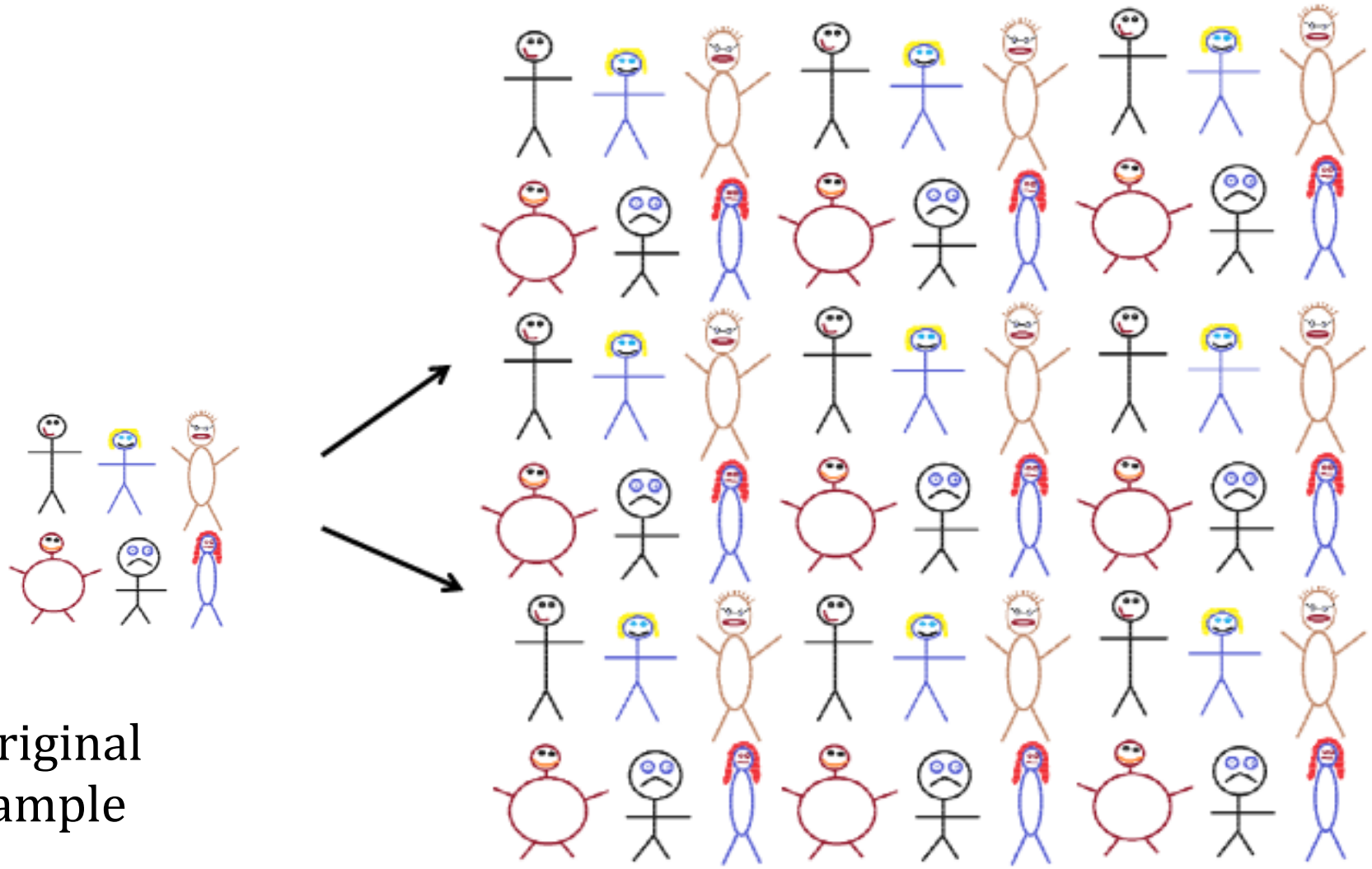
BOOTSTRAP!

"Population"

- Imagine the "population" is many, many copies of the original sample

- (What do you have to assume?)

Suppose we have a random sample of 6 people:



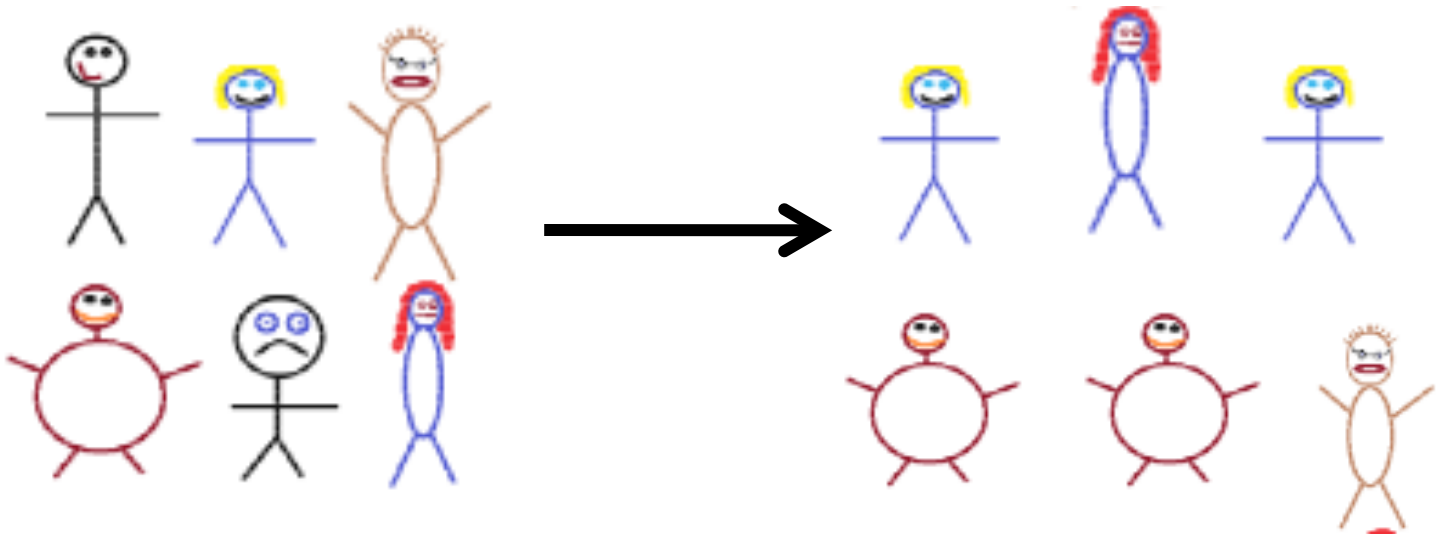
Original Sample

A simulated "population" to sample from

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)

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



Original Sample

Bootstrap Sample

Reese's Pieces

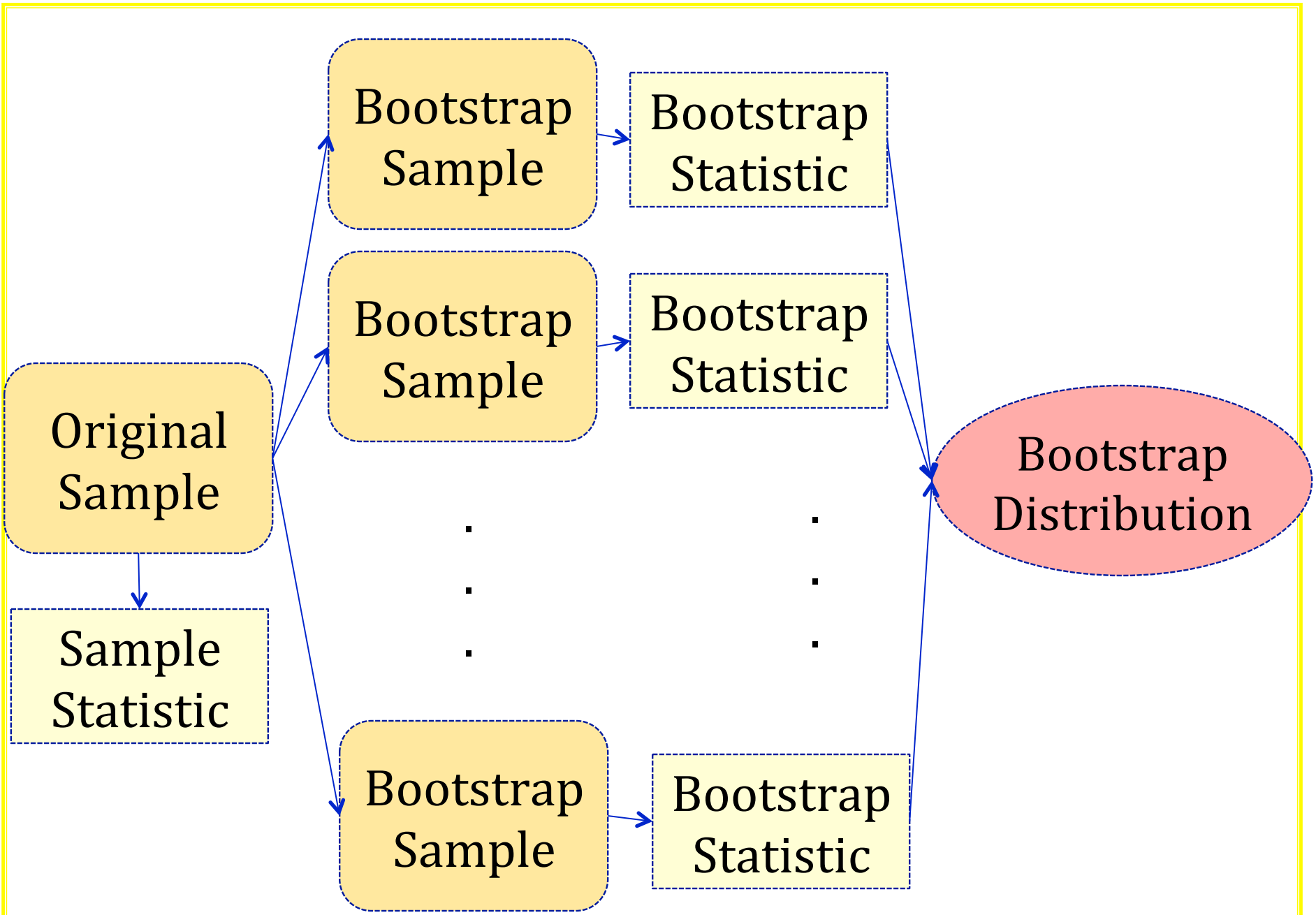
- Take a bootstrap sample from your sample of Reese's Pieces

Bootstrap

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

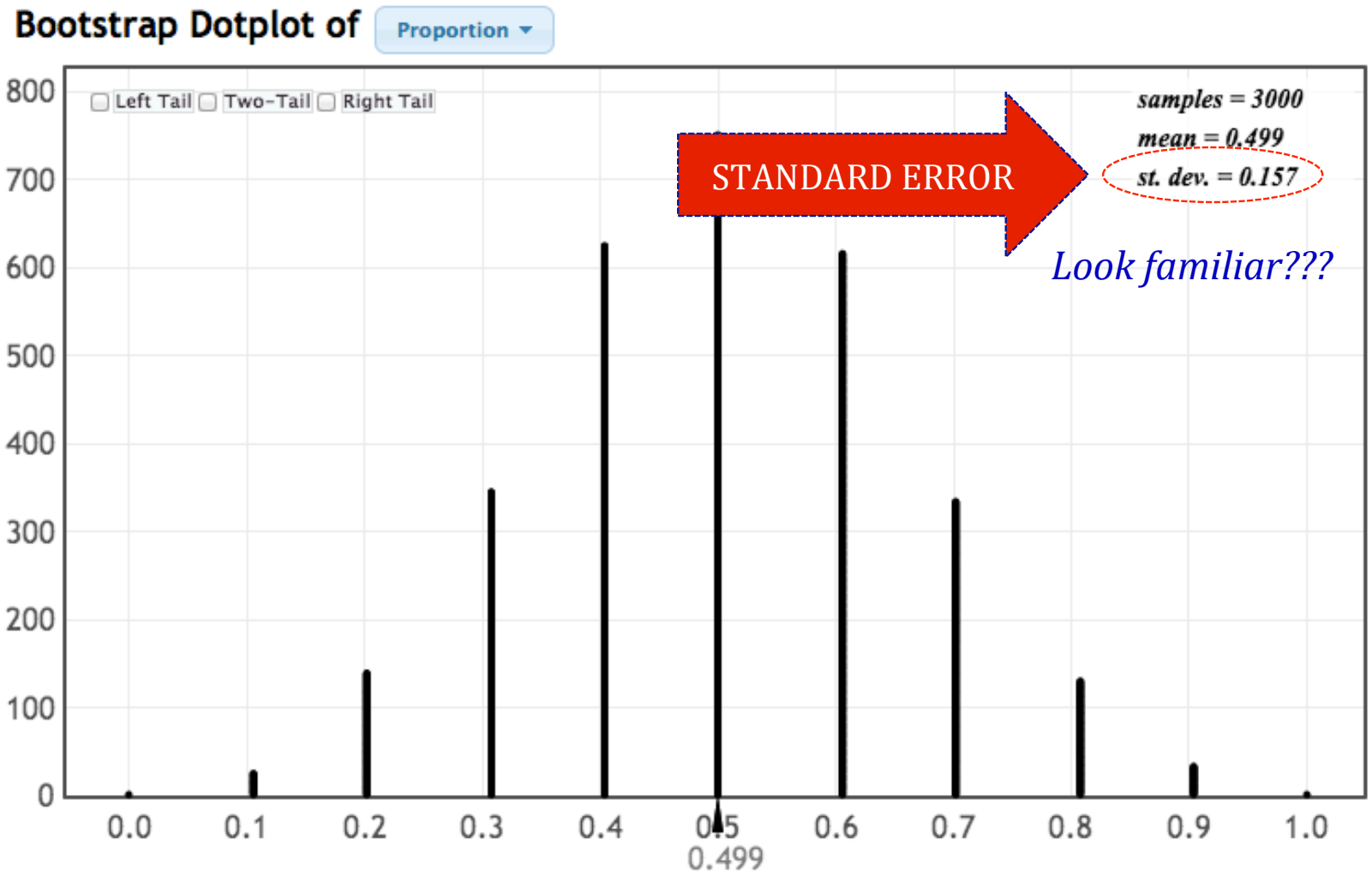

Bootstrap Sample
Bootstrap Statistic
Bootstrap Sample
Bootstrap Statistic
Original Sample
Bootstrap Distribution
Sample Statistic
Bootstrap Sample
Bootstrap Statistic

Lots of simulations!

- We need many more simulations!

www.lock5stat.com/statkey

Bootstrap Distribution



Why "bootstrap"?

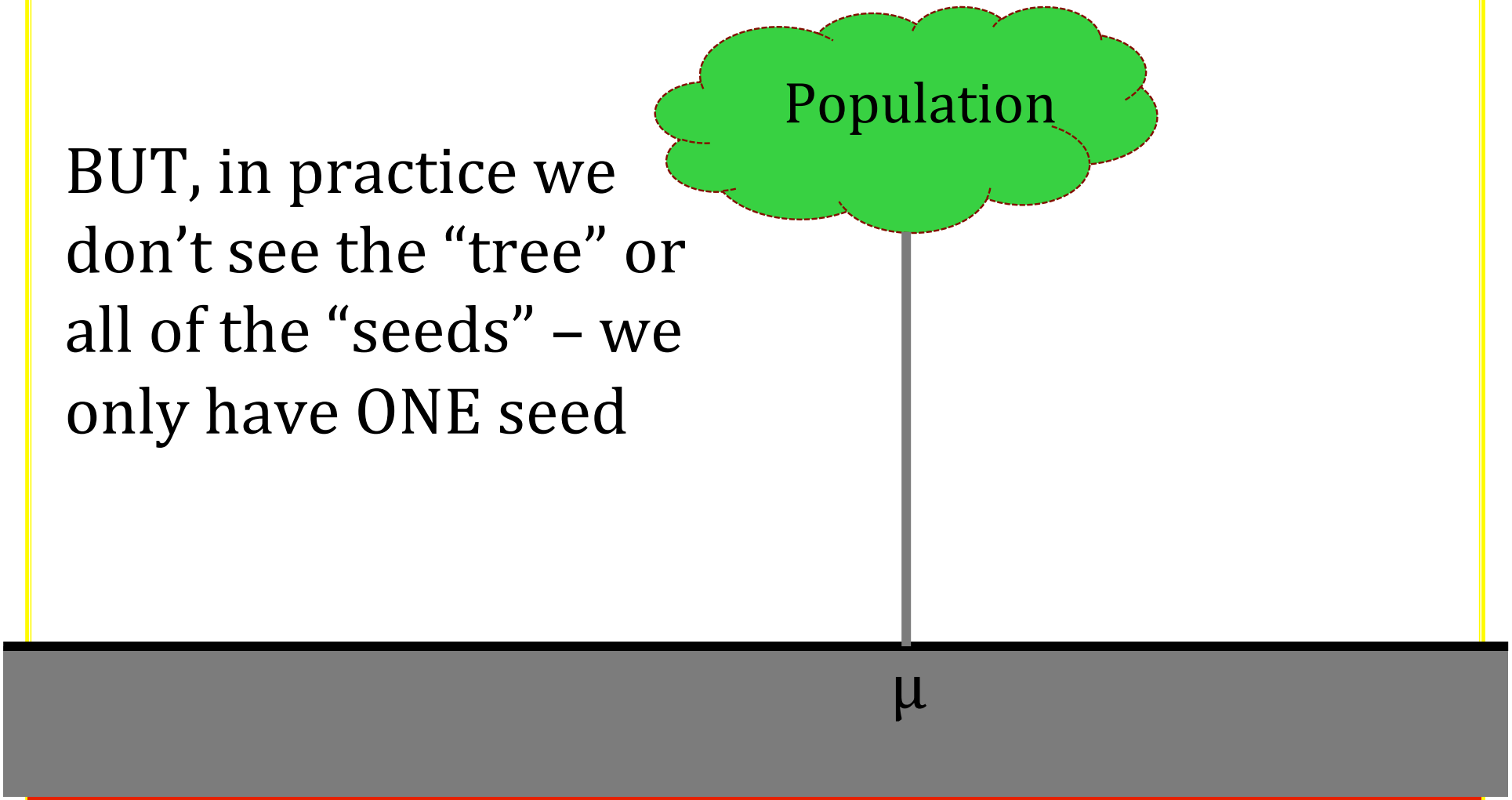
"Pull yourself up by your bootstraps"

- 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

Sampling Distribution

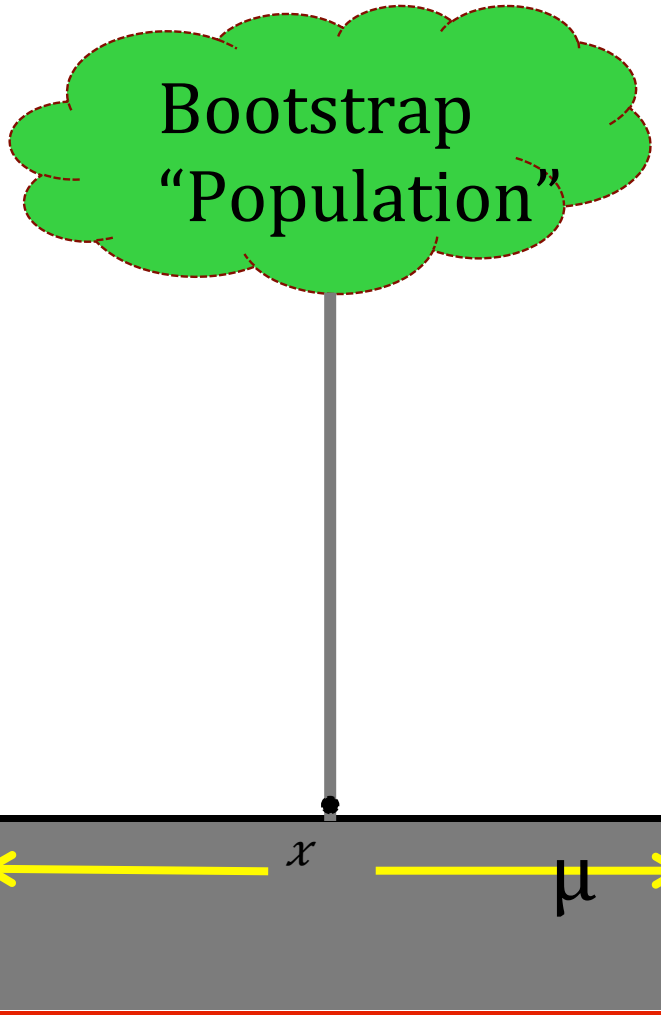
BUT, in practice we don't see the "tree" or all of the "seeds" – we only have ONE seed



Bootstrap Distribution

What can we do with just one seed?

Grow a NEW tree!



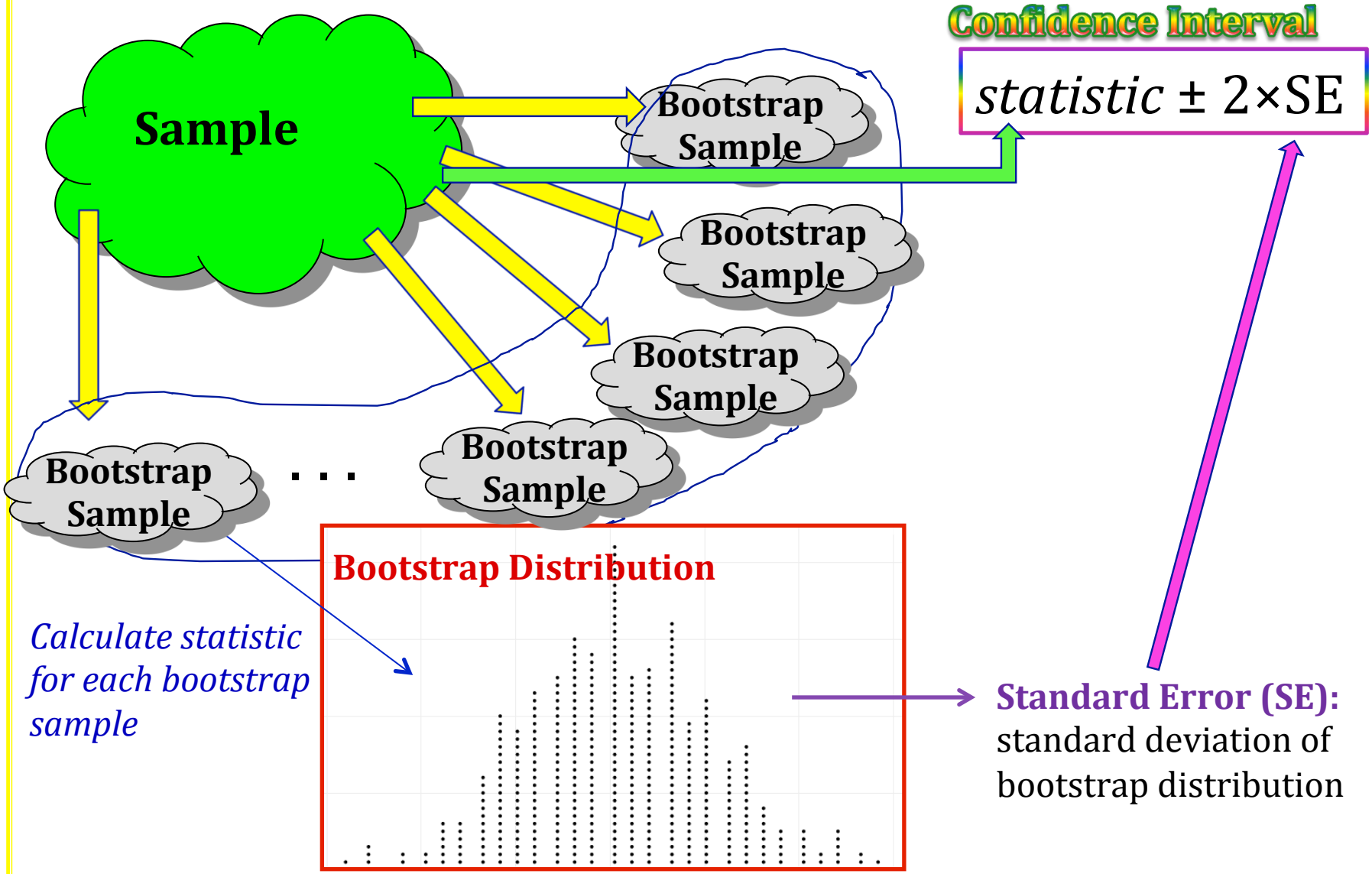
Estimate the distribution and variability (SE) of x's from the bootstraps

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



The Magic of Bootstrapping

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

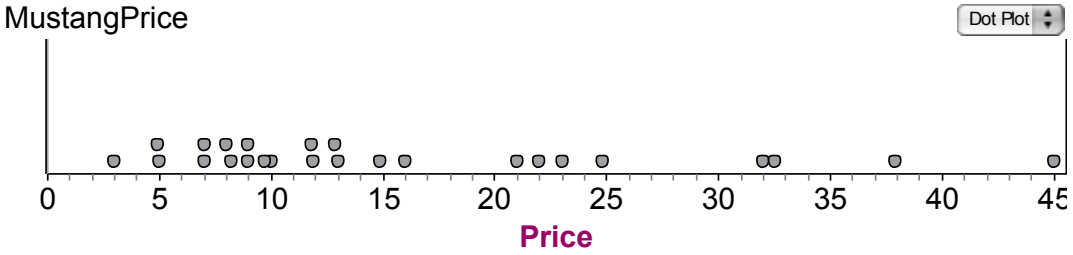
- If we have sample data, we can use bootstrapping to create a 95% confidence interval for any parameter!

(well, almost…)

Used Mustangs

- What's the average price of a used Mustang car?
- Select a random sample of n = 25 Mustangs from a website (autotrader.com) and record the price (in $1,000's) for each car.

Sample of Mustangs:



n=25 $\bar{x}$=15.98 s=11.11

Our best estimate for the average price of used Mustangs is $15,980, but how accurate is that estimate?

BOOTSTRAP!

Original Sample

1. Bootstrap Sample

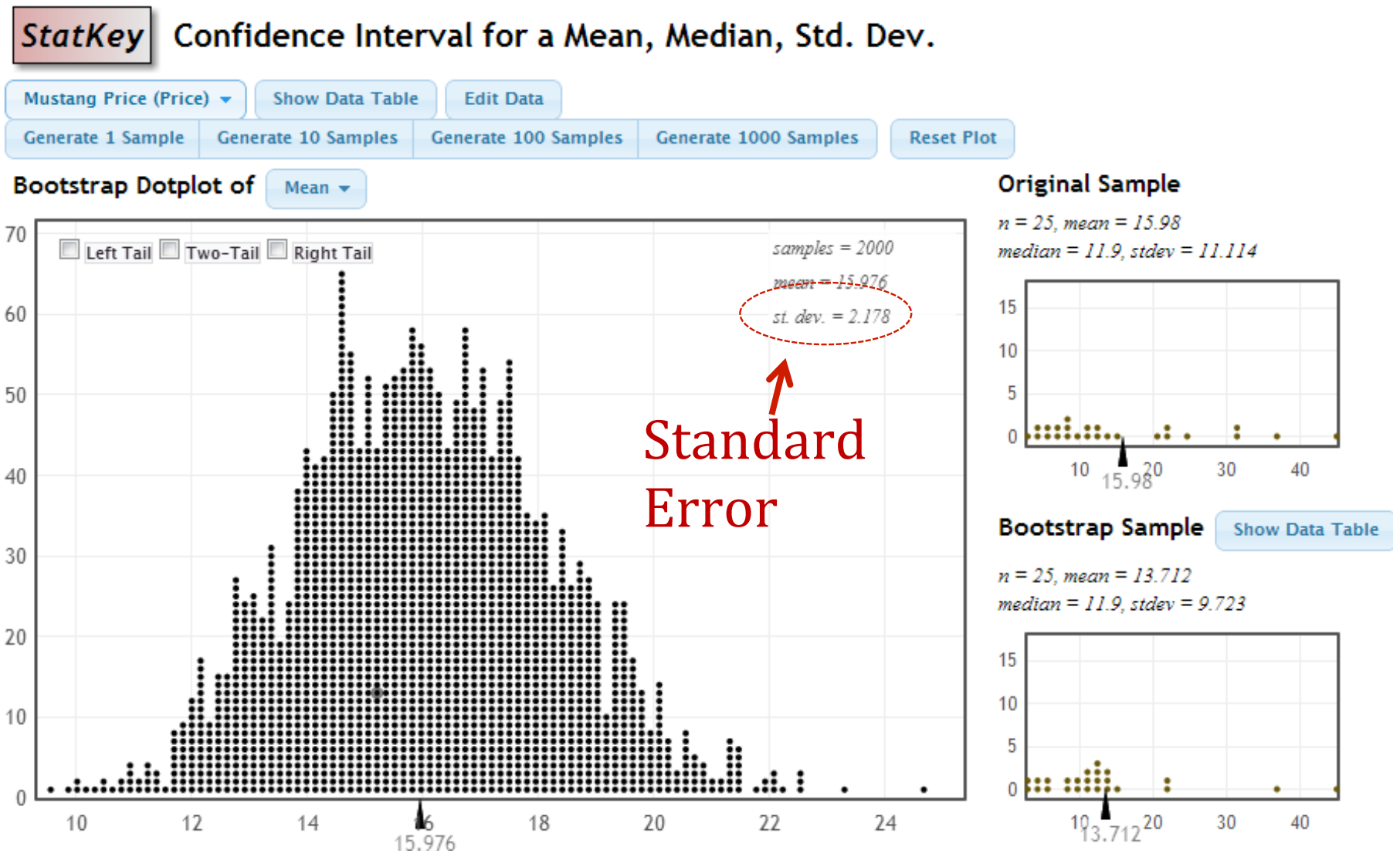
2. Calculate mean price of bootstrap sample

3. Repeat many times!

Used Mustangs

- Use StatKey (www.lock5stat.com/statkey) to generate your own 95% confidence interval for the price of used mustangs on autotrader.com.

Used Mustangs



Used Mustangs

- 95% CI:

$statistic \pm 2 \cdot SE$

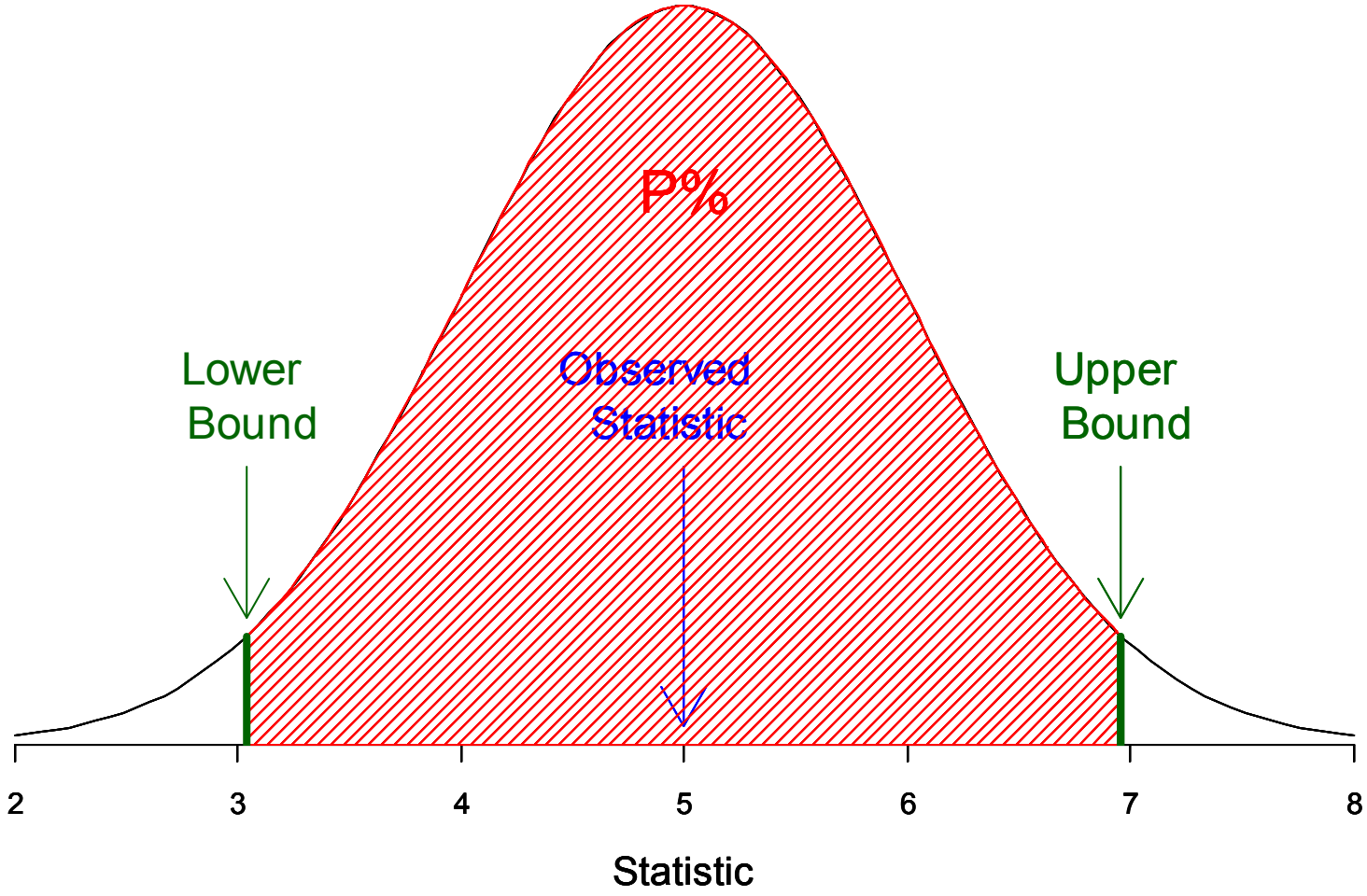
$\$15{,}980 \pm 2 \cdot \$2{,}178$

($11,624, $20,336)

- *We are 95% confident that the average price of a used Mustang on autotrader.com is between $11,624 and $20,336*

Other Levels of Confidence

- For a P% confidence interval:

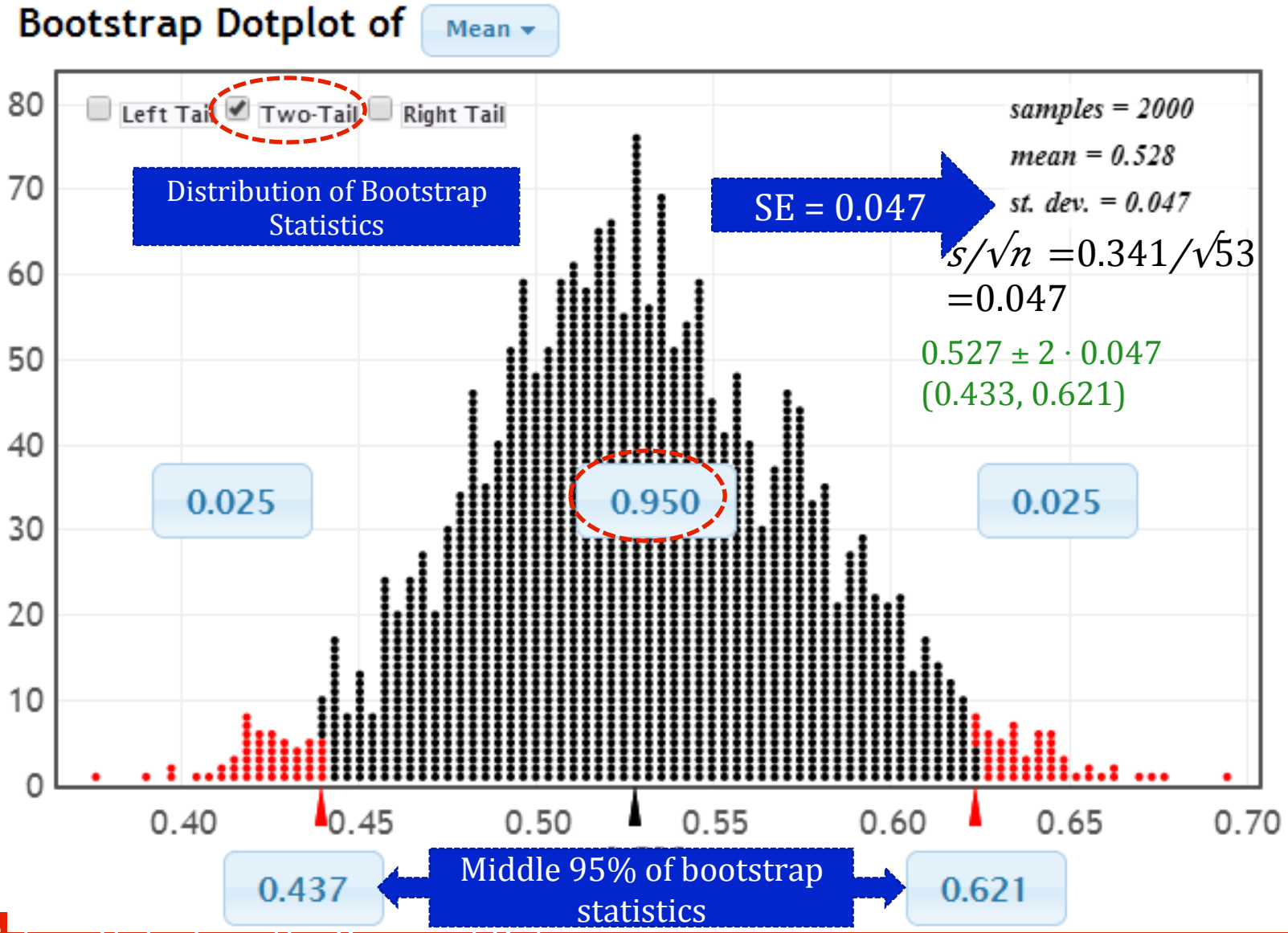


Mercury in Fish

- What is the average mercury level of fish (large mouth bass) in Florida lakes?
- Sample of size $n = 53$, with ppm.
- Give a confidence interval for true average.
- **Key Question:** ***How much can statistics vary from sample to sample?***
- www.lock5stat.com/statkey

Lange, T., Royals, H. and Connor, L. (2004). Mercury accumulation in largemouth bass (Micropterus salmoides) in a Florida Lake. Archives of Environmental Contamination and Toxicology, 27(4), 466-471.

Bootstrap Confidence Interval



Bootstrap CI

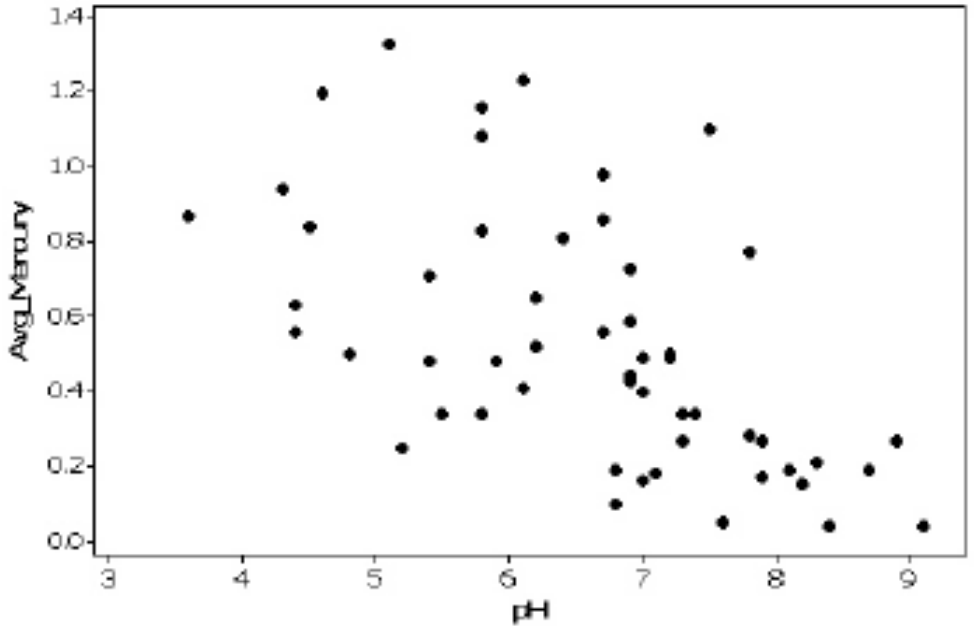
Option 1: Estimate the standard error of the statistic by computing the standard deviation of the bootstrap distribution, and then generate a 95% confidence interval by

$$statistic \pm 2 \times SE$$

Option 2: Generate a P% confidence interval as the range for the middle P% of bootstrap statistics

Mercury and pH in Lakes

- For Florida lakes, what is the correlation between average mercury level (ppm) in fish taken from a lake and acidity (pH) of the lake?

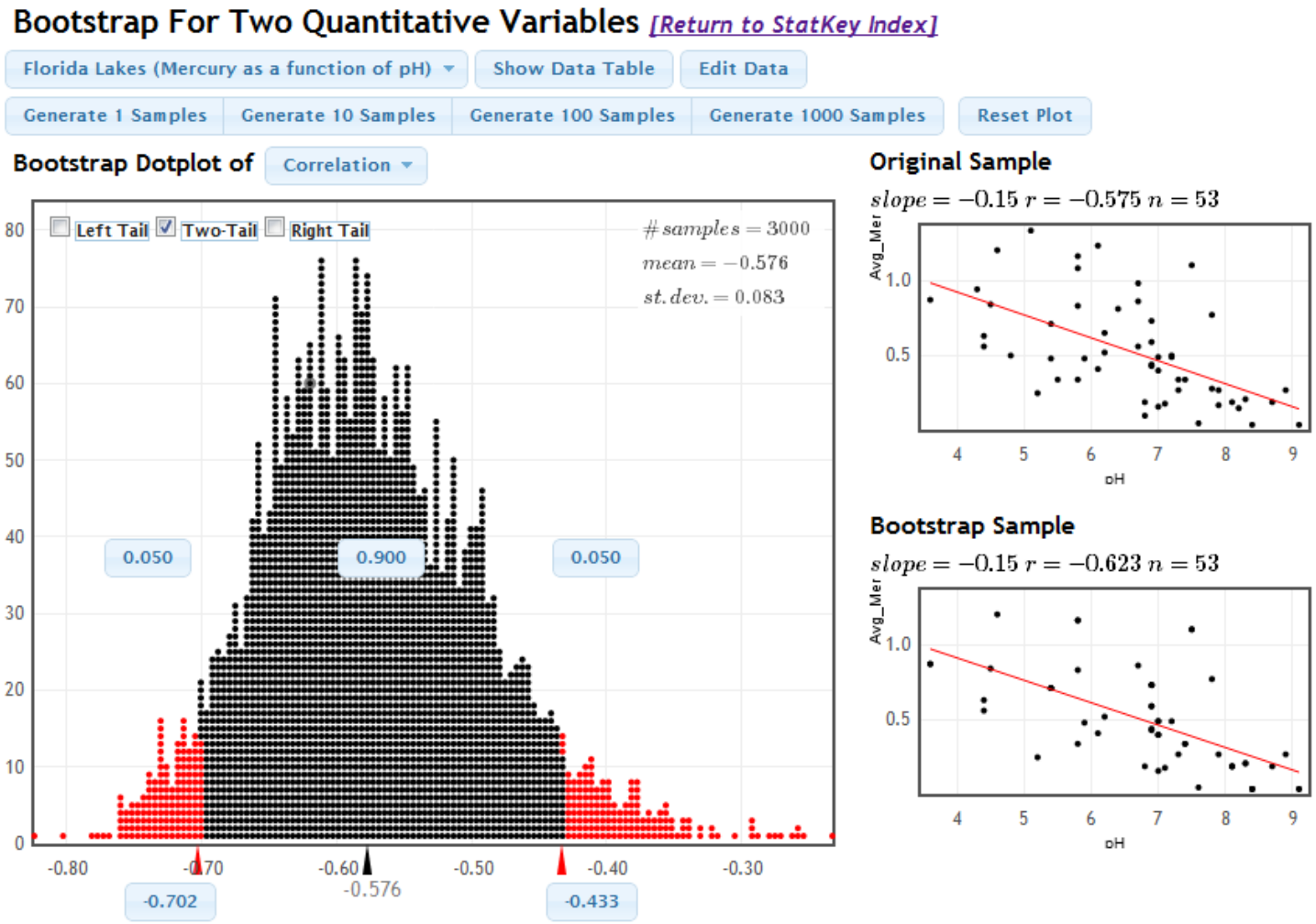


$r = -0.575$

Give a 90% CI for ρ

Lange, Royals, and Connor, Transactions of the American Fisheries Society (1993)

Mercury and pH in Lakes



We are 90% confident that the true correlation between average mercury level and pH of Florida lakes is between -0.702 and -0.433.

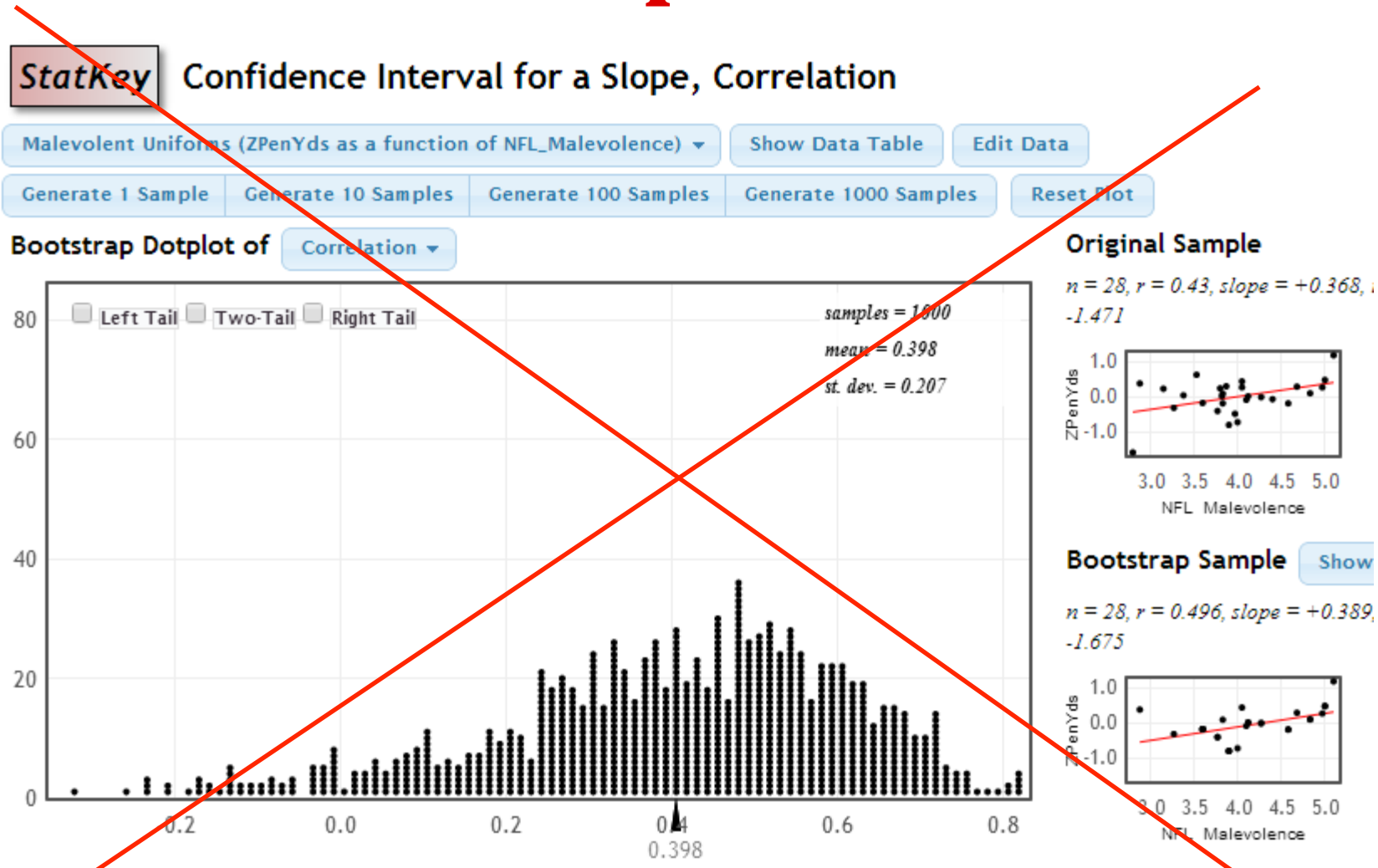
Bootstrap Cautions

- These methods for creating a confidence interval work whenever the bootstrap distribution is smooth and symmetric

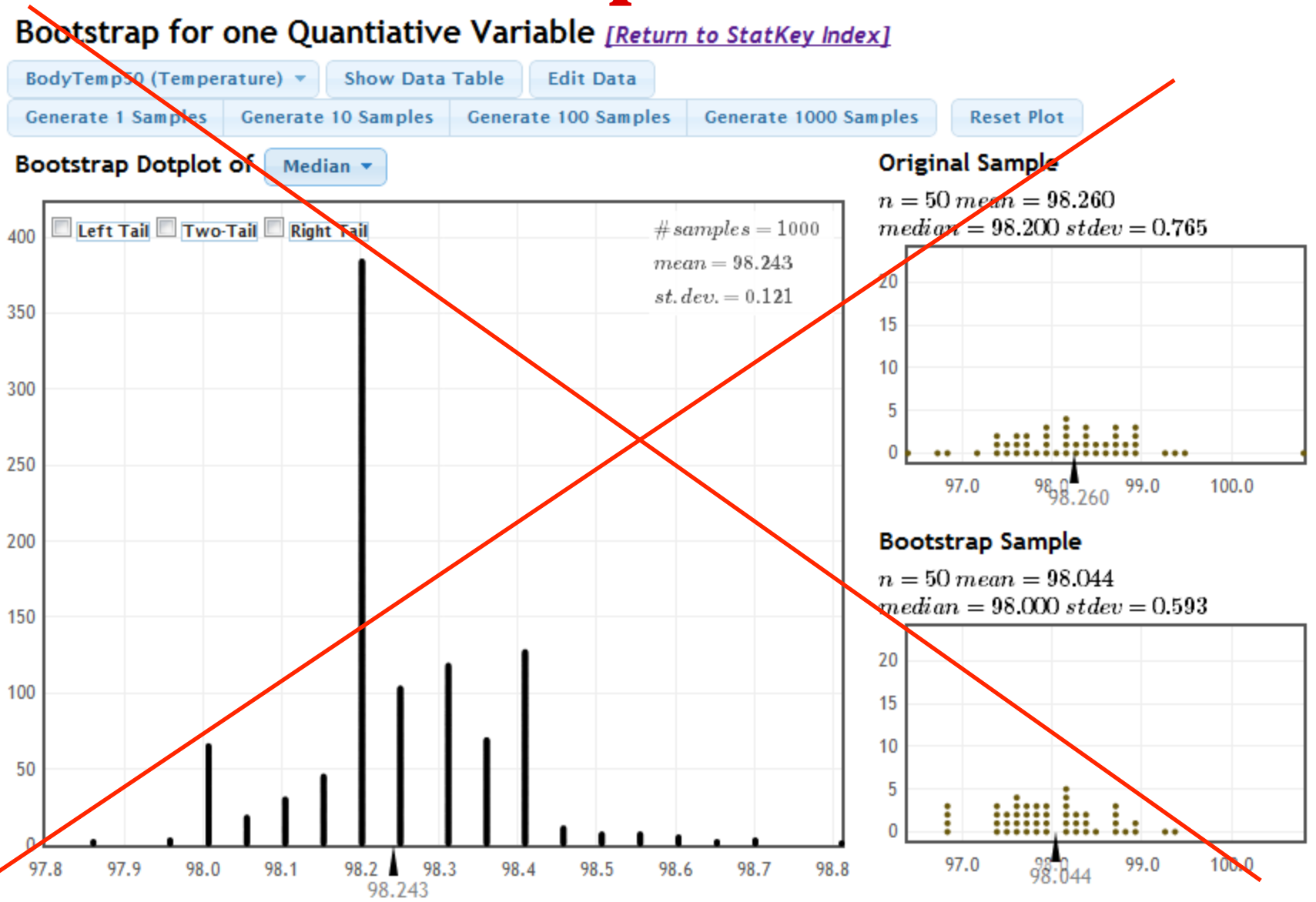
- ALWAYS look at a plot of the bootstrap distribution!

- If the bootstrap distribution is skewed or looks "spiky" with gaps, you will need something more advanced

Bootstrap Cautions



Bootstrap Cautions



Summary

- The standard error of a statistic is the standard deviation of the sample statistic, which can be estimated from a bootstrap distribution
- Confidence intervals can be created using the standard error or the percentiles of a bootstrap distribution
- Confidence intervals can be created this way for any parameter, as long as the bootstrap distribution is approximately symmetric and continuous