Estimating Sampling Variability Through Bootstrapping
Supplement to Sections 2.2 and 3.3: Inference for a Single Mean

Stacey Hancock

1 The Big Picture: 3S Strategy

As discussed in our textbook, simulation-based hypothesis testing can be thought of as three main steps:

1. **Statistic**: Calculate an observed statistic—a number that summarizes the data.

2. **Simulate**: Create a simulated distribution of potential statistics we could have seen if the null hypothesis was true.

3. **Strength of evidence**: Compare the observed statistic to the null distribution of simulated statistics and assess the strength of evidence against the null hypothesis by quantifying how far the observed statistic falls from the center of the null distribution.

When our statistic is a sample proportion, we can simulate a distribution of sample proportions by spinning a spinner n times and measuring the proportion of spins that landed in the shaded area, where the shaded proportion of the spinner is equal to our hypothesized true proportion, π_0.

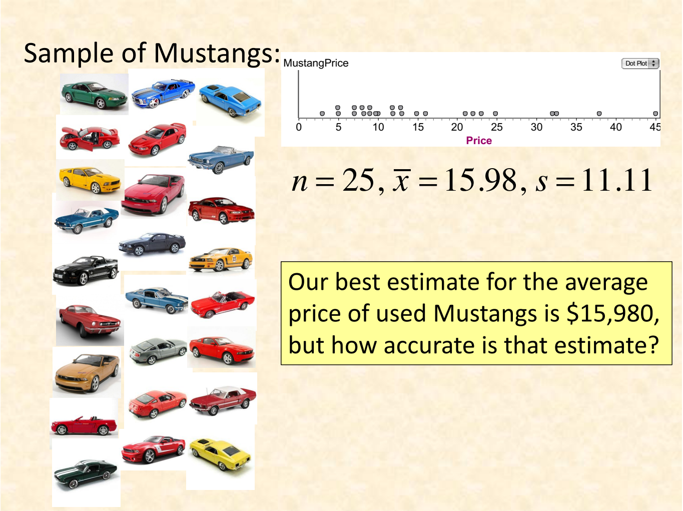
However, for the one mean scenario, where we measure a quantitative variable on each observational unit, we need to be more creative about the "Simulate" step—we can't just flip a coin, draw cards, or spin a spinner. We need to create a made-up population that has a mean equal to the null value and has variability similar to the sample, then simulate random samples of size n from this population. But how can we create such a population?

One option, described in this reading, shifts the original data to be centered at the null value, then samples *with replacement* from the original data n times. Simulating samples by sampling with replacement (or "resampling") from the original sample, then using these samples to estimate sampling variability of a statistic, is called **bootstrapping**.

2 Bootstrapping

What is the average price of a used Mustang car? To answer this question, you collect a random sample of $n = 25$ Mustangs from a website (autotrader.com) and record the price (in $1,000's) for each car (See Figure 1). How can we use this sample to estimate the average price for *all* used Mustang cars? The sample mean of $15,980 provides a *point estimate* for the parameter $\mu =$"mean price of all used Mustang cars", but how close is $15,980 to μ? If we have a representative sample of the population, we can imagine a made-up population of all used Mustang cars that is comprised of many, many copies of the original sample. We can simulate from this made-up population by *resampling* 25 cars, *with replacement*, from our sample of 25 cars. For example, one of our bootstrap resamples may look like the sample shown in Figure 2. Note that some of the cars in the original sample were not selected for the resample, but some of the cars were selected more than once.

Figure 1: Original sample of 25 used Mustang cars (Lock Morgan, 2014).



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?

Figure 2: Bootstrap resample of 25 used Mustang cars (Lock Morgan, 2014).

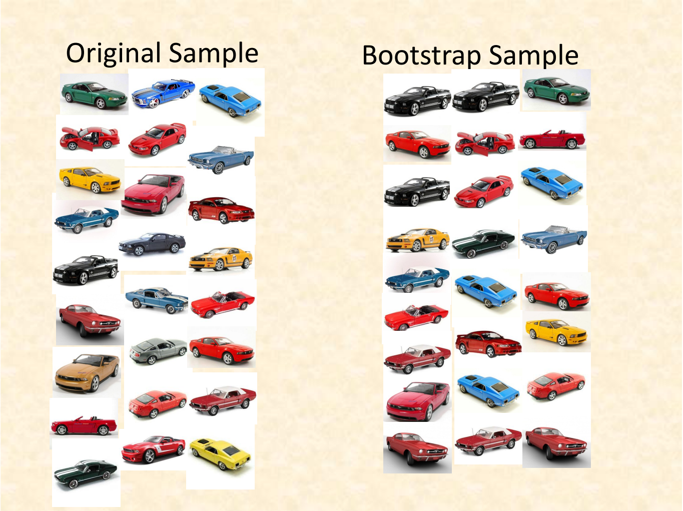
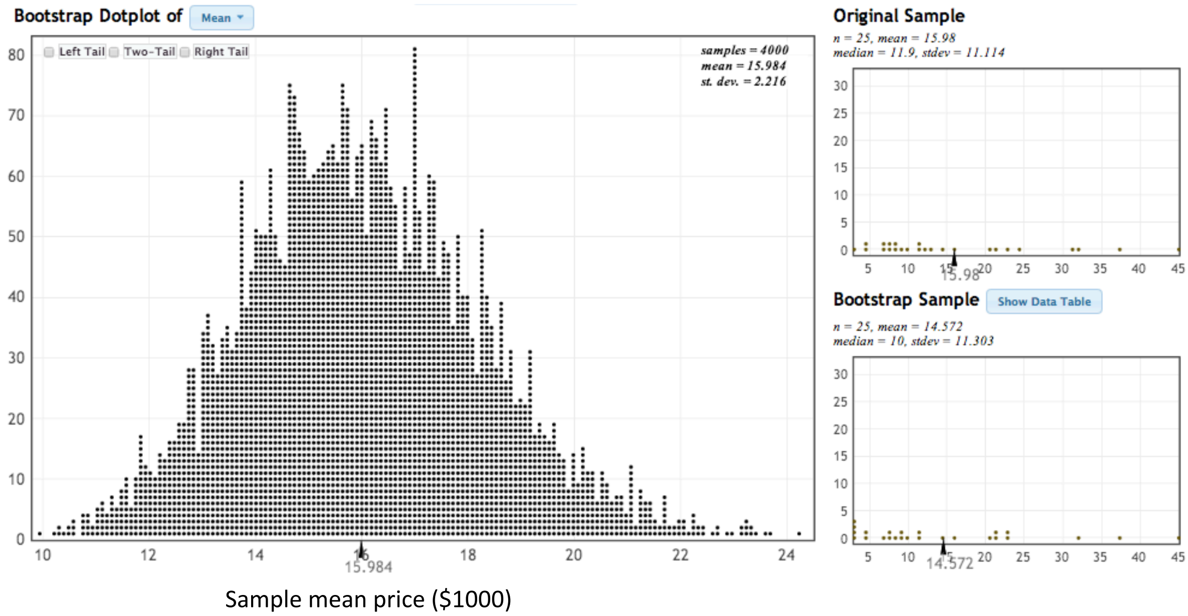


Figure 3 plots the sample means from 4000 bootstrap resamples of the original data. Now, we can use the standard deviation of these simulated sample means (2.216) as our measure of how much sample means vary from sample to sample. We can calculate an approximate 95% confidence interval for μ using the 2SD Method by: $15.98 \pm 2 \times 2.216 = 15.98 \pm 4.432 = (11.55, 20.41)$. That is, we are 95% confident that the mean price for *all* used Mustang cars is between $11,550 and $20,410.

Figure 3: Sample means from 4000 resamples of 25 used Mustang cars from the original sample. The top right dotplot displays the original sample. The bottom right dotplot displays the last resample. Note that the dotplot of sample means is centered close to the original sample mean of $\bar{x} = 15.98$. Why? (Lock Morgan, 2014).



3 Bootstrapping Null Distributions

Consider the following study on arsenic poisoning:

> Symptoms of low–level arsenic poisoning include headaches, confusion, severe diarrhea and drowsiness. When the poisoning becomes acute, symptoms include vomiting, blood in the urine, hair loss, convulsions, and even death. A 2007 study by Peter Ravenscroft found that over 137 million people in more than 70 countries are probably affected by arsenic poisoning from drinking water.[1] Scientists can assay toe nail clippings to measure a person's arsenic level in parts per million (ppm). They did this assay on 19 randomly selected individuals who drink from private wells in New Hampshire (data displayed in Table 1 and summarized in Figure 4). An arsenic level greater than 0.150 ppm is considered hazardous. The research question is, "Is there evidence that people drinking the ground water in New Hampshire are suffering from arsenic poisoning?"

The research question leads to the following hypotheses:

$$H_0 : \mu = 0.15 \qquad \text{versus} \qquad H_a : \mu > 0.15$$

where μ is the true mean arsenic level of all individuals who drink from private wells in New Hampshire.

[1] Ravenscroft, P. (2007). The global dimensions of arsenic pollution of groundwater. *Tropical Agriculture Association*, **3**.

Table 1: Arsenic levels measured via to nail clipping assay on 19 randomly selected individuals who drink from private wells in New Hampshire.

0.119	0.118	0.099	0.118	0.275	0.358	0.080	0.158	0.310	0.105
0.073	0.832	0.517	0.851	0.269	0.433	0.141	0.135	0.175	

Figure 4: Histogram, dotplot and boxplot of distribution of arsenic level measurements.

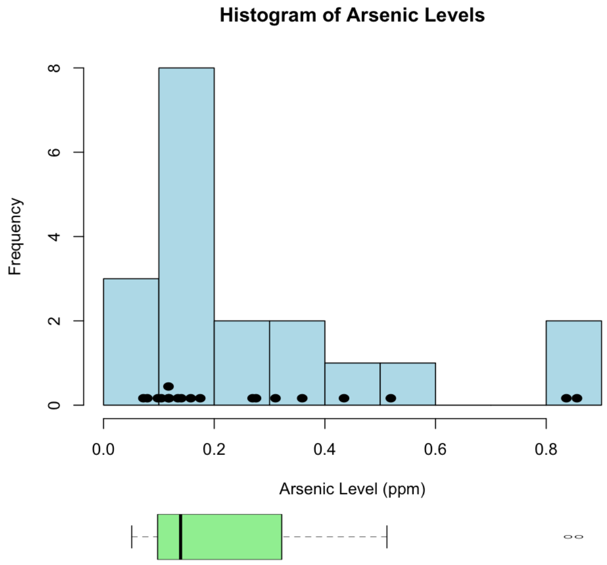
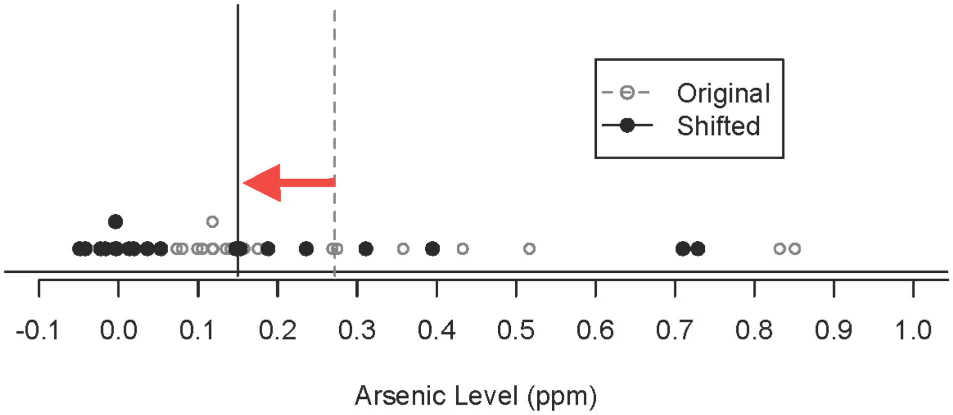


Figure 5: Shifted data for arsenic study obtained by subtracting 0.122 (the distance the sample mean lies above the null value) from each observation. The shifted data now has a sample mean equal to the null value, 0.15.

0.119 – 0.122	0.118 – 0.122	...	...	...	...	...	...	...	0.105 – 0.122
0.073 – 0.122	...	...	...	...	...	...	...	0.175 – 0.122	

⇩

-0.003	-0.004	-0.023	-0.004	0.153	0.236	-0.042	0.036	0.188	-0.017
-0.049	0.710	0.395	0.729	0.147	0.311	0.019	0.013	0.053	



We can now apply the 3S strategy to evaluate the strength of evidence for the research hypothesis, that the true mean arsenic level is greater than 0.15 ppm:

1. **Statistic**: The mean arsenic in our sample of 19 individuals is $\bar{x} = 0.272$ ppm (with standard deviation $s = 0.2365$ ppm).

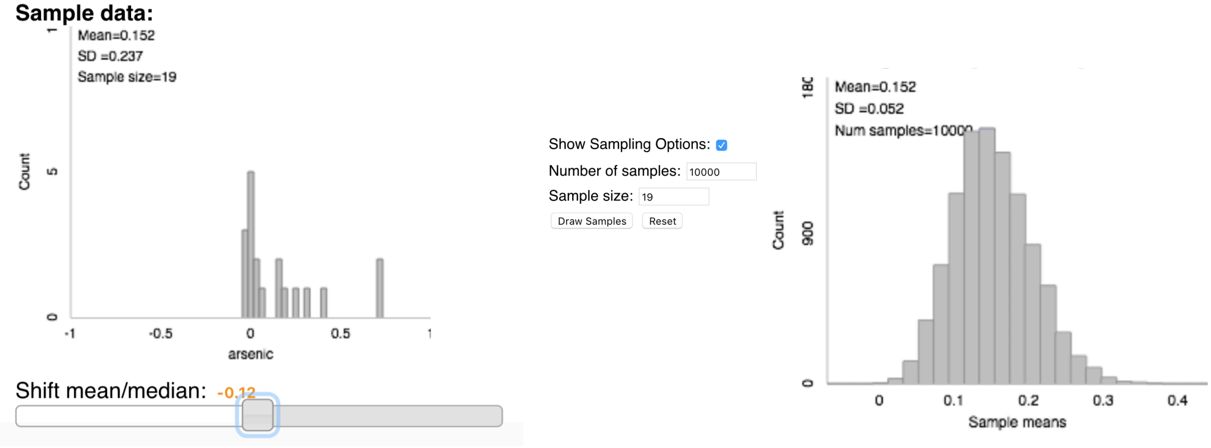
2. **Simulate**: Our goal is to simulate thousands of sample means under the assumption that $\mu = 0.15$. We could take samples by bootstrapping the original data, but then our distribution of sample means would not be centered at the null value of 0.15. (Where would it be centered?[2]) Instead, we first need to *shift the data* so that it is centered around the null value; then take bootstrap resamples.

   How could we change the values so that they are consistent with the null, $\mu = 0.15$? Our sample mean is 0.122 above the null value $(0.272 - 0.15)$, so we could shift the data by subtracting 0.122 from each arsenic level, shown in Figure 5.

   Next, using the **Bootstrapping One Mean** applet, we will resample (with replacement) from the shifted data. This simulation is shown in Figure 6.

[2] A simulated distribution of sample means by resampling from the original data should be centered at the original sample mean.

Figure 6: Bootstrapping One Mean applet: Shifted data (left) and bootstrap distribution of simulated sample means from resampling the shifted data (right).



Sample data:

Mean=0.152
SD =0.237
Sample size=19

Shift mean/median: -0.12

Show Sampling Options: ☑
Number of samples: 10000
Sample size: 19

Mean=0.152
SD =0.052
Num samples=10000

3. **Strength of evidence** – Now that we have a null distribution of simulated sample means, we can assess strength of evidence against H_0 by finding our sample mean $\bar{x} = 0.272$ on the null distribution. Only 176 out of 10,000 simulated sample means were as larger or larger than 0.272, giving us a p-value of 0.0176 (shown in Figure 7). This provides strong evidence against H_0. Thus, we have strong evidence that the mean arsenic level of all individuals who drink from private wells in New Hampshire is greater than 0.15 ppm.

Practice: Try working through this example on your own; copy and paste data from the Stat 216 course webpage into the applet.

4 Summary

When testing $H_0 : \mu = \mu_0$, we can simulate a null distribution of sample means by bootstrapping from the shifted sample. General steps are as follows. Suppose our data are $x_1, x_2, \ldots, x_n$ with sample mean $\bar{x} = (x_1 + \cdots + x_n)/n$.

1. Calculate how far the null value is from the sample mean: $d = \mu_0 - \bar{x}$.

2. Add the value d to each value in the original data to create a sample of shifted data:

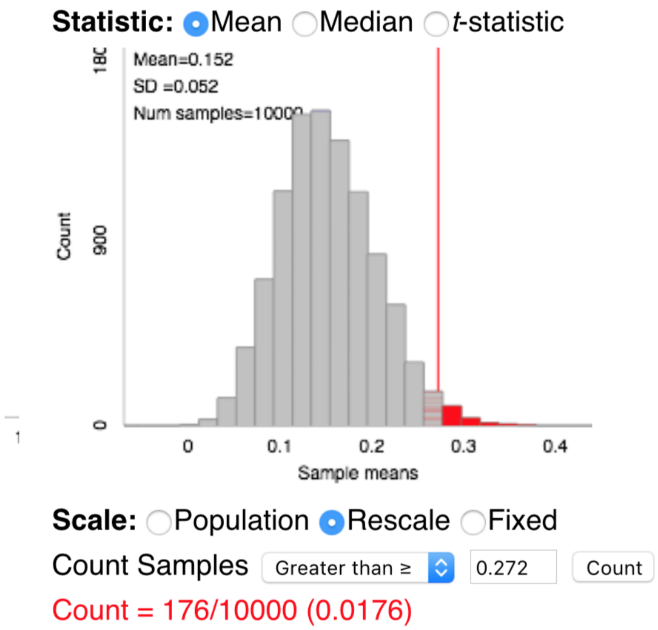
$$x_1 + d, \;\; x_2 + d, \ldots, \;\; x_n + d$$

   (Note that if d is a negative number, you will be subtracting $|d|$ from each value in the sample.)

3. Generate a bootstrap resample distribution of sample means from the shifted data. This is our null distribution of sample means.

4. Calculate the p-value by finding the proportion of resampled sample means in the null distribution that are as or more extreme than $\bar{x}$.

Figure 7: P-value calculation using the bootstrap null distribution from the Bootstrapping One Mean applet.



Statistic: ● Mean ○ Median ○ *t*-statistic

Scale: ○ Population ● Rescale ○ Fixed

Count Samples Greater than ≥ 0.272 Count

Count = 176/10000 (0.0176)

5 Extra Resources

For an additional explanation of the bootstrapping method, watch the first five minutes of this video:

http://www.lock5stat.com/videos/BootstrapIntro.mp4

6 References

- Lock Morgan, K. (2014). Estimating Parameters - Bootstrap Confidence Intervals. Workshop at the International Conference on Teaching Statistics.
- Lock, R. (2017). Lock5Data: Datasets for “Statistics: UnLocking the Power of Data”. R package version 2.8. https://CRAN.R-project.org/package=Lock5Data
- Penn State Stat 555: Statistical Analysis of Genomics Data – 15.3 - Bootstrapping. https://onlinecourses.science.psu.edu/stat555/node/119. Accessed Sep 18, 2017.
- Robison-Cox, J. (2016) *Stat 216 Course Pack Fall 2016: Activities and Notes.* License: Creative Commons BY-SA 3.0.
- Tintle, Nathan, Beth Chance, George Cobb, Allan Rossman, Soma Roy, Todd Swanson, Jill VanderStoep. *Introduction to Statistical Investigations for Montana State University.* Wiley Custom Select, 2016-11-15. VitalBook file.