

**Part 1** If we're trying to predict the results of the Clinton vs. Trump presidential race, what is the population of interest?

The people who can vote in the election.



**Part 2** What is the sampling frame?

The people who have a phone.



### 0.0.1 Question 5

Why can't we assess the impact of the other two biases (voters changing preference and voters hiding their preference)?

Note: You might find it easier to complete this question after you've completed the rest of the homework including the simulation study.

We cannot figure out if people will change their minds after the poll, or tell us false ideas. This requires a post-election survey to get the result.

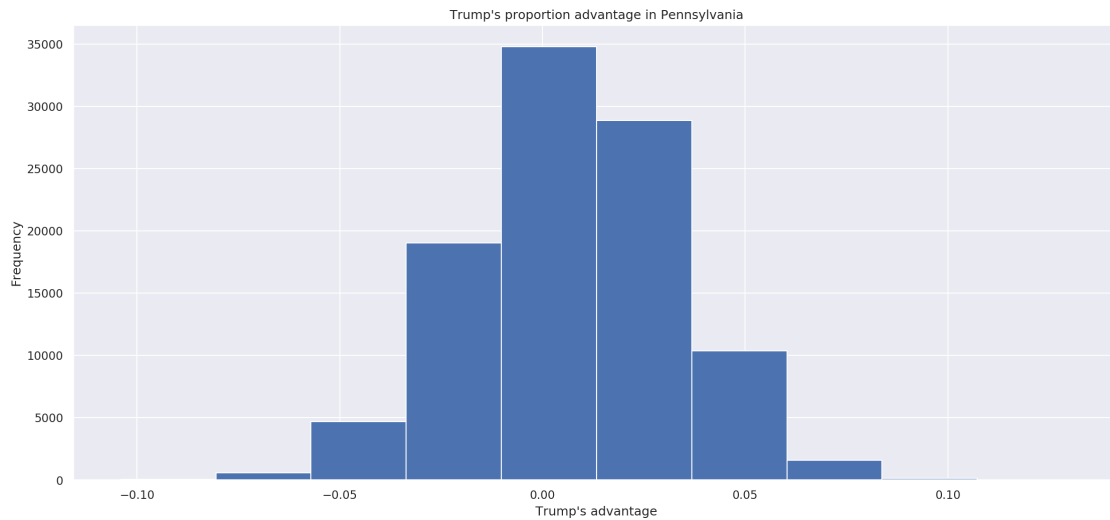


**Part 4** Make a histogram of the sampling distribution of Trump's proportion advantage in Pennsylvania. Make sure to give your plot a title and add labels where appropriate. Hint: You should use the `plt.hist` function in your code.

Make sure to include a title as well as axis labels. You can do this using `plt.title`, `plt.xlabel`, and `plt.ylabel`.

```
In [38]: plt.hist(simulations)
         plt.title("Trump's proportion advantage in Pennsylvania")
         plt.xlabel("Trump's advantage")
         plt.ylabel("Frequency")
```

```
Out[38]: Text(0, 0.5, 'Frequency')
```







**Part 2** Make a histogram of the new sampling distribution of Trump's proportion advantage now using these biased samples. That is, your histogram should be the same as in Q6.4, but now using the biased samples.

Make sure to give your plot a title and add labels where appropriate.

```
In [ ]: plt.hist(biased_simulations)
        plt.title("Trump's proportion advantage in Pennsylvania using biased samples")
        plt.xlabel("Trump's advantage")
        plt.ylabel("Frequency")
```



**Part 3** Compare the histogram you created in Q7.2 to that in Q6.4.

Those histogram looks very similar. In histogram in Q7.2, it is a little bit shift to the left compare to 6.4 because we introduced bias, the advantages of trump decrease. This shows the effect of bias on the results.



Write your answer in the cell below.

As the sample size increase, the sampling error for the unbiased sample will decrease. In the biased case, the more sample size, the greater the influence of bias. Therefore, we can see when we increase the sample size in the unbiased case, Trump's winning percentage is rising, and in the biased case, his winning percentage is going down.



### 0.0.2 Question 9

According to FiveThirtyEight: "... Polls of the November 2016 presidential election were about as accurate as polls of presidential elections have been on average since 1972."

When the margin of victory may be relatively small as it was in 2016, why don't polling agencies simply gather significantly larger samples to bring this error close to zero?

Larger samples can reduce the sampling error. However, if there has bias, the prediction is close to the bias estimate.

