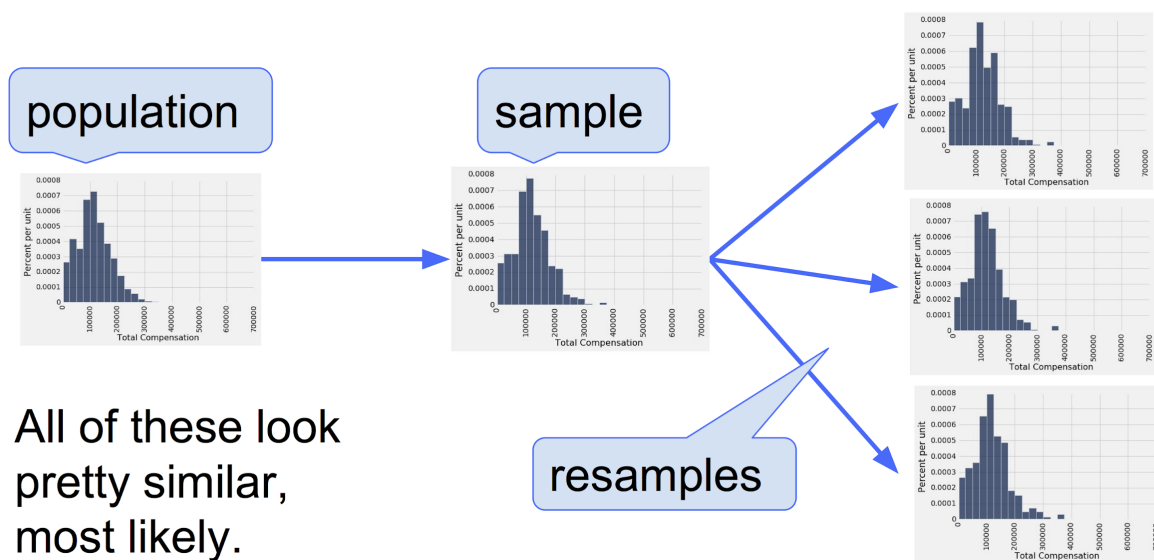


# Lab 08: Bootstrap, Confidence Intervals

## CIS 124 Discussion Worksheet

Suppose we are trying to estimate a *population parameter*. Whenever we take a random sample and calculate a statistic to estimate the parameter, we know that the statistic could have come out differently if the sample had come out differently by random chance. We want to understand the *variability* of the statistic in order to better estimate the parameter. However, we don't have the resources to collect multiple random samples. In order to solve this problem, we use a technique called *bootstrapping*.



1. **Warm-Up:** What is the difference between a parameter and a statistic? Which of the two is random?

2. **Sampling Techniques:** Assume we have one large, random sample. How could we generate another sample that resembles the population if we don't have the resources to sample again from the population?

**3. Tennis Time:** Ciara is interested in the heights of female tennis players. She's collected a sample of 100 heights of professional women's tennis players. She wants to use this sample to estimate the true interquartile range (IQR) of all heights of professional women's tennis players.

*Hint:* We defined the interquartile range (IQR) to be: **75th percentile - 25th percentile**

- a. In order to construct a 99% confidence interval for the IQR, what should our upper and lower percentile endpoints be?
- b. Define a function `ci_iqr` that constructs a 99% confidence interval for the IQR as follows. The function takes the following arguments:
  - `tbl`: A one-column table consisting of a random sample from the population; you can assume this sample is large
  - `reps`: The number of bootstrap repetitions

*Hint: To find the 25th and 75th percentile of an array, you can use the `percentile` function*

```
def ci_iqr(tbl, reps):  
    stats = _____  
    for _____ :  
        resample_col = _____  
        new_iqr = _____  
        stats = _____  
    left_end = _____  
    right_end = _____  
  
    return make_array(left_end, right_end)
```

- c. Say Ciara recruited 500 of her friends to perform the same bootstrapping process she did. In other words, each of her friends drew a large, random sample of 100 heights from the population of professional women's tennis players and constructed their own 99% confidence intervals. Approximately how many of these CI's do we expect to contain the actual IQR for the heights of professional women's tennis athletes?