PSTAT 5A: Homework 4

Please note, you are responsible for completing your homework and submitting it to Gradescope. Though the Syllabus allows for the option to not submit homework, it is highly recommended.

1. The U.S. Department of Housing and Urban Development defines a person or household to be “rentburdened” if 30% or more of the individual/household’s income is spent on housing. A recent survey revealed that 42%of households in a representative sample of 150 households were rent-burdened.
   a. Define the parameter of interest.

   b. Define the random variable of interest.

   c. Construct a 95% confidence interval for the true proportion of rent-burdened households, and interpret your interval in the context of the problem.

   d. Would you expect an 80% confidence interval for the true proportion of rent-burdened households to be wider or narrower than the 95% confidence interval you constructed in part (c)?

e. Explain briefly.

f. Construct an 80% confidence interval for the true proportion of rent-burdened households, and interpret your interval in the context of the problem.

2. In a particular iteration of PSTAT 5A, scores on the final exam had an average of 89 and a standard deviation of 40 The exact distribution of scores is, however, unknown. Suppose a representative sample of 100 students is taken, and the average final exam score of these 100 students is recorded.
   a. Identify the population.

   b. Identify the sample.

   c. Define the parameter of interest. Use the correct notations for the distribution.

d. Define the random variable of interest. Use the correct notations for the distribution.

e. What is the sampling distribution of the random variable you defined in part (d) above? Be sure to check any conditions that might need to be checked!

f. What is the approximate probability that the average score of these 100 students lies within 5 points of the true average score of 89?

3. Quinn is interested in performing inference on the average weight of Granny Smith apples in the Santa Barbara location of Bristol Farms. To that end, he takes a representative sample of 52 apples; the mean weight of his sample was 83g and the standard deviation of weights in his sample was 17g.
   a. Identify the population

   b. Identify the sample

c. Define the parameter of interest. Use the correct notations for the distribution.

d. Define the random variable of interest. Use the correct notations for the distribution.

e. What distribution do we use to construct confidence intervals for the true average weight of a Granny Smith apple at the Santa Barbara location of Bristol Farms?

f. Construct a 95% confidence interval for the true average weight of a Granny Smith apple at the Santa Barbara location of Bristol Farms.

4. Meta recently launched the social media app Threads. As the new resident Data Scientist for Meta's Santa Barbara division (congratulations!), you would like to determine the true proportion of Santa Barbara residents that have made a Threads account. Your supervisor believes that 47% of all Santa Barbara residents have made a Threads account; in a representative sample of 120 residents, however, you observe that only 48 of these sampled individuals have made a Threads account. You would like to use your data to test your supervisor's claims against a two-sided alternative, at a 5% level of significance.
   a. Define the parameter of interest.

   b. Define the random variable of interest.

   c. State the null and alternative hypotheses in terms of the parameter of interest.

   d. What is the observed value of the test statistic?

   e. What distribution does the test statistic follow, assuming the null is correct?

f. What is the critical value of the test?

g. Conduct the test, and phrase your conclusion in the context of the problem.

5. **[EXTRA CREDIT 6 points] (Deriving the Lower-Tailed Test of Proportions)**. Consider testing the set of hypothesis

$H_0: p = p_0$

$H_A: p < p_0$

at an arbitrary α level of significance. Define the test statistic TS to be

$$TS = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

a. Show that TS ~ N(0, 1) under the null hypothesis. If your answer depends on a set of conditions to be true, explicitly state those conditions.

b. Argue, in words, that the test should be of the form

$$\texttt{decision(TS)} = \begin{cases} \texttt{reject } H_0 & \text{if TS} < c \\ \texttt{fail to reject } H_0 & \text{otherwise} \end{cases}$$

for some constant c. As a hint, look up the logic we used in lecture to derive the two-tailed test, and think in terms of statements like "$\hat{p}$ is far away from p_0." You do not have to find the value of c in this part.

c. Now, argue that c must be the $\alpha \times 100$th percentile of the standard normal distribution (NOT scaled by negative 1), thereby showing that the full test takes the form

$$\texttt{decision(TS)} = \begin{cases} \texttt{reject } H_0 & \text{if TS} < z_\alpha \\ \texttt{fail to reject } H_0 & \text{otherwise} \end{cases}$$

where z_α denotes the (α)×100th percentile of the standard normal distribution.