Name: ________________________________ **Date:** ___________________________________

PSTAT 5A: Homework 4

Summer Session B 2025, with *Annie Adams*

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.

   Let p denote the proportion of households that are rentburned.

   (b) Define the random variable of interest.

   Let $\hat{P}$ denote the proportion of households in a representative sample of 150 that are rent burned.

(c) Construct a 95% confidence interval for the true proportion of rentburdened households, and interpret your interval in the context of this problem.

Our first task is to identify the sampling distribution of $\hat{p}$, which entails checking the success-failure conditions. Since we don't know the value of p, we use the substitution approximation:

1) $n\hat{p} = (150) \cdot (0.42) = 63 \geq 10 \checkmark$

2) $n(1-\hat{p}) = (150) \cdot (1 - 0.42) = 87 \geq 10 \checkmark$

Since both conditions are met, we can invoke the Central Limit Theorem for Proportions to conclude that $\hat{p}$ will be approximately normally distributed. Hence, our Confidence Interval will take the form

$$\hat{p} \pm z \cdot \sqrt{\frac{p(1-p)}{n}}$$

where z is the appropriately-selected quantile of the normal distribution. Since we want a 95% confidence interval, we take z to be negative one times the $(1 - 0.95)/2 \times 100\% = 2.5^{\text{th}}$ percentile of the standard normal distribution, which we know is around 1.96. Hence, our confidence interval is

$$(0.42) \pm (1.96) \cdot \sqrt{\frac{(0.42)(1-0.42)}{150}} = (0.42) \pm (1.96) \cdot (0.0402)$$

$$\approx [0.341208, 0.498792]$$

The interpretation of this interval is:

We are 95% certain that the true proportion of rent-burdened households is between 34.12% and 49.88%.

(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? Explain briefly.

We know that higher confidence intervals lead to wider intervals, meaning an 80% confidence interval should be narrower than a 95% one.

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.

The population is the set of all students in the aforementioned iteration of PSTAT 5A.

(b) Identify the sample

The sample is the 100 students that were selected.

(c)Define the parameter of interest, using the correct notation.

We use μ to denote population means; as such, let μ denote the true average final exam score of PSTAT 5A students.

(d) Define the random variable of interest, using the correct notation.

We use $\bar{X}$ to denote sample means; as such, let $\bar{X}$ denote the average final exam score of 100 randomly-selected students from PSTAT 5A.

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

The first question we ask ourselves is: is the population normally distributed? The answer is no. As such, we then ask ourselves: is the sample size greater than 30? The answer is yes. As such, we finally ask ourselves: is the population standard deviation known? The answer is yes. Hence, $\bar{X}$ will be normally distributed; specifically,

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \sim N\left(89, \frac{40}{\sqrt{100}}\right) \sim N(89, 4)$$

(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?

We seek $P(84 \leq \bar{X} \leq 94)$. As such, we compute

$$P(84 \leq \bar{X} \leq 94) = P(\bar{X} \leq 94) - P(\bar{X} \leq 84)$$

$$= P\left(\frac{\bar{X} - 89}{4} \leq \frac{84 - 89}{4}\right) - P\left(\frac{\bar{X} - 89}{4} \leq \frac{94 - 89}{4}\right)$$

$$= P(Z \leq 1.25) - P(X \leq -1.25) = 0.8944 - 0.1056 = 78.88\%$$

3. 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.

Let p denote the true proportion of Santa Barbara residents that have made a Threads account.

(b) Define the random variable of interest.

Let $\hat{P}$ denote the proportion of Santa Barbara residents in a sample of 120 that have made a threads account.

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

Our null hypothesis is that $p = 0.47$; we are told to adopt a two-sided alternative, meaning our hypotheses take the form

$$\begin{cases} H_0 : p = 0.47 \\ H_A : p \neq 0.47 \end{cases}$$

(d) What is the observed value of the test statistic?

First, note that $\hat{p} = (48/120) \cdot 100 = 0.4$

Thus,

$$\mathrm{ts} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.4 - 0.47}{\sqrt{\frac{0.47(1-0.47)}{120}}} \approx -1.54$$

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

We know that TS will be normally distributed under the null, provided that:

1) $np_0 = (120)(0.47) = 56.4 \geq 10 \checkmark$

2) $n(1 - p_0) = (120) \cdot (1 - 0.47) = 63.6 \geq 10 \checkmark$

Since both conditions hold, we can conclude that $\mathrm{TS} \overset{H_0}{\sim} N(0, 1)$

(f) What is the critical value of the test?

Recall that for a general α level of significance, the critical value is found to be :

- -1 times the $(\alpha/2) \times 100^{\text{th}}$ percentile of the standard normal distribution, which is equivalent to the $[1 - (\alpha/2)] \times 100^{\text{th}}$ percentile of the standard normal distribution

Since $\alpha = 0.05$, this leads us to a critical value of 1.96.

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

We reject the null only when the absolute value of the observed value of the test statistic exceeds the critical value. Here, $|\mathrm{ts}| = |-1.54| = 1.54 < 1.96$ meaning we fail to reject the null:

At an $\alpha = 0.05$ level of significance, there was insufficient evidence to reject the null that 47% of Santa Barbara residents have a Threads account, in favor of the alternative that the true proportion is not 47%.