

Assignment 1. Opinion Polling

In this assignment, you will be expected to analyze a dataset on your own and answer questions about your findings.

Due date: Friday Oct 25, 2019

Completed assignments will be collected in class.

Scenario:

You've been hired as a consultant to predict how a state school board election will turn out.

1. There are three candidates and all voters must vote for one of them: Pearle Goodman, Masako Holley, Genevieve Gallegos.
2. The candidate with the final highest vote count wins the election.
3. You are given the list of registered voters here:
<https://github.com/sjyk/cmssc21800/blob/master/voters.csv>
4. The state gives you two samples of data one collected by SurveyMonkey and one collected by Qualtrics:
https://github.com/sjyk/cmssc21800/blob/master/survey_monkey.csv
<https://github.com/sjyk/cmssc21800/blob/master/qualtrics.csv>

Q1. The SurveyMonkey data shows Genevieve Gallegos winning 59% vote of 100 people polled and the Qualtrics data shows her losing with 42% vote of 50 people polled. Which of the following best describes the likelihood that a difference this large ($>17\%$) happened purely by random chance and not an error in the polling process?

- a) 20% chance of a variation greater than 17% in independent polls.
- b) 10% chance of a variation greater than 17% in independent polls.
- c) $<5\%$ chance of a variation greater than 17% in independent polls.
- d) Not enough information to determine this.

Explanation:

Q2. The data provider suspects that the SurveyMonkey dataset is biased. What do you think?

- a) Yes, the SurveyMonkey dataset shows a clear bias in data collection
- b) No, the observed bias is likely due to the natural variation in randomly sampled data
- c) The sample size is too small to determine this

Explanation:

Q3. Which of the following best describes the margin of error of the Qualtrics poll:

- a) +/- 10% with 99% confidence
- b) +/- 15% with 95% confidence
- c) +/- 20% with 95% confidence

Explanation:

Q4. A news report suggests that Pearle Goodman is dropping out of the election. Is it clear which candidate benefits from her departure?

Explanation:

Q5. How likely are the following scenarios to affect the polling results (choose between “in favor of Genevieve Gallegos”, “in favor of Masako Holley”, or statistically insignificant/unclear), explain.

- a) Only half of the registered voters from Mountain Farm County turn out to vote.

- b) The elections are held during the regional college’s final exam week leading to a poor turnout for the 18-25 and 26-35 age group.

- c) A women’s organization in Mountain Farm County endorsed Masako Holley.