

Simulation-Based Inference

STA 198: Introduction to Health Data Science

Yue Jiang

July 06, 2020

The following material was used by Yue Jiang during a live lecture.

Without the accompanying oral comments, the text is incomplete as a record of the presentation.

Brief review: interval estimation

In statistical inference, we're interested in making a statement about a population parameter, often by using observed sample statistics.

We saw how the CLT could be used to construct confidence intervals for means and proportions.

Confidence intervals for the mean

In order to construct a confidence interval, we needed to quantify the variability of our sample statistic.

The variability of the sample mean was specified through the standard error, which gives us a measure of how much we expect the *sample mean* to vary from one sample to the next.

Bootstrapping

To construct a CI, we need to quantify the variability of the sample mean, but in practice we often only have one sample available.

How can we use only one sample to tell us how the sample mean varies from sample to sample?

The bootstrap principle

All the information we have about the population is contained in the sample. Let's suppose it's representative.

"The population is to the sample as the sample is to the bootstrap sample" – Fox, 2008

What is a bootstrap sample?

The bootstrap procedure

1. Take a **bootstrap sample**: a random sample taken *with replacement* from the original sample, *of the same size* as the original sample.
2. Calculate the **bootstrap statistic**: the statistic you're interested in (the mean, the median, the correlation, etc.) computed on the bootstrap sample.
3. Repeat steps 1 and 2 many times to create a **bootstrap distribution**.
4. Calculate the bounds of a confidence interval using the quantiles from the bootstrap distribution.

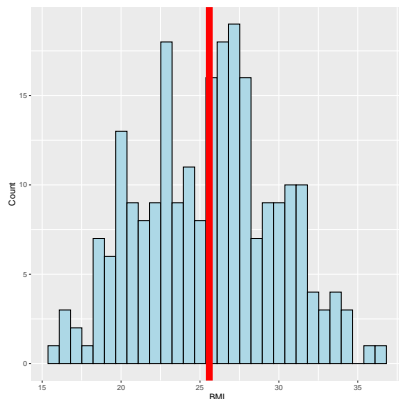
Discussion questions

1. Why do we take a random sample with replacement of the same sample size for each bootstrap sample?
2. What advantage(s) might the bootstrap procedure have over using the CLT?
3. What drawback(s) might the bootstrap procedure have over using the CLT?

Example: bootstrapping a mean

Let's consider the licorice dataset we've used on HW 05 and HW 06. Calculate a 95% confidence interval for the mean BMI among all patients.

Example: bootstrapping a mean



$$\bar{X} = 25.59, s_X = 4.27, n = 235$$
$$\bar{X} \pm t_{0.975, 234}^* \times s_X / \sqrt{n} = (25.04, 26.14)$$

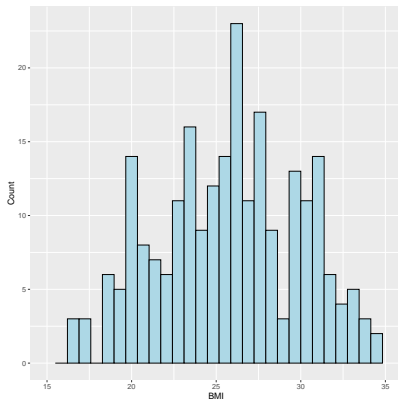
Example: bootstrapping a mean

Using the CLT, we have a 95% confidence interval given by

$$\bar{X} \pm t_{0.975, 234}^* \times s_X / \sqrt{n} = (25.04, 26.14)$$

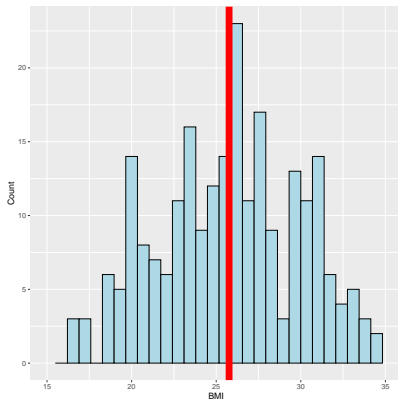
Example: bootstrapping a mean

Step 1: Take a bootstrap sample (random sample with replacement of same size as the original sample):



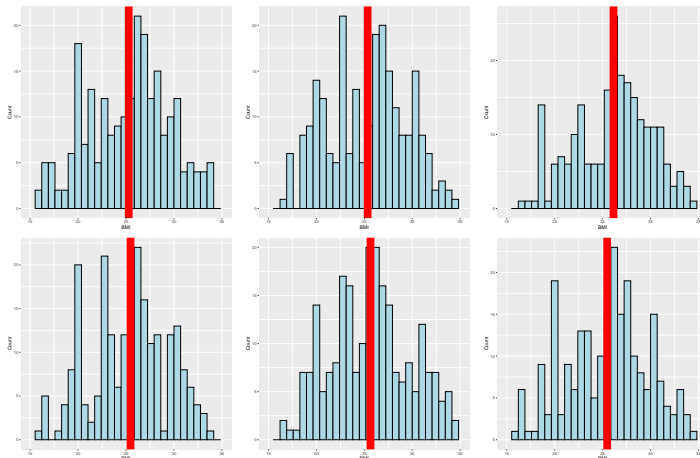
Example: bootstrapping a mean

Step 2: Calculate the bootstrap statistic (in this case, the sample mean of the bootstrap distribution):



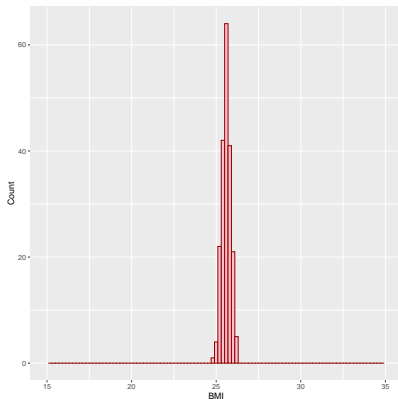
Example: bootstrapping a mean

Step 3: Repeat steps 1 and 2 over and over again:



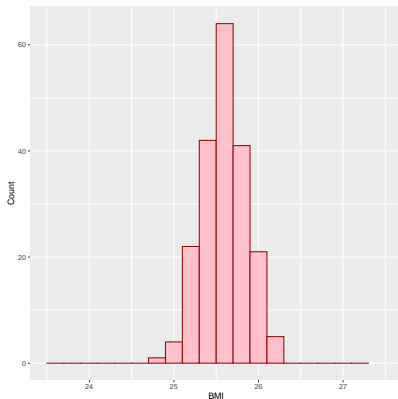
Example: bootstrapping a mean

Step 3: We now have the **bootstrap distribution** of bootstrap sample means. What is each observation in the histogram below?



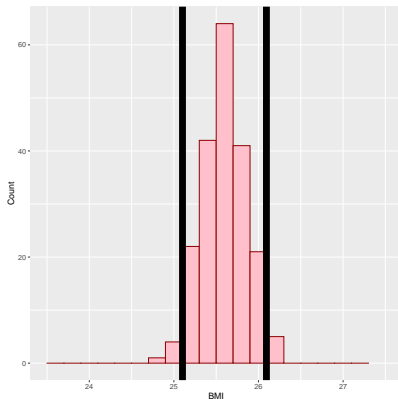
Example: bootstrapping a mean

Step 3: We now have the **bootstrap distribution** of bootstrap sample means. We've zoomed in the plot from the previous slide.



Example: bootstrapping a mean

Step 4: Calculate the bounds of the bootstrap interval by using the 2.5th and 97.5th percentile of the bootstrap distribution



Example: bootstrapping a mean

We used 200 bootstrap samples to construct the bootstrap interval of **(25.10, 26.09)**.

Compare this to the CLT interval of (25.04, 26.14).

Discussion

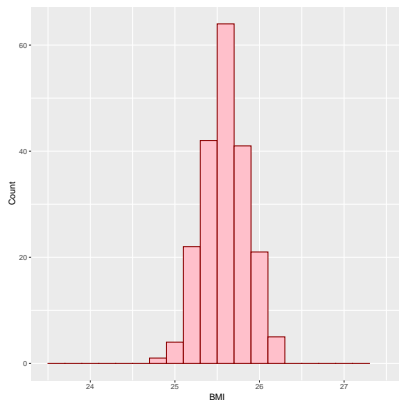
- ▶ The more bootstrap replicates we use to generate the bootstrap distribution, the better, but this may become computationally prohibitive.
- ▶ Bootstrap confidence intervals are very flexible and don't require many assumptions on the underlying data
- ▶ However, they should **not** be constructed when the original sample size is very small (<5 or so)
- ▶ Bootstrap intervals are also inappropriate for estimating the maximum or minimum (unfortunately, we can't easily create CIs for these statistics!)

Hypothesis tests using bootstrap

We can construct a bootstrap distribution to tell us the variability of the sample statistic (in the previous case, the sample mean).

Here, the bootstrap distribution gives us the distribution of the sample mean.

The bootstrap distribution



The hypothesis testing paradigm

Remember, in hypothesis testing, we always assume that the null hypothesis is true, and calculate the conditional probability of seeing our observed data under this assumption.

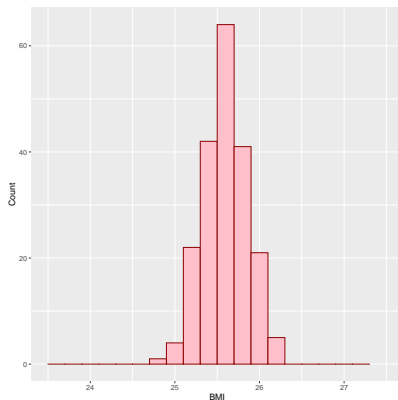
Suppose we're interested in testing whether $\mu = 25$. How can we test this hypothesis using our bootstrap distribution?

Randomization tests for means

1. Construct a bootstrap distribution for the sample mean.
2. *Shift* the distribution to be consistent with the null hypothesis.
3. Calculate the proportion of simulations that yield a sample mean as extreme or more than our originally observed sample mean.

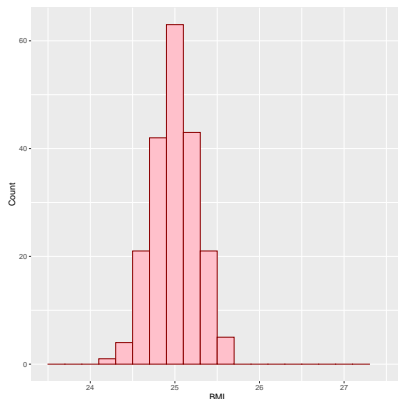
The bootstrap distribution

Step 1: Construct a bootstrap distribution for the sample statistics (in this case, the sample mean)



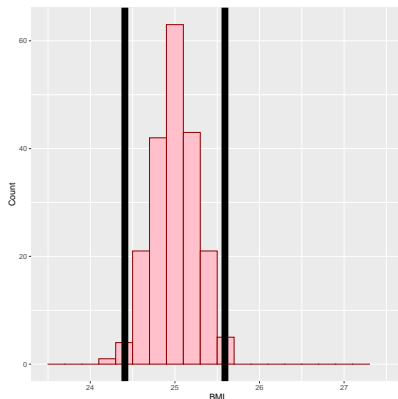
The bootstrap distribution

Step 2: Shift the distribution to be consistent with the null hypothesis



The bootstrap distribution

Step 3: Calculate the proportion of simulations with a sample mean as extreme or more than the *original* sample mean of 25.59



Hypothesis test results

In this case, 5 / 200 (2.5%) of the observations were in this region.
Compare this with our calculated p-value of 0.035 from the t-test.