**Problem Set 4**

This problem set covers material from classes 8 and 9, including material from OIS sections 4.1, 4.3, and 5.1. It makes use similar data to that we discussed in class. Partial credit may be given for answers that are correct in part, but not in full. This problem set is due on Gradescope by Wednesday September 22, 2021 at 11:59 PM.

**Part I: A Normal Election? (24 pts.)**

1. This November, two states—New Jersey and Virginia—will hold regularly scheduled gubernatorial elections. Here, we are going to work with data on recent gubernatorial elections in Virginia, our neighbor to the North. Below is a histogram of the percentage of the vote received by the Republican candidate in all 14 Virginia gubernatorial elections since 1969, the first gubernatorial election Republicans won in Virginia since 1885. Since then, these elections have traditionally been competitive, with Republicans never getting below 40 or above 60 percent of the vote. Below is a histogram (bin width = 5 percentage points) and a Q-Q plot of the distribution of the vote percentage. (24 pts.)

|  |  |
| --- | --- |
|  |  |

1. Does this data look like it follows a normal distribution? Discussing both plots, what evidence is there that supports your answer? (8 pts.)
2. Virginia is alone among states in currently term-limiting governors to a single consecutive term. However, former governors are theoretically allowed to run for their old seats again, which Democratic former Governor Terry McAuliffe is doing this year. When he won his first term in 2013, McAuliffe’s Republican opponent Ken Cuccinelli received 45.2% of the vote. In this period, the average Republican vote percentage was 50.5%, with a standard deviation of 5.2%. Setting aside whatever you answered for part a, if we were to assume normality, what is Cuccinelli’s, z-score? What is its percentile for Republican vote? Please show your work or explain how you got the value. (8 pts.)
3. The most recent election in which a Republican won the governorship of Virginia was in 2009, when then- VA Attorney General Bob McDonnell won the office with 58.6% of the vote. Using the same assumptions and overall mean and standard deviation as part b, what is McDonnell’s z-score and percentile? Please show your work and discuss how you found the percentile. (8 pts.)

**Part II: The Most Ambitious Crossover Event in the History of Stats Assignments (24 pts.)**

1. At the end of the movie *Avengers: Infinity War*, when Thanos snaps his fingers, half of all life remains and half disappears. Imagine that each trial is independent (i.e., there is a 50/50 chance of any one living thing disappearing or remaining) and that each living thing has a 50% chance of disappearing. (24 pts.)
   1. There are 7 Avengers/Avenger allies on Titan. Five of these Avengers disappear, while two remain. What is the probability of *exactly* two Avengers remaining? What is the probability of *at least* two remaining? (12 pts.)
   2. At the beginning of *Avengers: Endgame*, Hawkeye is standing in a field with his wife and three children. When Thanos snapped his fingers, Hawkeye remains, but the rest of his family now disappears. What is the probability of only one person of five in this field remaining when Thanos snapped his fingers? (6 pts.)
   3. In the post-credits scene of Infinity War, both Nick Fury and Agent Maria Hill are snapped. What is the probability that both of them will disappear? (6 pts.)

**Part III: Does Congress approve of itself? (22 pts.)**

1. Congress is historically unpopular with the American people, with the late Senator John McCain once joking that approval was down to [paid staff and blood relatives](https://www.mcall.com/la-pn-john-mccain-mom-congress-20131113-story.html). Yet a [Gallup Poll](https://news.gallup.com/poll/329945/congressional-job-approval-highest-nearly-years.aspx) earlier this year of 1,021 adults noted that approval was up to a recent high of 35%. Is it possible that members of Congress themselves were indicating support (or lack of support) for their institution in the poll? I would like you to determine the probability that exactly one member of Congress was contacted for this poll and the probability that at least one member of Congress was contacted for the survey.

Some relevant information for this question:

* To simplify this problem, you don’t have to consider non-response or the lack of replacement after a person is polled.
* According to Census data compiled by the website InfoPlease, there are approximately 209,128,094 adults United States. At the time this survey has conducted, there were 532 members of Congress including both the House and Senate (there were 3 vacancies).
* Hint: First calculate the probability of a random draw of registered voters including a student from this class.

**Part IV: You can’t spell Clinton, Ohio without CLT… (30 pts.)**

1. My hometown of Clinton, Ohio last held a mayoral election in 2019. (Clinton is a rural, really small town south of Akron with a population of about 1200…. so small that when in 2016 the Akron newspaper did a series of stories about the top 10 things to do in each municipality in Summit County, they could only come up with [eight](https://www.beaconjournal.com/article/20160730/NEWS/307309599) things to do in Clinton—and one was a bicentennial parade that was only happening once!) In 2019, Clarissa Allega received 152 votes for mayor and Bud McDaniel received 129 votes for Mayor. Imagine that someone wanted to conduct a series of polls of these voters. Based on the population size, what are two difficulties that the pollster would face based on the Central Limit Theorem? What size should their samples be? Be as specific as possible. (12 pts.)
2. In 2020 in the village of Clinton, Donald Trump received 476 votes and Joe Biden received 207 votes. Imagine that a pollster who is interested in the proportion of the vote Trump has in the village decided to conduct a series of polls of sample size 50 of this election right before it took place using random sampling. Based on the CLT, what would µp̂ and SEp̂ be equal to here? (Please calculate these.) What percent of the time would you expect these polls to show Trump with at least 70% of the two-party vote (i.e., the percent who prefer Trump of those who are voting for either Trump or Biden)? Please show all work and explain how you got your answers for this problem. (18 pts.)