

## Bootstrapping a Mean

If we do not have all of the data for the population of interest, we can still estimate the mean of that population. To do so we use the **bootstrap method**.

This method builds up a picture of what the population could look like, based on a sample of size  $n$ . We begin by taking a sample of size  $n$  from our sample, using sampling with replacement. The mean of that first sample is computed. The process is then repeated a large number of times, creating a distribution of these sample means.

- If you have a sample of size  $n$ , draw a sample of size  $n$  from those values using sampling with replacement.
- Compute the mean of that sample.
- Repeat the process many times. (We will repeat the process 10,000 times.)
- Find the 2.5<sup>th</sup> percentile and the 97.5<sup>th</sup> percentile for these 10,000 means. These values separate the lower 2.5% and upper 2.5% of the means from the rest. In other words, these two values separate the middle 95% of all means from the rest.
- The same procedure can be used for estimating a population median.

These critical values will create a **95% confidence interval** for the population mean or median. 95% of the intervals created in this fashion will contain the unknown population mean or median. We can then evaluate claims about population means or medians as plausible (it is contained by the interval) or unusual (it is outside the interval).

## Using StatCrunch to Bootstrap a Population Parameter

1. Load the data into a StatCrunch spreadsheet.
2. Press Applets; highlight Resampling, then Bootstrap a statistic. Be sure the "From data table:" radio button is selected. Select the column that contains the data. Select "Mean" (or "Median") under the Statistic drop-down menu. Click Compute!.
3. Click 1000 times ten different times to generate 10,000 Bootstrap samples with replacement and determine the mean of each sample.
4. The needed percentiles will be labeled as 2.5th perc. and 97.5th perc.

## Problem

The mean wait time at a drive-thru of a fast-food restaurant is 90 seconds. A new system is implemented to reduce the wait time.

Here are the wait times (in seconds) for 10 randomly selected drive-thru customers:

109      67      58      76      65      80      96      86      71      72

*(Data can be found as "AMATYC 2020" in the "Woodbury Math 21" StatCrunch group.)*

a) Generate 10,000 bootstrap samples and compute their means. Find the 2.5<sup>th</sup> percentile and 97.5<sup>th</sup> percentile for these means.

b) Based on your results from part a), does it seem plausible that the new system has reduced the waiting time for drive-thru customers?