Name: _____ **Date:** _____

PSTAT 5A: Homework 04

Summer Session A 2023, with *Ethan P. Marzban*

1. 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.

   Final Answer(s): No numerical answer.

   (b) Identify the sample.

   Final Answer(s): No numerical answer.

   (c) Define the parameter of interest. Use the notation discussed in Lecture 12.

   Final Answer(s): No numerical answer.

   (d) Define the random variable of interest. Use the notation discussed in Lecture 12.

   Final Answer(s): No numerical answer.

   (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!

   Final Answer(s): $\mathcal{N}(89,\ 4)$

   (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?

   Final Answer(s): 78.88%

2. 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.

Final Answer(s): No numerical answer.

(b) Identify the sample.

Final Answer(s): No numerical answer.

(c) Define the parameter of interest. Use the notation discussed in Lecture 12.

Final Answer(s): No numerical answer.

(d) Define the random variable of interest. Use the notation discussed in Lecture 12.

Final Answer(s): No numerical answer.

(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*?

Final Answer(s): Use the t_{51} distribution.

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

Final Answer(s): $[78.26147\ ,\ 87.73853]$

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.

Final Answer(s): No numerical answers.

(b) Define the random variable of interest.

Final Answer(s): No numerical answers.

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

Name: _____ **Date:** _____

Final Answer(s): No numerical answers.

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

Final Answer(s): -1.54

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

Final Answer(s): $\mathcal{N}(0,\ 1)$

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

Final Answer(s): 1.96

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

Final Answer(s): Fail to reject the null.

4. (**Deriving the Lower-Tailed Test of Proportions**). Consider testing the set of hypothesis

$$\begin{cases} H_0: & p = p_0 \\ H_A: & p < p_0 \end{cases}$$

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

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

(a) Show that $\text{TS} \overset{H_0}{\sim} \mathcal{N}(0,\ 1)$. If your answer depends on a set of conditions to be true, explicitly state those conditions.

Final Answer(s): No numerical answer.

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

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

for some constant c. As a hint, look up the logic we used in Lecture 13 to derive the two-tailed test, and think in terms of statements like "$\widehat{p}$ is *far away* from p_0". **You do not have to find the value of c in this part.**

Name:_____ **Date:**_____

Final Answer(s): No numerical answer.

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

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

where z_α denotes the $(\alpha) \times 100^{\text{th}}$ percentile of the standard normal distribution.

Final Answer(s): No numerical answer.

PLEASE NOTE: You may be expected to use this result on future homework/quizzes/exams.

