Point Estimates and Sampling Variability

Mini-Assignment - MTH 361 A/B - Spring 2023

Instructions:

- Please provide complete solutions for each problem. If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- Note that some problems have multiple solutions or ways to solve it. Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- Creativity and collaborations are encouraged. Use all of the resources you have and what you need to complete the mini-assignment. Each student must take personal responsibility and submit their work individually. Please abide by the University of Portland Academic Honor Principle.
- **Please save your work as one pdf file, don't put your name in any part of the document, and submit it to the Teams Assignments for this course. Your document upload will correspond to your name automatically in Teams.**
- If you have questions or concerns, please feel free to ask the instructor.

I. Null and Sampling Distributions

Materials

The exercises below are derived from the textbook OpenIntro Statistics (4th edition) by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr.

Exercises

1. **Unexpected expense.** In a random sample 765 adults in the United States, 322 say they could not cover a $400 unexpected expense without borrowing money or going into debt.
    a. What population is under consideration in the data set?
    b. What parameter is being estimated and what is the point estimate for the parameter?
    c. What is the name of the statistic we use to measure the uncertainty of the point estimate?
    d. Compute the value from part (c) for this context.
    e. A cable news pundit thinks the value is actually 50%. Should she be surprised by the data?
    f. Suppose the true population value was found to be 40%. If we use this proportion to recompute the value in part (d) using $p = 0.4$ instead of $\hat{p}$, does the resulting value change much?

2. **Repeated water samples.** A nonprofit wants to understand the fraction of households that have elevated levels of lead in their drinking water. They expect at least 5% of homes will have elevated levels of lead, but not more than about 30%. They randomly sample 800 homes and work with the owners to retrieve water samples, and they compute the fraction of these homes with elevated lead levels. They repeat this 1,000 times and build a distribution of sample proportions.
    a. What is this distribution called?
    b. Would you expect the shape of this distribution to be symmetric, right skewed, or left skewed? Explain your reasoning.
    c. If the proportions are distributed around 8%, what is the variability of the distribution?
    d. What is the formal name of the value you computed in (c)?
    e. Suppose the researchers' budget is reduced, and they are only able to collect 250 observations per sample, but they can still collect 1,000 samples. They build a new distribution of sample proportions. How will the variability of this new distribution compare to the variability of the distribution when each sample contained 800 observations?

3. (Outstanding Question) **Online communication.** A study suggests that 60% of college student spend 10 or more hours per week communicating with others online. You believe that this is incorrect and decide to collect your own sample for a hypothesis test. You randomly sample 160 students from your dorm and find that 70% spent 10 or more hours a week communicating with others online. A friend of yours, who offers to help you with the hypothesis test, comes up with the following set of hypotheses in mathematical notations.
$$H_0 : \hat{p} < 0.6$$
$$H_A : \hat{p} > 0.7$$
    a. Indicate any errors you see and provide a corrected hypothesis in mathematical notations.
    b. What is the null value and the point estimate of the parameter?
    c. What are the null and sampling distributions? What are their centers? Explain your reasoning.
    d. Measure the uncertainty of the point estimate and measure the uncertainty of the null value.
    e. Compare the values computed in (d) and (e). Are there any significant differences in these values? Explain your reasoning.
    f. Compute the Z-score of the point estimate under the assumption that the null hypthesis is true. Should you be surprised by the data?