Math 242, Polling Project Part 1: Due Monday, September 26 by 5pm, submit electronically

Your name:

As the Presidential Election draws near, we will use what we are learning about statistics to model the election in various ways. First, we have to understand what has already been done. We won’t be conducting our own polls (that would take far too long), so we’ll be relying on polls others have conducted. It’s important to think about potential biases and to get access to trustworthy polls. We will call a poll **trustworthy** if it is set up in such a way that, if a simple random sample of the population is polled, the results will be representative of the overall population. We will call a poll **biased** if there is a systematic way in which the results of the poll are not representative of the population.

(1) What are some reasons that polling organizations might be biased, i.e. might want results that are not representative?

(2) Name two concrete ways that the bias of the polling organization might affect the polling results. For each, discuss why that behavior of the polling organization would lead to a result different from the truth.

(3) Go to <http://fivethirtyeight.com/>, click “See Polls and Forecasts”, and find 2 graphics you think are particularly helpful for understanding the election. Take screenshots of the graphs and include them here. For each, write 2-3 sentences about what you liked about the graph, why you thought it did a good job summarizing important information, and how it could be better.

(4) Different websites that aggregate polls will give you different impressions of what is going on. Here are several aggregators:

<http://projects.fivethirtyeight.com/2016-election-forecast/national-polls/>

<http://www.realclearpolitics.com/epolls/latest_polls/elections/>

<http://www.politico.com/2016-election/polls/>

<http://www.nytimes.com/interactive/2016/us/elections/polls.html>

<http://www.gallup.com/topic/politics.aspx>

Choose 2 of the above, and record how many of the top 20 polls you can see had Trump leading, and how many had Clinton leading:

Did the two sources agree? If not, do you think this had more to do with random variation or more to do with bias?

Please find two more poll aggregators that you feel represent some bias, and record how many polls they show had Trump leading and how many had Clinton leading. Also, include the links to the aggregators you found:

(5) Find a confidence interval that was reported as a result of a poll. Try to find a source that gives more than just the final result, e.g. that gives the sample size and original standard deviation. Identify the population (e.g. all voters, likely voters, etc) and what kind of confidence interval it is (e.g. proportion, mean, difference of proportions). What factors might change your confidence in this poll between now and election day?