Section 3.4

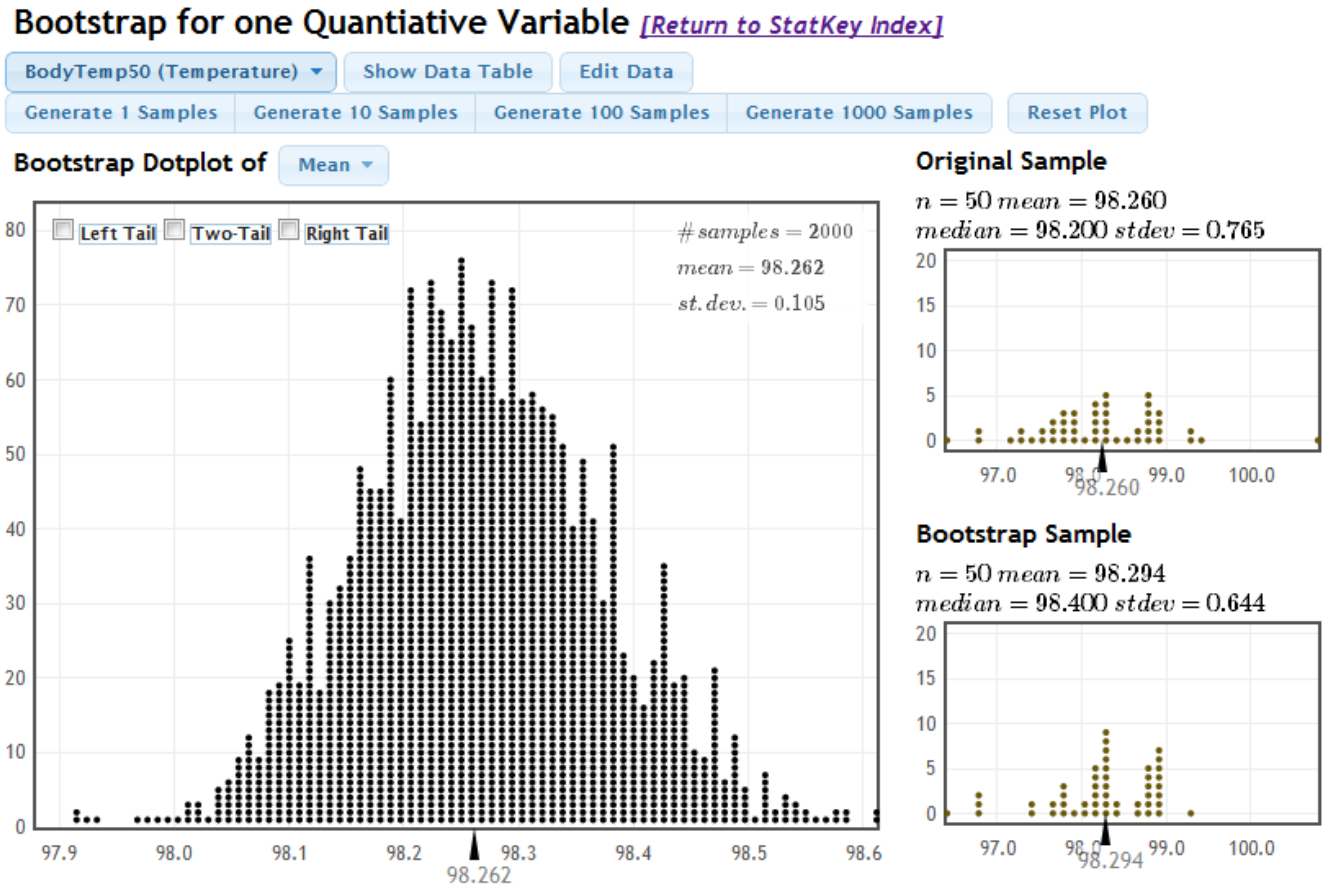
Bootstrap Confidence Intervals using Percentiles

Outline

- Confidence intervals based on bootstrap percentiles
- Different levels of confidence
- Impact of sample size and level of confidence on interval width
- Cautions for bootstrap intervals

Body Temperature

What is the average body temperature of humans?

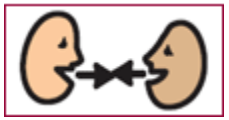


$$98.26 \pm 2 \times 0.105$$
$$= 98.26 \pm 0.21$$
$$= (98.05, 98.47)$$

We are 95% sure that the average body temperature for humans is between 98.05° and 98.47°

98.6° ???

Shoemaker (1996). "What's Normal: Temperature, Gender and Heartrate", Journal of Statistics Education, 4(2).



Other Levels of Confidence

- What if we want to be more than 95% confident?

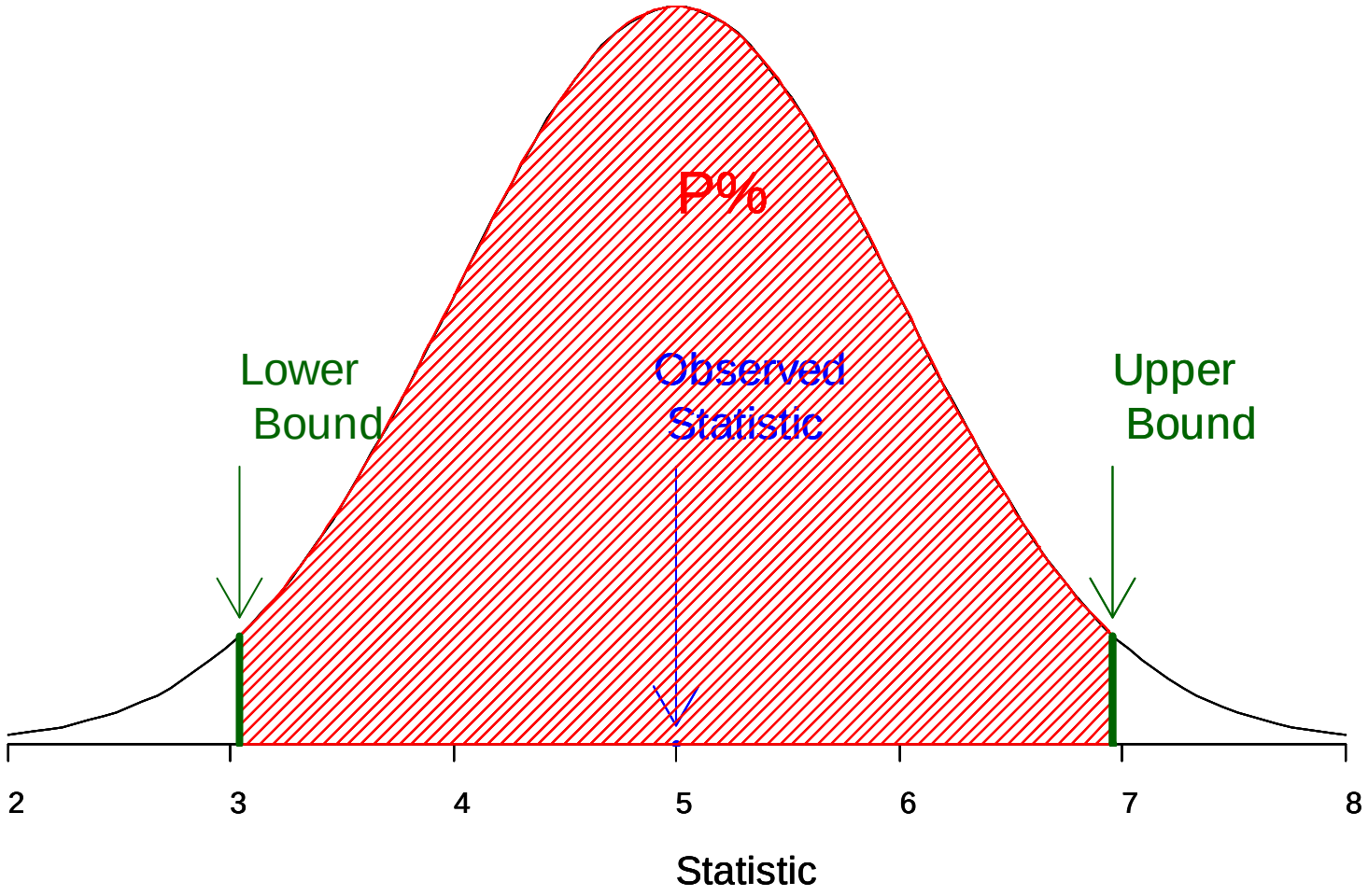
- How might you produce a 99% confidence interval for the average body temperature?

Percentile Method

If the bootstrap distribution is approximately symmetric, we can construct a confidence interval by finding the percentiles in the bootstrap distribution so that the proportion of bootstrap statistics between the percentiles matches the desired confidence level.

Bootstrap Distribution

- For a P% confidence interval:



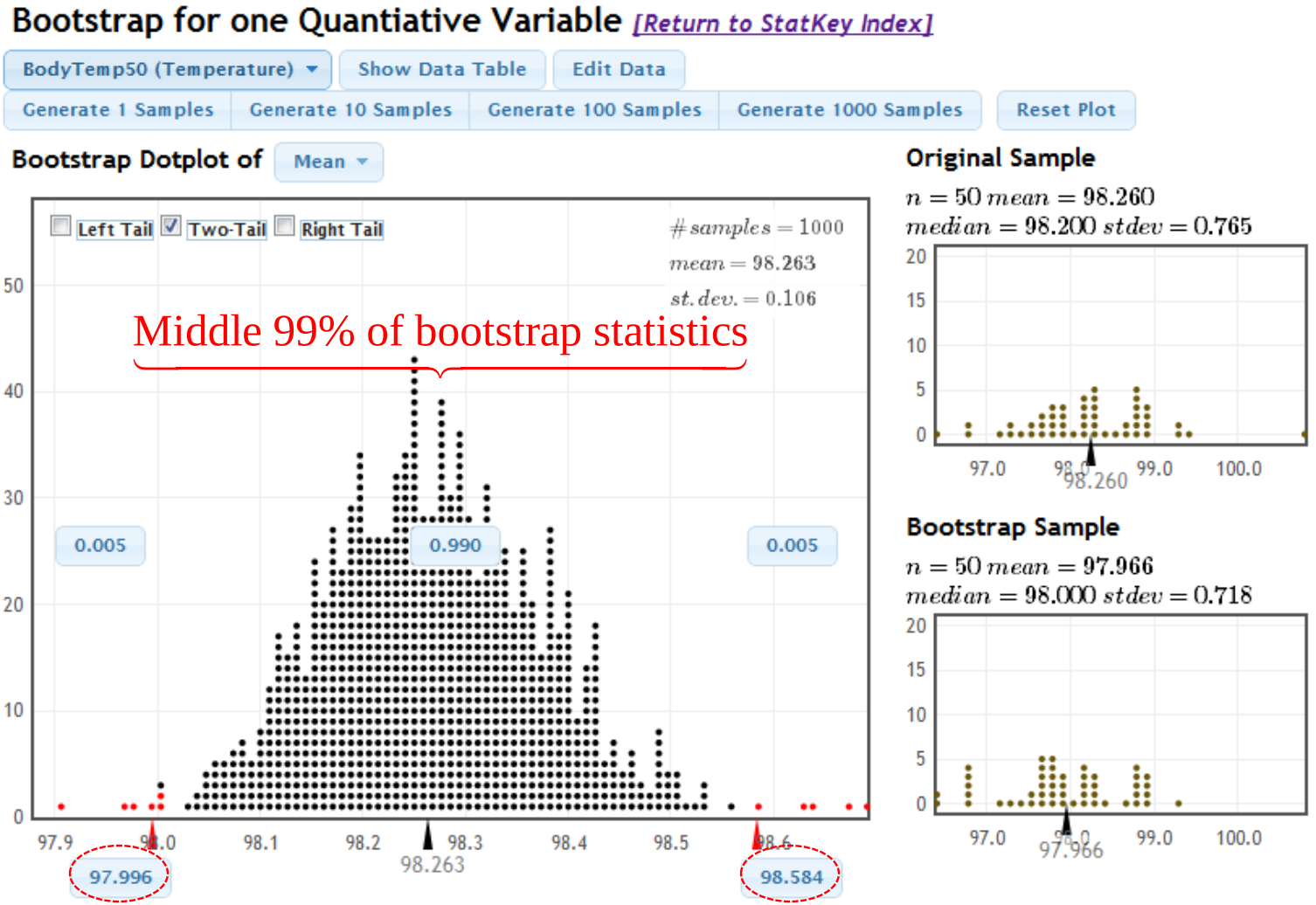
Percentile Method

- For a P% confidence interval, keep the middle P% of bootstrap statistics
- For a 99% confidence interval, keep the middle 99%, leaving 0.5% in each tail.
- The 99% confidence interval would be

(0.5th percentile, 99.5th percentile)

where the percentiles refer to the bootstrap distribution.

Body Temperature



We are 99% sure that the average body temperature is between 98.00° and 98.58°.

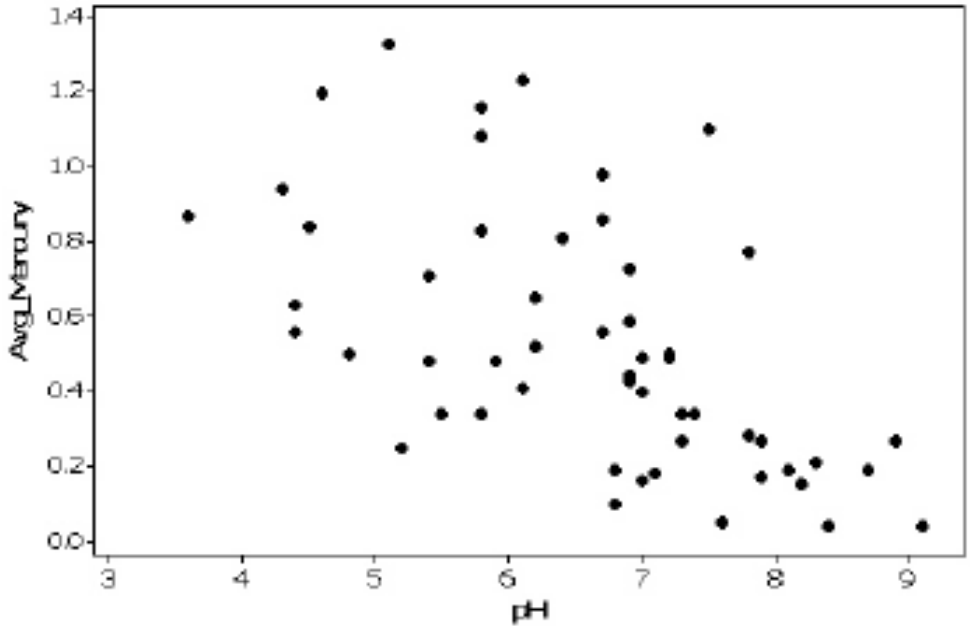
Level of Confidence

Which is wider, a 90% confidence interval or a 95% confidence interval?

A 95% CI contains the middle 95%, which is more than the middle 90%

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?

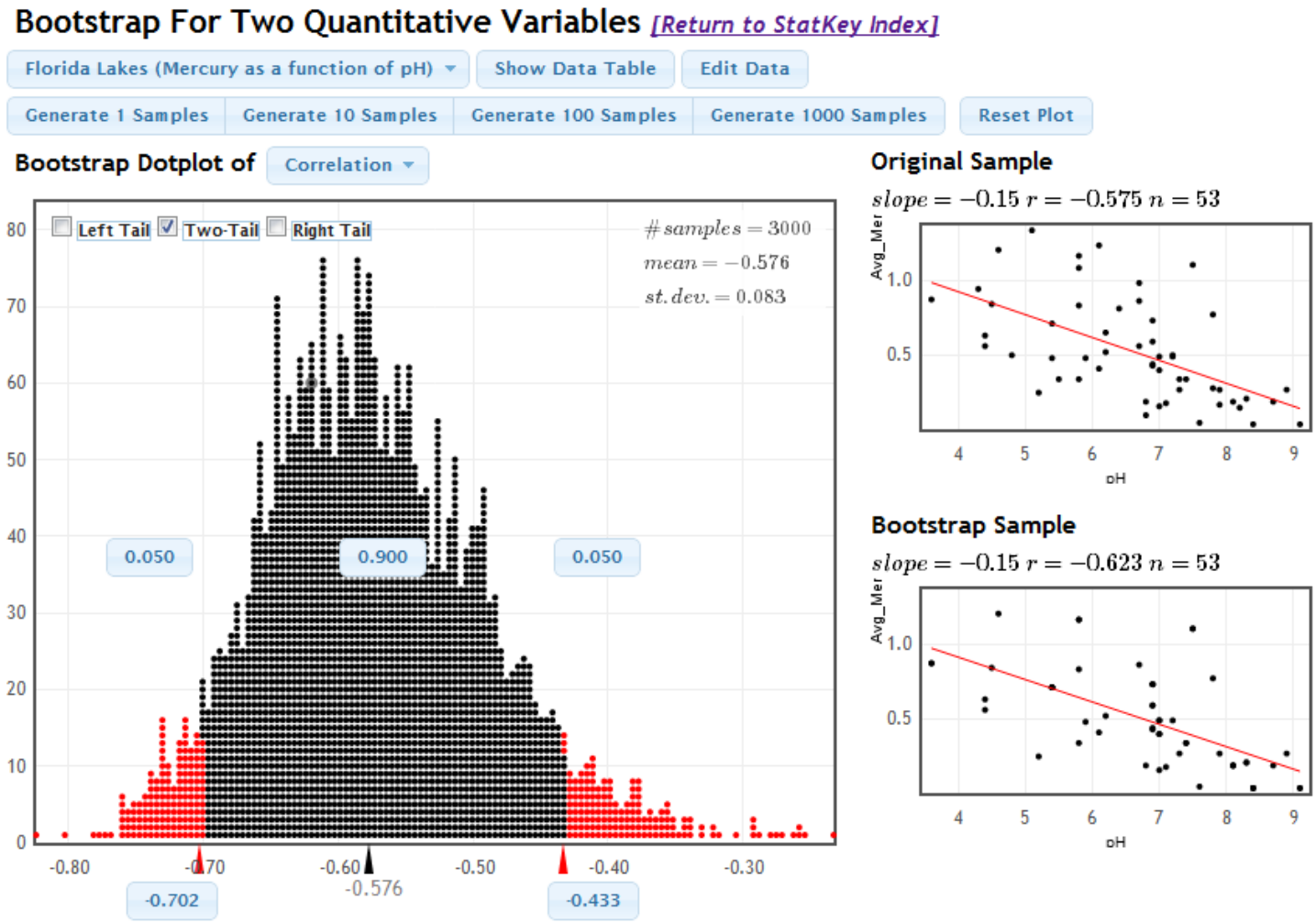


$r = -0.575$

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.

Sample Size

- Remember the effect of sample size?
- The larger the sample size the

(a) wider

(b) narrower

the confidence interval.

The larger the sample size the smaller the variability in the bootstrap distribution, which will make the interval narrower. The larger the sample size, the more precise the estimate.

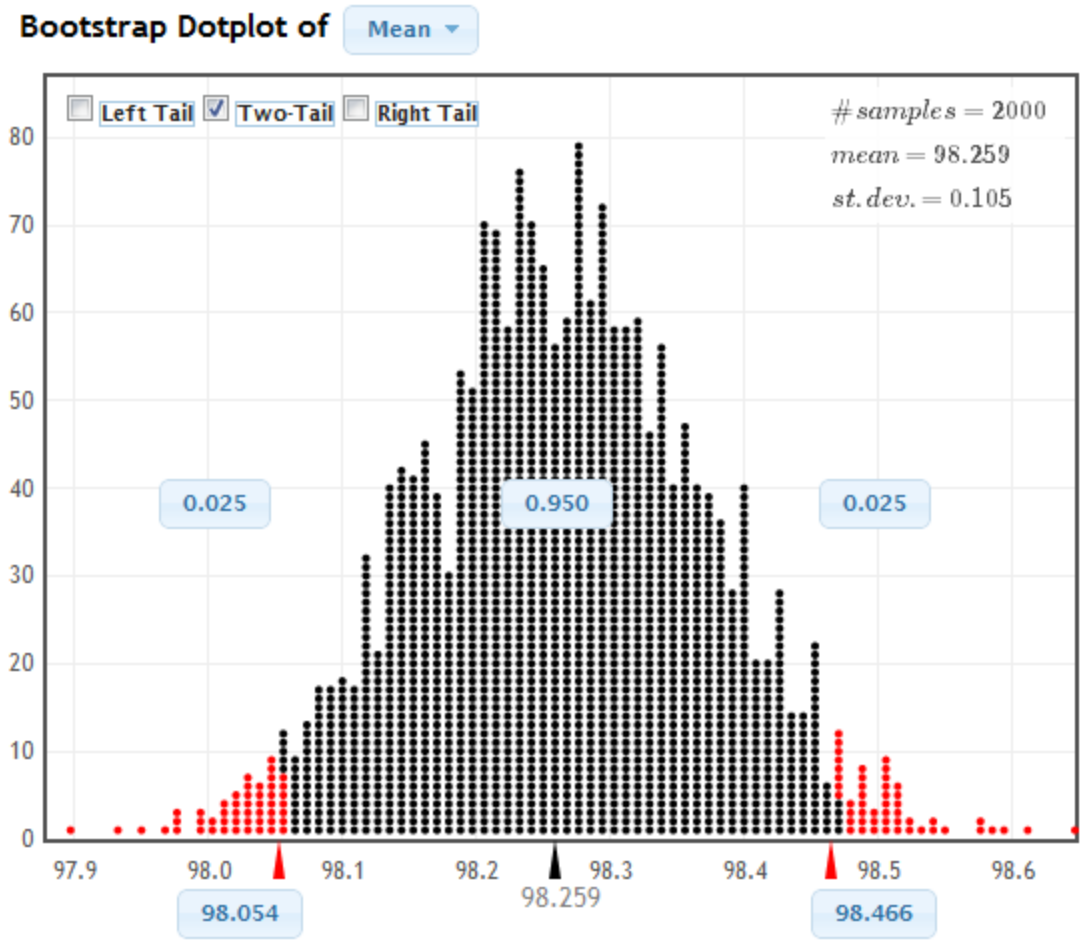
Bootstrap CI

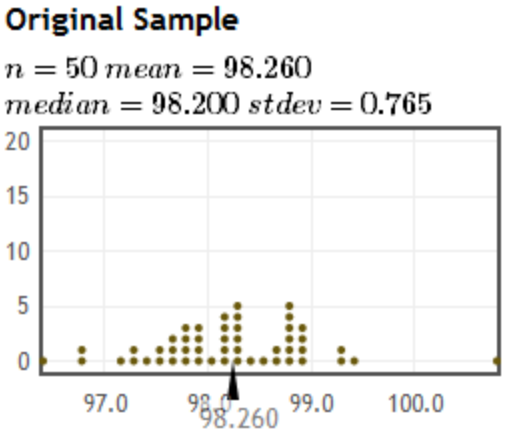
Option 1: Estimate the standard error of the statistic by computing the standard deviation of the bootstrap distribution, and then generate a 95% confidence interval by

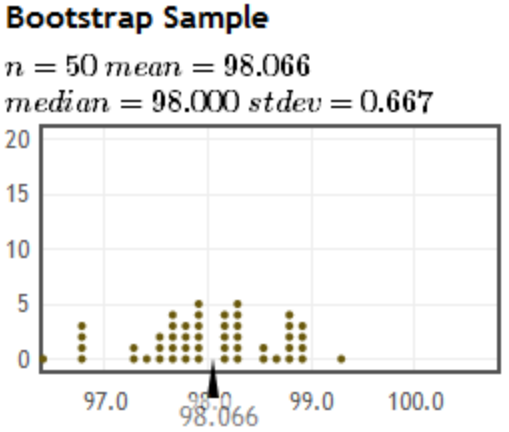
$$statistic \pm 2 \times SE$$

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

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

- Either the standard error method or the percentile method will give similar 95% confidence intervals

- If a level of confidence other than 95% is desired, use the percentile method

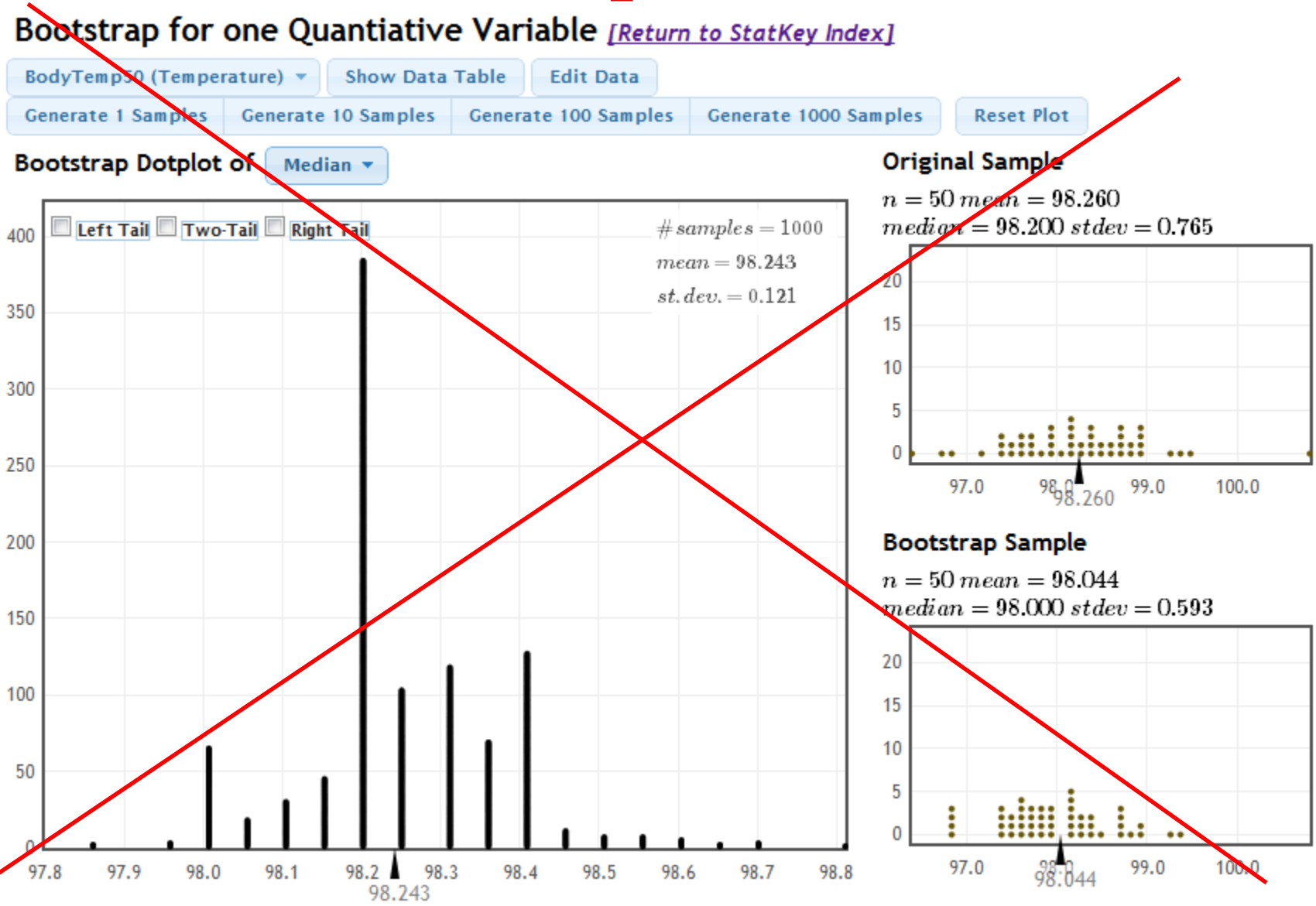
Bootstrap Cautions

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

- ALWAYS look at a plot of the bootstrap distribution!

- If the bootstrap distribution is highly skewed or looks "spiky" with gaps, you will need to go beyond intro stat to create a confidence interval

Bootstrap Cautions



Number of Bootstrap Samples

- When using bootstrapping, you may get a slightly different confidence interval each time. This is fine!

- The more bootstrap samples you use, the more precise your answer will be.

- For the purposes of this class, 1000 bootstrap samples is fine. In real life, you probably want to take 10,000 or even 100,000 bootstrap samples

Summary

- 95% confidence intervals can be created using the standard error or the percentiles of a bootstrap distribution
- For any other desired level of confidence, the percentile method may be used
- A confidence interval from bootstrap percentiles can be created for any parameter, as long as the bootstrap distribution is approximately smooth and symmetric