

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

The population of interest are those in the US population who is eligible to vote.



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

The sampling frame would include every person they could contact with a phone in the US service area.



### 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 assess the impact because we cannot differentiate between whether voters had changed their preferences or hid their preferences.

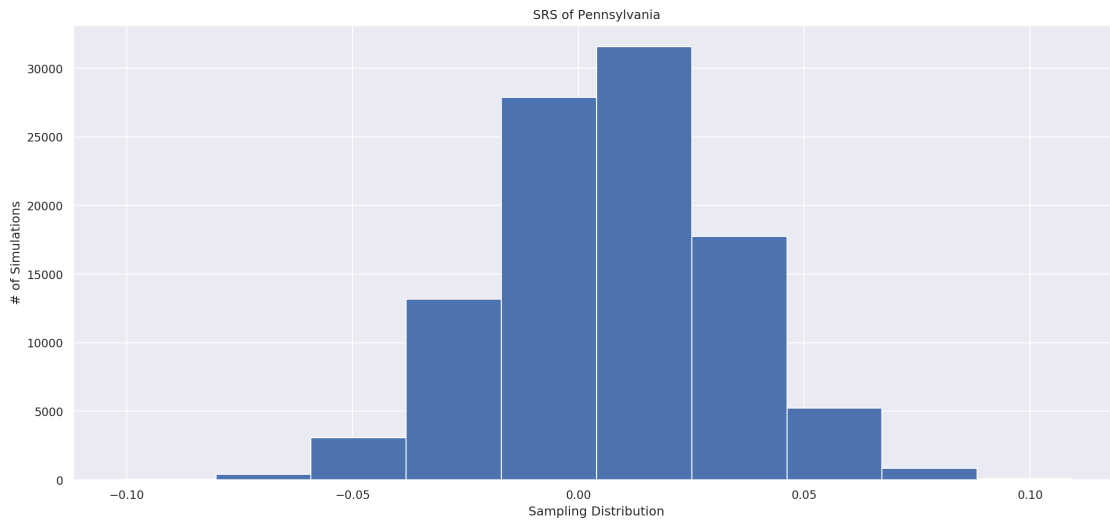


**Part 4** Make a histogram of the sampling distribution of Trump's proportion advantage in Pennsylvania. Make sure to give your plot a title and add labels where appropriate. Hint: You should use the `plt.hist` function in your code.

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

```
In [176]: plt.hist(simulations)
          plt.title('SRS of Pennsylvania')
          plt.xlabel('Sampling Distribution')
          plt.ylabel('# of Simulations')
```

```
Out[176]: Text(0, 0.5, '# of Simulations')
```





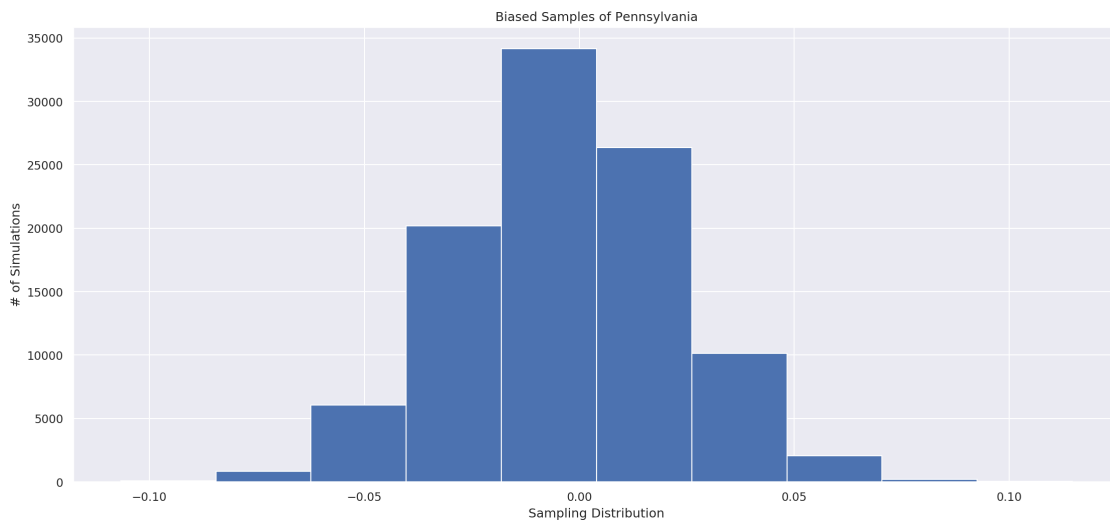


**Part 2** Make a histogram of the new sampling distribution of Trump's proportion advantage now using these biased samples. That is, your histogram should be the same as in Q6.4, but now using the biased samples.

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

```
In [203]: plt.hist(biased_simulations)
          plt.title('Biased Samples of Pennsylvania')
          plt.xlabel('Sampling Distribution')
          plt.ylabel('# of Simulations')
```

```
Out[203]: Text(0, 0.5, '# of Simulations')
```





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

Overall, the histograms between Q7.2 and Q6.4 seem similar and mostly symmetrical. However, the histogram in Q7.2 is skewed to the left when compared to Q6.4, most likely due to the bias that was implemented.



Write your answer in the cell below.

With larger sample size, the spread in the histogram decreases, which in turn decreases the sampling error and results in a more accurate prediction for the unbiased sample and a decrease in the prediction for the bias sample.



### 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?

It has been observed that larger samples decrease sampling error and increase the accuracy of the predictions. However, in the process of getting a large enough sample, if bias is introduced, then regardless of the size of the sample, the prediction will still be influenced by the bias. Therefore, it is best to take a sample that is large enough and controlled for any biases.

