Programming Assignment 5
STAT 311

Please complete the following problems and submit a file named `STAT311-HW5.R` to Gradescope. You should start from the provided `STAT311-HW5.R` file on Canvas.

Overview

Address each of the following questions. As always, you should start from the provided skeleton code.

Question 1 - (4 points)

Construction a function called `ci.for.proportion` which calculates confidence intervals for proportions. Your function should take three arguments, `ci.for.proportion(phat, n, conf)` where `phat` is the sample proportion, `n` is the sample size, and `conf` is the level of confidence, expressed as a real number between 0 and 1.

Your function should return a list of two items, representing the lower and upper bound of a confidence interval for the specified level of confidence.

Question 2 - 5 points (d, f, g graded together)

The state of Florida is in the midst of a homeowners insurance crisis, with insurance costs skyrocketing and pricing many homeowners out of the market. More and more homeowners are resorting to self-insurance, saving money that would be spent on insurance in the hopes that the savings can cover any issues that arise that would normally fall under an insurance policy. (Some people might call this 'not having insurance', but sure, let's go with 'self-insurance'.)

A survey of `487` Florida homeowners found that `82` of those surveyed reported relying on self-insurance.

a) Based on the sample proportion found, what is the estimated standard error of the distribution of $\hat{p}$? Save your answer in the variable `q2.a`

b) Calculate a 90% confidence interval for the true proportion of Florida homeowners who rely on self-insurance. (Your answer should be formatted similarly to the function output, if done by hand). Save your answer in the variable `q2.b`

c) If the true proportion of Florida homeowners relying on self-insurance were $p = .177$, what is the true standard error of the distribution of $\hat{p}$? Save your answer in the variable `q2.c`

d) If the true proportion of Florida homeowners relying on self-insurance were $p = .177$, what is the probability that the interval you found in `q2.b` contains the true proportion? Save your answer in the variable `q2.d`

e) If the true proportion of Florida homeowners relying on self-insurance were $p = .202$, what is the true standard error of the distribution of $\hat{p}$? Save your answer in the variable `q2.e`

f) If the true proportion of Florida homeowners relying on self-insurance were $p = .202$, what is the probability a 95% interval constructed from the sample used in `q2.b` would contains the true proportion? Save your answer in the variable `q2.f`

g) If the true proportion of Florida homeowners relying on self-insurance were $p = .202$, what is the probability a 95% interval constructed from a new sample of size `487` would contains the true proportion? Save your answer in the variable `q2.g`

Question 3 - 7 points

A 2022 survey of high school students found that a reported 14.1% of students used e-cigarettes. Researchers attempting to show that the proportion of students using e-cigarettes has decreased in 2023. They take a sample of `189` high school students and find that `19` of the high school students used e-cigarettes.

Researchers use this sample to conduct a hypothesis test to determine if there is evidence that less students are using e-cigarettes than the previous year. They plan to test their hypothesis at the $\alpha = 5\%$ level.

Your answers should be exact, utilizing the binomial distribution.

a) Based on the proposed hypothesis test of the researcher, what is the assumed value of the population parameter p. Save your answer in the variable `q3.a`

b) How many standard deviations (or standard errors) away from the mean is the observed sample proportion? Save your answer in the variable `q3.b`

c) What is the probability under the null hypothesis of observing an estimate of the parameter as or more extreme than the observed $\hat{p}$? (This is the p-value). Save your answer in the variable `q3.c`

d) What is the maximum number of students reporting using e-cigarettes that would lead the researchers to reject the null hypothesis? (The sample size remains unchanged). Save your answer in the variable `q3.d`

e) What is the exact probability of committing a type 1 (type I) error if the null hypothesis is true? Save your answer in the variable `q3.e`

A much larger study of high school students found that 10.0% of students used e-cigarettes. Assume this is the true value of the parameter.

f) Based on the true value of the parameter, what is the probability of committing a type 1 (type I) error? Save your answer in the variable `q3.f`

g) Based on the true value of the parameter, what is the probability of observing a level of e-cigarette use that would causes the researchers to reject the null hypothesis? (This is the power of the test). Save your answer in the variable `q3.g`