Bootstrapping and the Central Limit Theorem are two ways to approximate the sampling distribution of statistics. But how good of an approximation are they?

Go to abray.shinyapps.io/sampling-dist

The Triptych

1. Using the app, create a large slightly skewed population, draw a sample of size 100, then simulate the sampling distribution of the mean. Sketch the three distributions as density plots below.

2. Of these three plots, which one(s) are observed in real-life settings?

3. What mathematical result describes the particular shape seen in the sampling distribution?

4. Experiment with the shape of the population, the sample size, and the statistic being studied. Report 3 different settings that lead to sampling distributions that look *non-normal* and sketch those non-normal shapes below.

Approximations

Using the Triptych settings from one of your non-normal sampling distributions, click into the “Approximations” tab at the top.

5. Import your empirical distribution to serve as the bootstrap population and then draw several bootstrap samples. How many observations are being visualized in each bootstrap sample? What is the highest and lowest value that can be seen in a bootstrap sample?

6. Generate a full bootstrap sampling distribution using the statistic specified in the Triptych. Of these three plots, which one(s) can be observed in a real-life setting?

7. Overlay on top of the bootstrap distribution the shape of the true simulated sampling distribution. Draw that figure below, labeling both distributions, and comment on how good of an approximation the bootstrap sampling distribution is to the true sampling distribution.

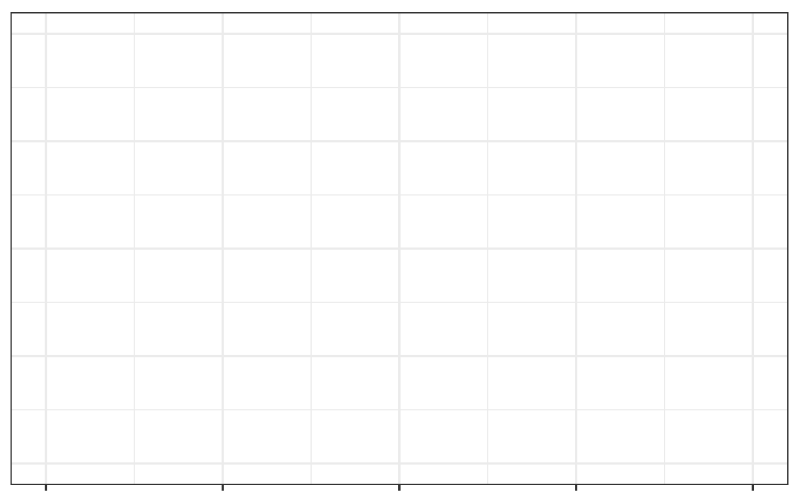
8. Use the app to develop general rules of thumb for each statistic for when the bootstrap algorithm can provide a good approximation to the shape of the sampling distribution. For each statistic, discuss the properties of the population and the sample.

Simulating Yawners

What kind of data would be observed if there was no association between these variables and if the only variation was caused by the process of randomly assigning subjects to the two conditions? Find out by *simulating* the process.

1. Create a deck of cards, 36 of which represent subjects who did not yawn, 14 of which represent subjects who yawned.
2. Shuffle the deck of cards to simulate the process of randomly assignment to the two conditions: being exposed to a yawn (stimulus) and not being exposed (no stimulus).
3. Deal them into two decks of size 16 and 34, representing the 1/3 of the subjects that were assigned to the no stimulus group and the 2/3 assigned to the stimulus group.
4. Calculate the difference in the proportion of yawners in the two group, $\hat{p}_s - \hat{p}_n$, and record it below.
5. Repeat process 5 more times and sketch the distribution of simulated statistics from the class in the blank plot.

Sim	$\hat{p}_s - \hat{p}_n$
1	
2	
3	
4	
5	
6	



(9) If there was in fact no relationship between yawning and the stimulus, what values of the statistic would you expect to see?

(10) What value of $\hat{p}_s - \hat{p}_n$ did the MythBusters actually observe?

(11) Is this data convincing evidence that yawning is contagious? Why or why not?