# Bootstrapping for the Mean with Minitab Express

When Do We Need Bootstrapping?

Gathering information about an entire population usually isn't an option. Instead, we take a random sample from the population and use inferential statistics to identify a range of likely values for parameters such as the population mean. With traditional inferential statistics, we typically use a t-distribution to construct this range of values, also referred to as a confidence interval. However, this approach assumes that the data follow a normal distribution. So what can we do if this assumption does not hold true? We can use bootstrapping.

How Does Bootstrapping Work?

Our sample data may be the best—in fact, the only—information we have about what the actual population looks like. Since most randomly chosen samples provide a good estimate of the distribution of the population they come from, bootstrapping randomly samples an observation from the sample itself, places that observation back into the pool of sample data, and repeats this process over and over again.
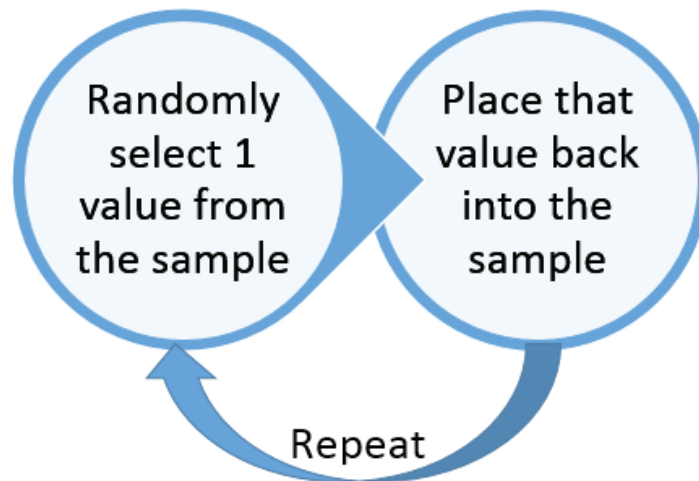
Bootstrapping treats the many samples of data as a surrogate population to approximate the sampling distribution of a statistic, such as the mean, and to calculate a confidence interval for that statistic.

Let's walk through how to use Minitab Express to create a thousand "bootstrap samples" by sampling, with replacement, from the sample data. We will then create a histogram of the bootstrap sample means to evaluate the bootstrap distribution and calculate a confidence interval for the mean.

Example

We collect a sample of data from a population we know nothing about by asking 5 randomly selected college students for their birthdate. We chose a small sample size of 5 for this example so you can easily type the data into Minitab Express and follow along. Here are their ages:

<div align="center">

17.17, 18.33, 21.08, 19.92, 19.58

</div>

We want to calculate a 95% confidence interval for the mean.

Using the 5 data points above, we sample with replacement (or "resample") to create 10 bootstrap samples, each of size 5—the same size as the original sample. Because we are sampling with replacement, the bootstrap samples will not be identical to the original sample.
We then calculate the mean for each of the 10 bootstrap samples.

Mean of sample 1: **18.416**	Mean of sample 6: **19.216**
Mean of sample 2: **20.248**	Mean of sample 7: **17.952**
Mean of sample 3: **19.980**	Mean of sample 8: **20.066**
Mean of sample 4: **18.434**	Mean of sample 9: **19.284**
Mean of sample 5: **20.780**	Mean of sample 10: **19.748**

We created and computed means for these 10 bootstrap samples above to illustrate the resampling, but the bootstrapping method requires many more samples. We can use Minitab Express to create 1,000 bootstrap samples, each of size 5, and calculate their corresponding means.

We can then use these 1000 sample means to compute a 95% confidence interval for the population mean, using the 2.5th and 97.5th percentiles (since 97.5% - 2.5% = 95%):
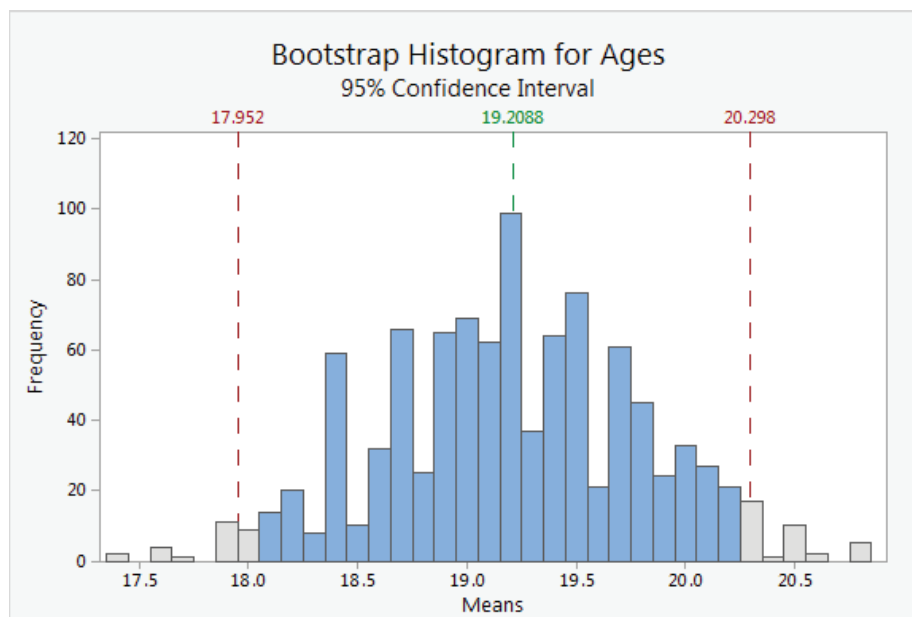
The graph indicates we can be 95% confident that the population's average age is between 17.952 and 20.298 years old.

Step-by-Step in Minitab Express

Follow these steps to conduct this analysis using Minitab Express 1.5.

1

Create the Ages column

C1
Ages
17.17
18.33
21.08
19.92
19.58

2

Open the bootstrapping for 1-sample mean dialog box

Mac: Statistics > Resampling > Bootstrapping for 1-Sample Mean
PC: STATISTICS > Resampling > Bootstrapping > 1-Sample Mean
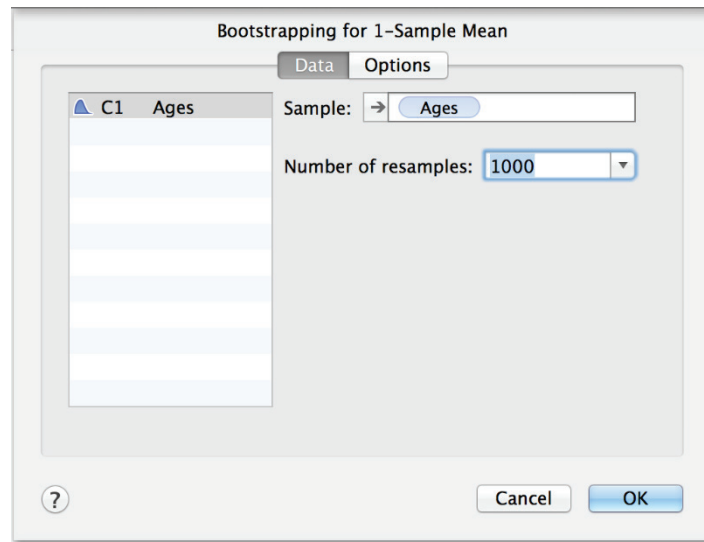
3

In Sample, enter Ages.

(Optional) Click **Options** and select **Store sample means in a column**.

(Optional) Click **Options** and for **Base**, enter 2 to randomly sample the same values as this article.

4

Click OK.

With bootstrapping and Minitab Express, we can easily create repeated random samples to estimate a range of likely values for the mean and draw conclusions about our population.

Minitab