**Problem Set 5**

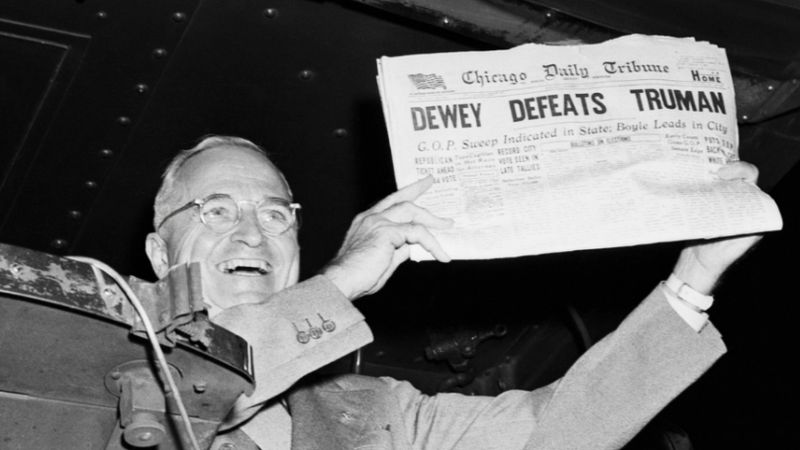
This problem set covers material from classes 13 and 14, including material from OIS sections 5.3 and 6.1 and the Denworth article. It makes use similar data to that we discussed in class. Partial credit may be given for answers that are correct in part, but not in full. This problem set is due on Gradescope by Wednesday October 13, 2021 at 11:59 PM.

**Part I: Hypothesis Testing Framework (24 pts.)**

1. You are a researcher who would like to examine a topic of interest. Write a sentence or two describing your topic. Then, write your null hypothesis and an alternative hypothesis for both a “not-equal” and either a “greater-than” or “less-than” scenario (based upon which you think makes more sense here). You may choose the topic you proposed for your project, but may also choose something else. You may collaborate with your group on this problem, but should use your own words (12 pts).
2. Having collected data, you would now like to evaluate the evidence. (12 pts.)
   1. After getting your results, what type of evidence would lead you to reject the null? Would you ever accept the null? (6 pts.)
   2. Why might you prefer to use a “not equal” alternative rather than a “greater-than”/ “less than alternative” here? (6 pts.)

**Part II: Confidence Intervals and Hypothesis Testing (66 pts.)**

1. In the years since the 1948 election, the below image became famous. Here, President Harry S Truman, who had just won reelection, is seen holding up a copy of the Chicago Daily Tribune declaring that Republican New York Governor Tom Dewey had won the election. In this problem, you will work with data from the final Gallup Poll from before that election (66 pts.).



Source: [Chicago Tribune](https://www.chicagotribune.com/featured/sns-dewey-defeats-truman-1942-20201031-5kkw5lpdavejpf4mx5k2pr7trm-story.html)

1. The final Gallup Poll before the 1948 election conducted of 1500 adults in late October of that year had Governor Dewey at 50% of the vote, President Truman at 45% of the vote, former Vice President Henry Wallace at 4%, and South Carolina Governor Strom Thurmond at 2% of the vote (does not add to 100% due to rounding; Source: Roper Center). This poll was conducted using quota sampling, but for the purposes of this problem, you can carry out calculations as if it was a simple random sample. Find and interpret a 95% confidence interval for support for President Truman and discuss whether you have evidence to suggest that his percentage differs from 50% and how you know this (12 pts.).
2. Now, please find and interpret the z-score for support for President Truman. (12 pts.)
3. After finding the z-score, what are two options for finding the p-value associated with this estimate? Using one of these, find and interpret the p-value (95% confidence level). Can we reject the null hypothesis here? (8 pts.)
4. Using a one-sided hypothesis test associated with the same z-score, do you have evidence to suggest with 95% confidence that Truman’s percentage in the poll is *less than* 50%? Ultimately, Truman received just under 50% on Election Day. Given your estimate here, is that result particularly surprising? Given the result provided earlier in this problem, what is one thing that may have caused this result? (10 pts.)
5. Find and interpret a 95% confidence interval around the support for Governor Dewey in the poll. Does this confidence interval overlap the 95% confidence interval in support for President Truman that you found in part a? Please discuss what you find. (12 pts.)
6. How small would the margin of error have to be for Truman to be within the margin of error of Dewey? (Hint: this comes from pre-exam material) Is Gallup’s sample size big enough for this result to be outside the margin of error? How big would they have had to make the survey it in order to have a margin of error smaller than this number? Please show your work, but double check using a calculator (or R). For these calculations, you can set *p* equal to 0.5. (12 pts.)

**Part III: P-Values and Confidence Levels (10 pts.)**

1. Is there any reason why we have to have a 95% confidence level when hypothesis testing? Denworth suggests using surprisals as an alternative to p-values. Do you find this approach intuitive? Why or why not? What would a p-value of 0.005 represent in surprisals? (10 pts.)