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 is voters in the 2016 presidential election.

Part 2 What is the sampling frame?

The sampling frame is all people who have phone numbers that can be generated randomly.

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 can't asses voters changing preference unless they were polled again after the vote. Logistcally, this is pretty hard as we would have to track down every person we polled before. We can;t asses voters hiding their preference because we can't force people to tell the truth nor answer, they can lie or simply not answer if they wanted to.

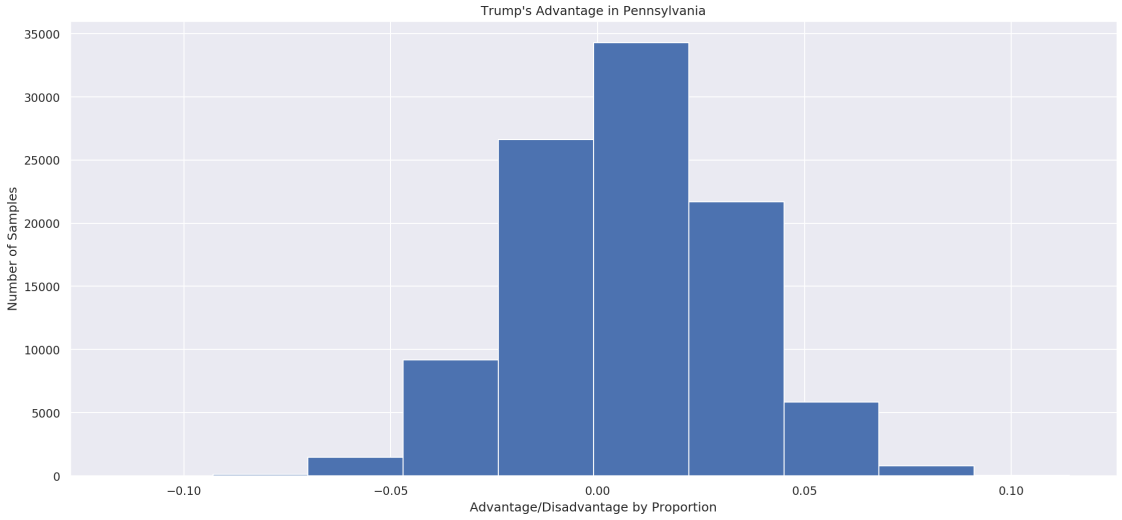
Part 4 Make a histogram of the sampling distribution of Trump's proportion advantage in Pennsylvania. Make sure to give your plot a title and add labels where appropriate. Hint: You should use the `plt.hist` function in your code.

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

```
In [90]: plt.hist(simulations)
         plt.title("Trump's Advantage in Pennsylvania")

         plt.xlabel('Advantage/Disadvantage by Proportion')
         plt.ylabel('Number of Samples')
```

```
Out[90]: Text(0, 0.5, 'Number of Samples')
```

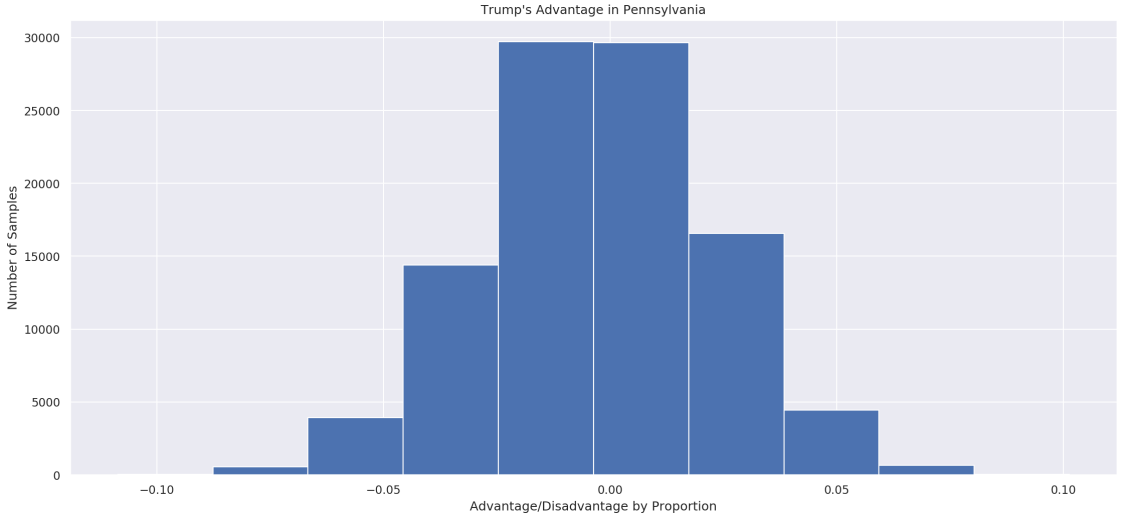


Part 2 Make a histogram of the new sampling distribution of Trump's proportion advantage now using these biased samples. That is, your histogram should be the same as in Q6.4, but now using the biased samples.

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

```
In [88]: plt.hist(biased_simulations)
         plt.title("Trump's Advantage in Pennsylvania")
         plt.xlabel('Advantage/Disadvantage by Proportion')
         plt.ylabel('Number of Samples')
```

```
Out[88]: Text(0, 0.5, 'Number of Samples')
```



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

The histogram in Q7.2 if centered around 0 and normal. Its similar to the histogram in Q6.4, except that distribution is not centered at 0, but a value above 0.

Write your answer in the cell below.

The biased sample was about the same no matter the sample size, 0.44692 vs. 0.46668, but the unbiased proportion increased by a much larger amount, 0.69413 vs. 0.8276. This shows that increasing sample size reduces the sampling error but has no effect on the bias of the 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's not easy to sample large amounts of people. These polls already sample a large amount of people, and sampling even more is difficult. Also, a larger sample will not address the bias already present in current polling methods, while attempting to obtain a larger sample could introduce even more bias in these polls.