Bootstrapping

ECON 5030

Will Achi, Peri Brimley, Sayandeep Paul

What is Bootstrapping?

- a statistical procedure that resamples a single dataset to create many simulated samples (sampling with replacement)

Different Types of Bootstrapping[1]

Parametric bootstrapping

- Resampling from a known distribution in order to compare sample results to what we know the distribution should look like.

Non-parametric bootstrapping

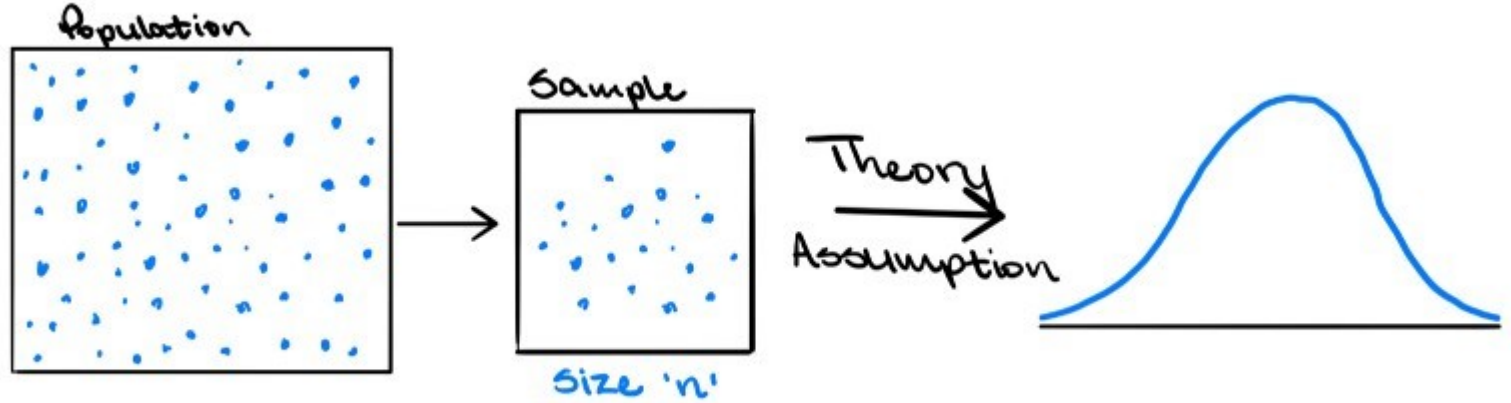
Bayesian bootstrapping

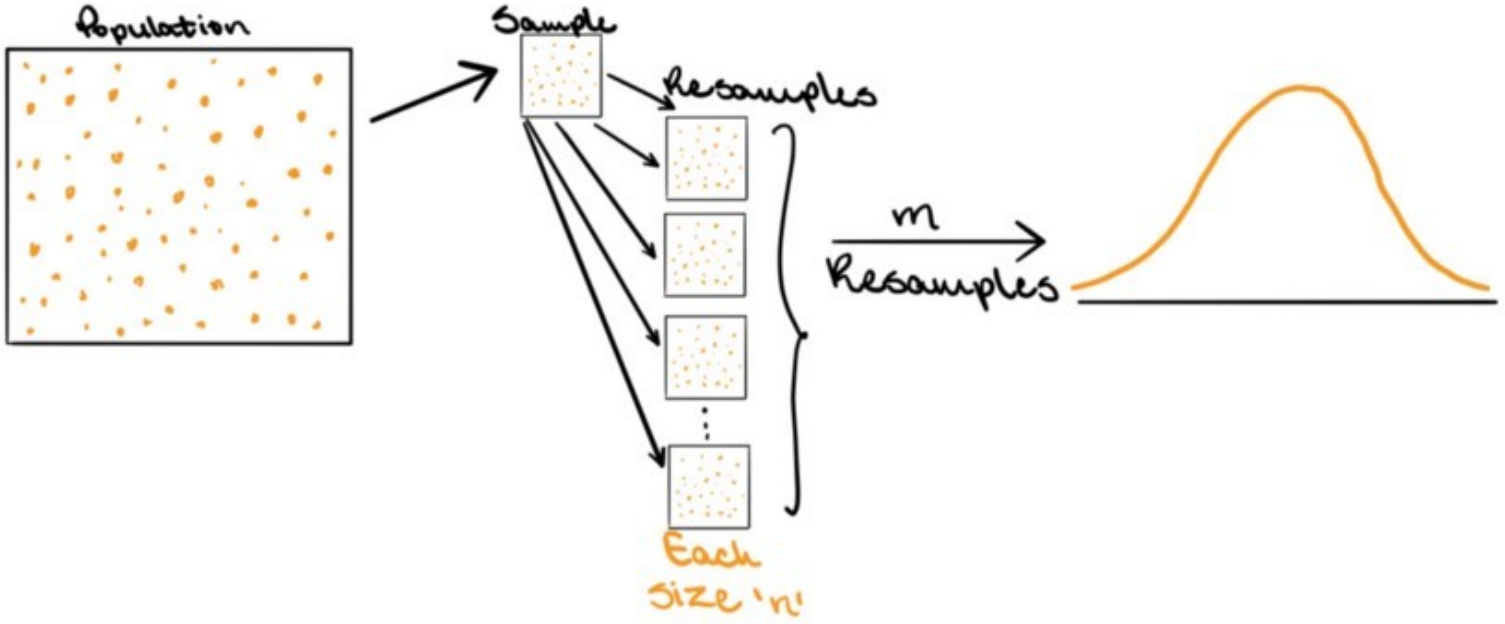
Smooth bootstrapping

1. https://jillian-green.medium.com/applications-of-bootstrapping-8240da9df6d7

How does bootstrapping work?[2]

Graphic illustrating large sample theory →





← Graphic illustrating bootstrapped sampling

2. https://towardsdatascience.com/bootstrapping-statistics-what-it-is-and-why-its-used-e2fa29577307

How does bootstrapping work?

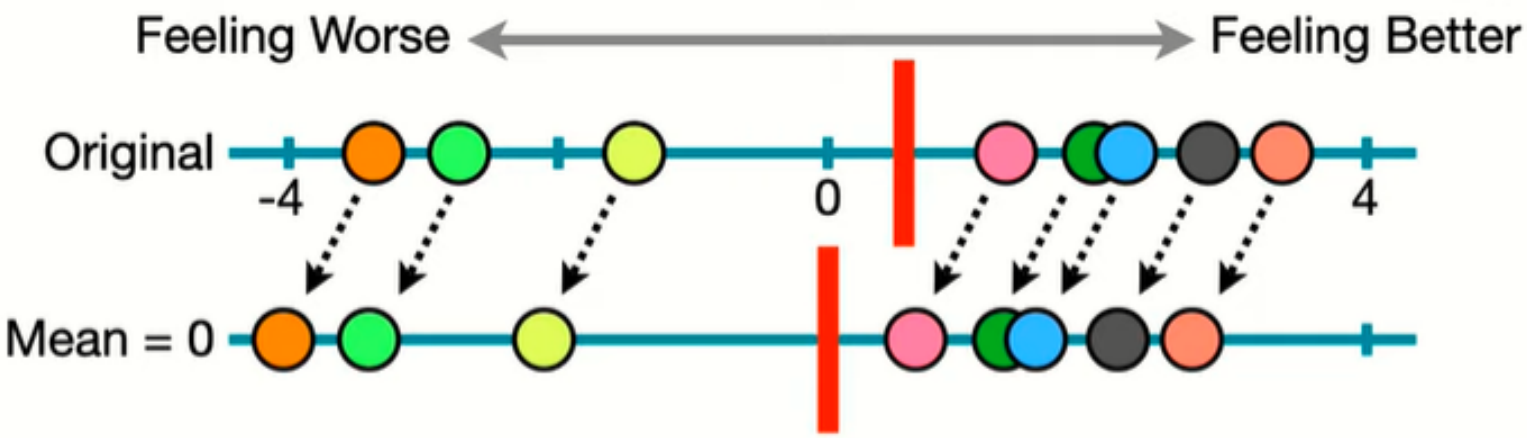
- 4-step process:
  - Step 1) Create a bootstrapped dataset
  - Step 2) Calculate some statistic
  - Step 3) Calculate that same stat. for each repetition
  - Step 4) Calculate test stats. for bootstrapped dataset by averaging results from all of your repetitions

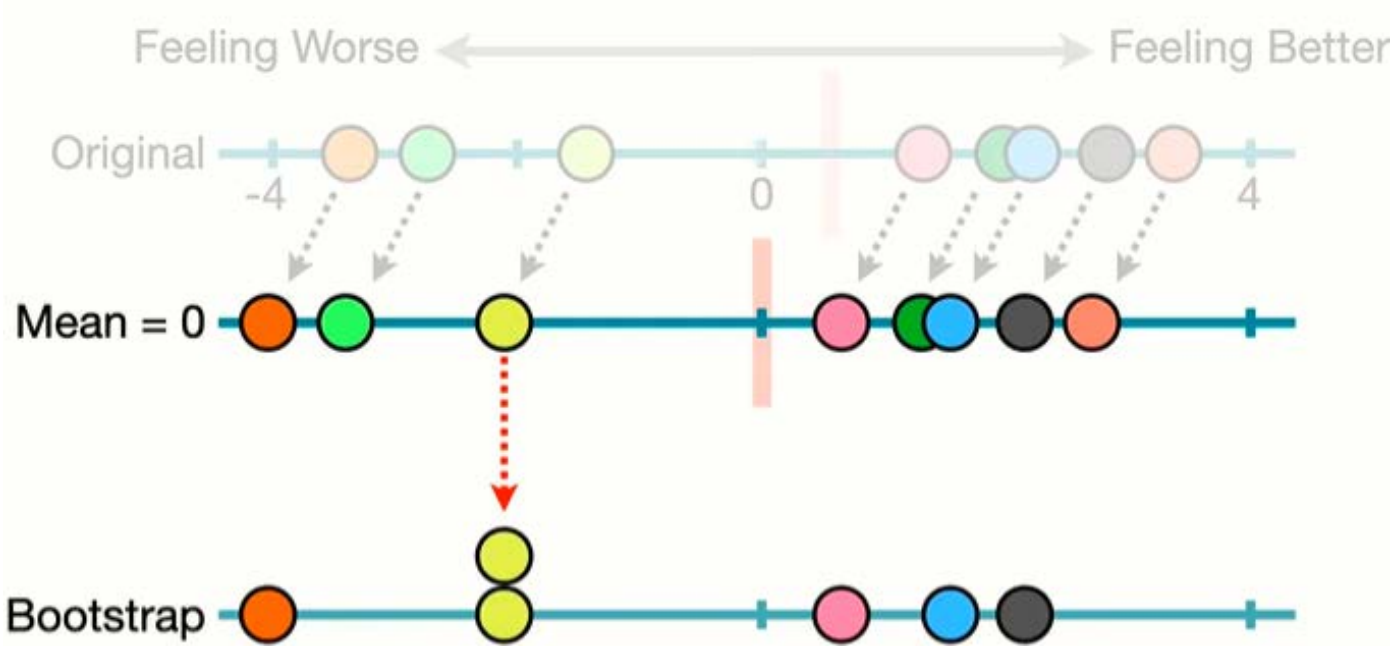
Basic Example of Bootstrapping

- Consider the total population $\{1,10,15,16,19,1,18,17,19,1\}$ where $\mathrm{n} = 10,\ \bar{x} = 11.7,\ s = 7.82,\ \mathrm{SE} = 2.47.$
- Next, consider the random sample $\{1, 10, 15, 16, 19\}$ where $\mathrm{n} = 5,$ $\bar{x} = 12.2,\ s = 7.05,\ \mathrm{SE} = 3.15.$
- We have $5^5 = 3{,}125$ possible permutations
- For simplicity, we'll repeat sampling five times, which yields the bootstrapped samples
  - $\{19, 19, 1, 19, 19\} \rightarrow \overline{x_2^*} = 15.4,\ s_2^* = 8.05,\ \mathrm{SE}_2^* = 3.60$
  - $\{15, 19, 1, 15, 19\} \rightarrow \overline{x_3^*} = 13.8,\ s_3^* = 7.43,\ \mathrm{SE}_3^* = 3.32$
  - $\{1, 1, 19, 16, 10\} \rightarrow \overline{x_4^*} = 9.4,\ s_4^* = 8.32,\ \mathrm{SE}_4^* = 3.72$
  - $\{16, 1, 16, 1, 10\} \rightarrow \overline{x_5^*} = 8.8,\ s_5^* = 7.53,\ \mathrm{SE}_5^* = 3.37$
  - $\{10, 10, 16, 1, 16\} \rightarrow \overline{x_6^*} = 10.6,\ s_6^* = 6.15,\ \mathrm{SE}_6^* = 2.75$
- That yields the bootstrapped mean $\overline{x^*} = 11.6,\ s^* = 7.496,\ \mathrm{SE}^* = 3.35$

Applications of bootstrapping

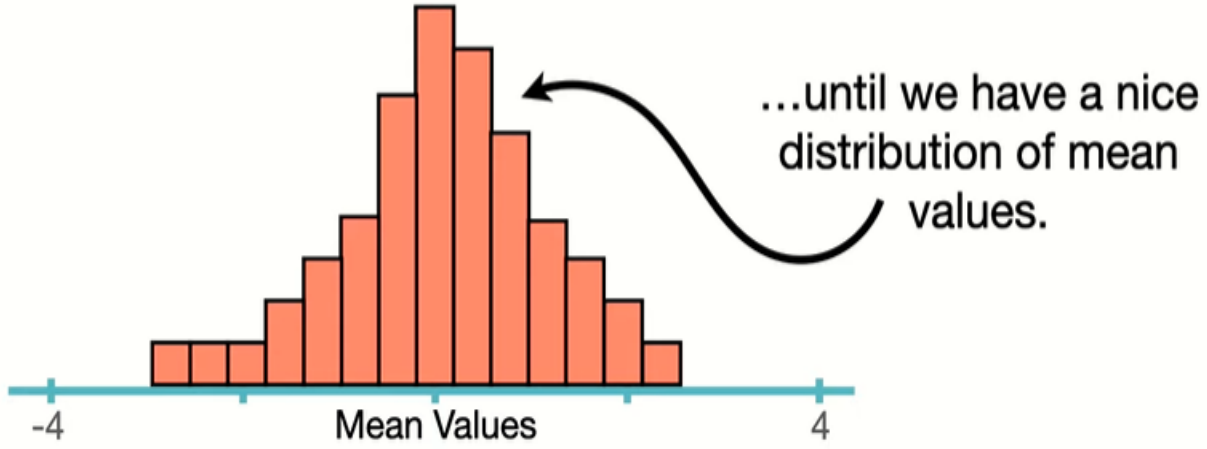
- **Hypothesis testing:** Bootstrapping can be used to help you reject, or fail to reject, a null hypothesis.
- **Example**) you are testing whether a drug works or not. You have a distribution of study participants who reported feeling worse, better, or no different from taking the drug.





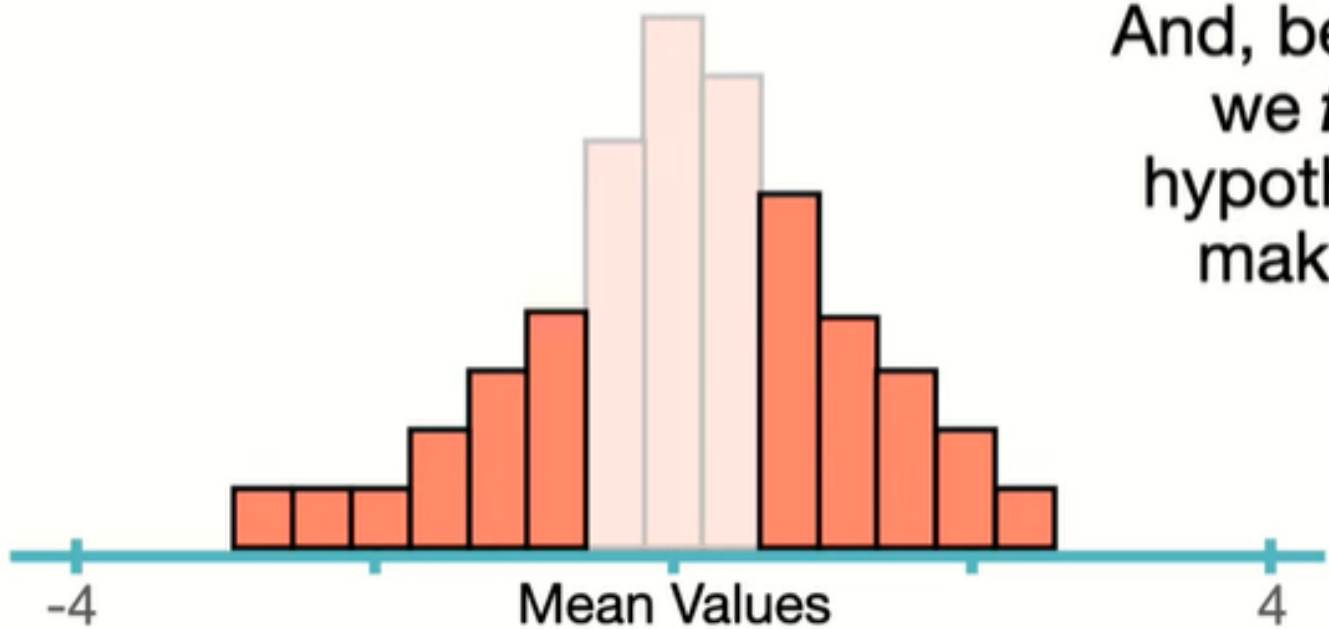
Create bootstrapped dataset from your mean-zero bootstrapped dataset

Calculate mean of each repeated sample and add to histogram



https://www.youtube.com/watch?v=N4ZQQqyIf6k&ab_channel=StatQuestwithJoshStarmer

p-value = 0.47 + 0.16 = 0.63



And, because **0.63 > 0.05**, we ***fail to reject*** the hypothesis that the drug makes no difference.

Applications of bootstrapping

- **Data collection**: Bootstrapping can be used when you do not have the resources to collect as much data as you may require, especially in methods such as surveying.
- **Other applications**: Bootstrapping is frequently used within real-world careers, such as finance.

Limitations of bootstrapping

- Bootstrap method cannot give you any new information about your data
- Bootstrap method is limited by your ability to create representative samples
- Bootstrap method may incorrectly value outliers if they are not “present” enough in your repetitions
- Bootstrap method is highly limited by computational power

Now Will is going to show you a MATLAB example about cows!