STAT 0116: Introduction to Statistical Sciences

Practice Exam 1 **Name:** _____

Fall 2026

Instructions

- This exam consists of 5 questions. The point value for each question is provided at the start of the prompt.
- You may use **one hand-written sheet of notes** (front and back).
- No phones or other electronic devices are permitted.
- Show all work clearly. Unsupported answers may not receive full credit.

I have neither given nor received unauthorized aid on this examination, nor have I concealed any violation of the Honor Code. I have not used any outside resources, including the internet or generative AI, to complete this exam, nor have I spoken about this exam with anyone else.

Signature: _____

Question	Points	Score
1	5	
2	6	
3	12	
4	12	
5	16	
Total	**51**	

1. [5 pt] In a coastal town, 36% of the residents fish regularly. Among those who fish regularly, 68% catch at least one fish on any given trip. For those who don't fish regularly, only 12% catch a fish if they happen to go out fishing. If a randomly selected person from the town is observed catching a fish, what is the probability that they are someone who fishes regularly?

2. [6 pt] In youth hockey in Canada, players are grouped into age divisions by calendar year, so a child born in January can be almost a full year older (and bigger) than a teammate born in December. Researchers studying this "relative age effect" suspected that players born early in the year are more likely to reach the professional level. To investigate, they took a random sample of 300 National Hockey League (NHL) players and recorded each player's birth month, position (forward, defense, or goalie), jersey number, and height (in inches). Of the 300 players, 105 were born in January, February, or March.

   a. [1 pt] Who or what are the observational units in this study?

   b. [2 pt] Classify each of the four variables the researchers recorded as categorical or quantitative.

   c. [2 pt] If birth month had nothing to do with reaching the NHL, about one quarter of players would be born in January–March. Write hypotheses, in words and in notation, that address the researchers' suspicion.

   d. [1 pt] Give the value of the statistic, with proper notation, and describe what it measures.

3. [12 pt] The amount of time Middlebury students spend waiting in line at the dining hall during the lunch rush is strongly right-skewed: most students wait only a few minutes, but a few wait much longer. Across all students, the mean wait time is $\mu = 6$ minutes and the standard deviation is $\sigma = 4$ minutes.

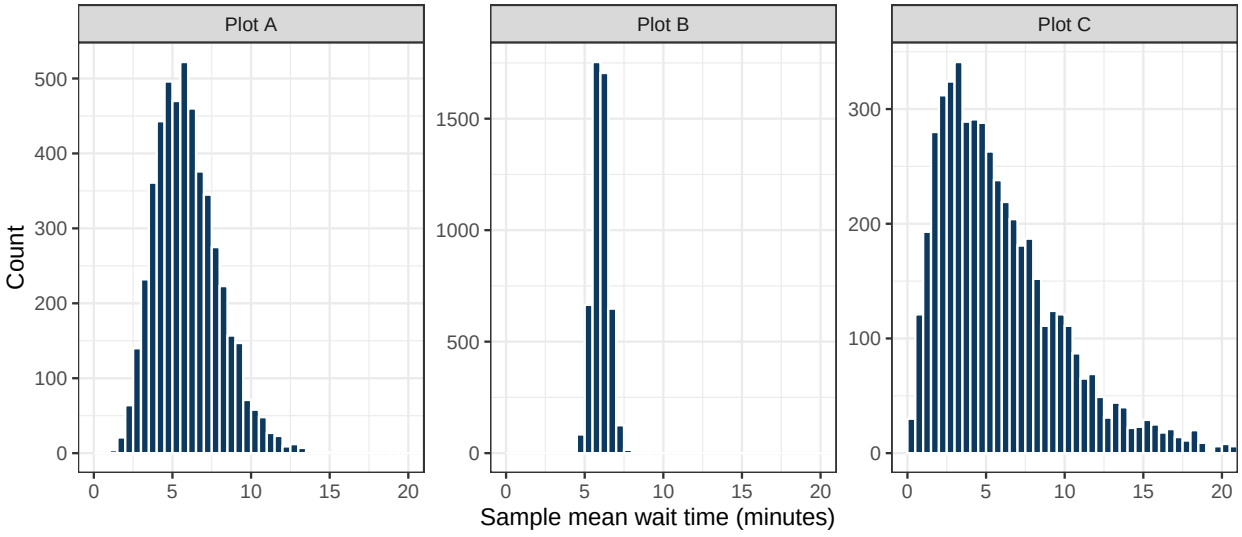
   a. [2 pt] Would it be appropriate to use a Normal distribution to find the probability that a single randomly selected student waits more than 10 minutes? Explain.

   b. [2 pt] Interpret the standard deviation in the context of the problem.

   c. [3 pt] Suppose we take a random sample of $n = 64$ students and compute their mean wait time, $\bar{x}$. Describe the sampling distribution of $\bar{x}$: give its shape, center, and spread, and justify your answer for the shape.

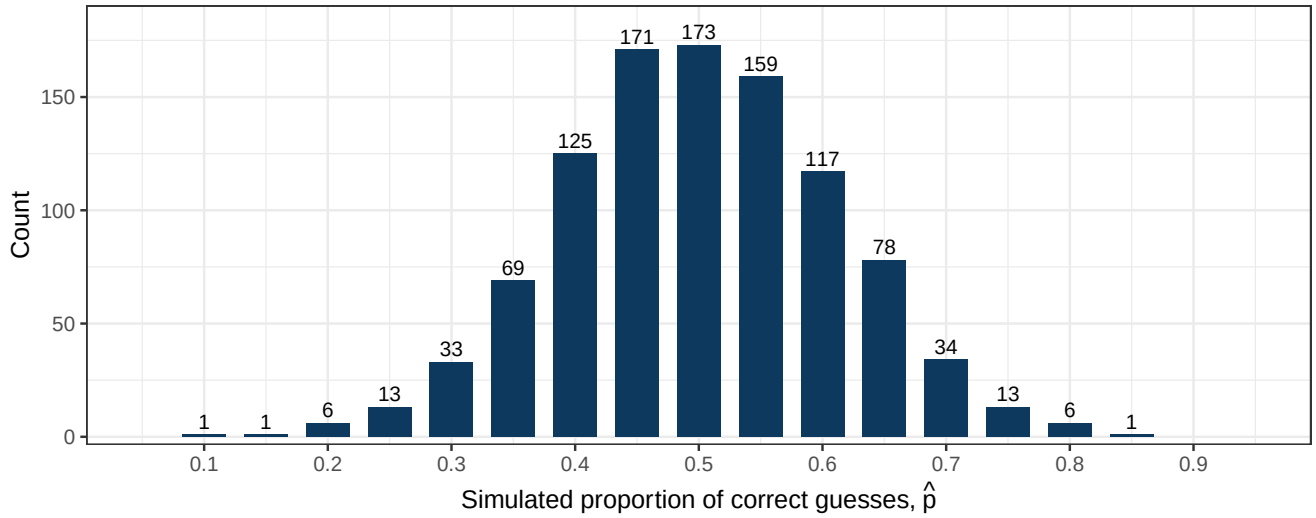
d. [2 pt] Between what two values would you expect the middle 95% of sample means (for samples of size 64) to fall?

e. [3 pt] The histograms below each show 5,000 sample means computed from repeated random samples from this population. One plot used samples of size $n = 1$, one used $n = 4$, and one used $n = 64$. Match each plot to its sample size and explain your reasoning.



4. [12 pt] A Middlebury student claims she can tell real Vermont maple syrup from imitation "pancake syrup" by taste alone. To test her claim, her roommate runs 20 blind trials. In each trial, the student is blindfolded and tastes two cups, one with each type of syrup, in an order decided by a coin flip. She then says which cup has the real maple syrup. She identifies the real maple syrup correctly in 15 of the 20 trials.

   The null distribution for this problem is provided below.



   a. [2 pt] Describe the parameter of interest in words, and calculate the observed statistic, $\hat{p}$.

   b. [2 pt] State the null and alternative hypotheses in words and in symbols.

c. [2 pt] Why is the simulated distribution centered at 0.5?

d. [2 pt] The null distribution is based on 1000 simulated samples. Find the p-value.

e. [2 pt] Interpret the p-value from part (d) in the context of this study.

f. [2 pt] State your conclusion in the context of the problem. Which type of error (Type I or Type II) could you have made? Describe what that error would mean here in the context of the problem..

5. [16 pt] In 2022, Middlebury's sustainability office found that 40% of students bought single-use bottled water at least once a week. Since then, the College has installed water-bottle refill stations in every academic building. This fall, the office surveyed a random sample of 150 Middlebury students, and 45 of them said they buy bottled water at least once a week. The office wants to know whether weekly bottled-water buying has become less common. It designed its test to have a significance level of $\alpha = 0.025$.

   a. [2 pt] What population does the sustainability office want to learn about? What is the parameter, and what statistic do we use to estimate it? Use notation, and give a value for anything you can compute.

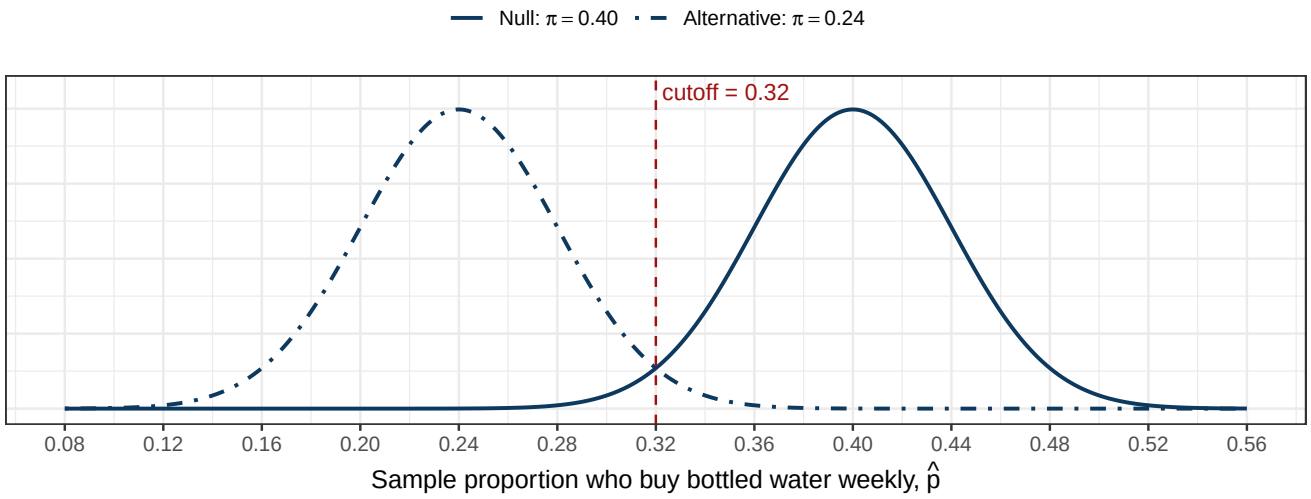
   b. [2 pt] Write the hypotheses in symbols, and explain in one sentence each what they claim.

   c. [2 pt] Can the Central Limit Theorem be used to approximate the null distribution of $\hat{p}$? Check the necessary conditions.

d. [3 pt] How many standard errors below the null value is the observed sample proportion? Show how you calculate the standard error.

e. [3 pt] Give a range of plausible values for the p-value. Then describe the strength of evidence against the null hypothesis, and state your conclusion in the context of the problem.

The plot below shows two possible sampling distributions of $\hat{p}$ for random samples of 150 students: one if H_0 is true ($\pi = 0.40$), and one if the true proportion has dropped to $\pi = 0.24$. Treat the standard error as approximately 0.04 for both. At $\alpha = 0.025$, the office rejects H_0 whenever $\hat{p}$ falls to the left of the dashed line.



f. [2 pt] On the plot, shade the area that represents the power of the test if $\pi = 0.24$. Then use the empirical rule to approximate the power.

g. [2 pt] Suppose the office had used $\alpha = 0.05$ instead of $\alpha = 0.025$. What would happen to the location of the cutoff, the power of the test, and the probability of a Type I error? Then name one change, other than changing α, that would increase the power of the test.