

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. According to a market research firm, the average duration of ownership for a laptop in the United States this year (2023) is estimated to be 4.5 years. To test this claim, Maria takes a representative sample of 50 people and notes that the average laptop ownership duration of these 50 people is 4.8 years, with a standard deviation of 0;75 years. Suppose Maria wishes to use a 5% level of significance to test that the average duration of ownership for a laptop in the United States is longer than 4.5 years.
 - a. State the null and alternative hypotheses. Be sure to use the proper notation and define your parameter.

b. Check the necessary conditions for this test.

c. Calculate the test statistic.

d. Write the Python code to calculate the p-value for this test.

e. Suppose the p-value is calculated to be 0.006. What would you tell Maria about their finding?

f. Calculate the 90% confidence interval. Remember to include interpretation.

- 2. Michael wondered whether apartments built in Austin in the 1990s are larger, on average, than apartments built in Austin in the 2020s. He pulled data from the Austin property search website, recording the living area (in square feed) and the decade the apartment was built (1990s or 2020s) for a random sample of 60 apartments built in Austin in the 1990s and another random sample of 40 apartments built in Austin in the 2020s. Michael's data set has two variables: living area in square feet and decade (which has two values: 1990s and 2020s. Suppose the mean living area in square feet of apartments from the 1990s is 1917.9 sqft with standard deviation of 622.5 sqft, and the mean living area in square feet of apartments from 2020s is 1737 sqft with standard deviation of 640.2 sqft.
 - a. State the null and alternative hypotheses. Be sure to use the proper notation and define your parameter.

b. Check the necessary conditions for this test.

c. Calculate the test statistic.

d. Write the Python code to calculate the p-value for this test.

e. Suppose the p-value is calculated to be 0.06. What would you tell Michael about their finding at the 5% significance level?

f. Would your conclusion change if Michael tested whether the average living area of the apartments from the 1990s is different from the average living area of those from the 2020s?

g. Calculate the 85% confidence interval. Remember to include interpretation.

3. Marissa would like to determine whether or not there is a significant difference between the average price of an apartment in Santa Barbara and the average price of an apartment in Los Angeles. Her initial beliefs are that the average prices in these two cities are the same. To test this claim, she takes a representative sample of 20 Santa Barbara apartments and another representative sample of 25 Los Angeles apartments; her data is summarized below (measurements are reported in thousands of dollars per month):

	Sample Mean	Sample Std. Dev.
Santa Barbara	2.69	0.67
Los Angeles	2.71	0.55

Assume that all independence and normality assumptions are met. Additionally, suppose Marissa decides to label "households in Santa Barbara" as Population 1 and "households in Los Angeles" as Population 2. Finally, assume Marissa wants to test her initial assumption against a two-sided alternative.

a. Define the parameters of interest, μ_1 and μ_2 .

b. State the null and alternative hypotheses.

c. Compute the observed value of the test statistic.

d. Assuming the null is correct, what is the approximate distribution of the sampling distribution? Be sure to include any / all relevant parameters.

e. What is the p-value of the observed test statistic? (You may need to use Python for this part).

f. What is the critical value of the test, if we use an $\alpha = 0.05$ level of significance? (You can use Python but you don't need to).

g. Now, carry out the test at $\alpha = 0.05$ level of significance. Be sure to phrase your conclusions in terms of the context of the problem.

h. Calculate the 95% confidence interval and include interpretation.

- 4. Faculty and students are all concerned about the amount of money that students need to spend on textbooks. In this question, we will examine the differences between the cost of new textbooks available at the UCLA bookstore and on Amazon, and whether a difference exists.
 - a. Define your parameter of interest.

b. Provided a dataset of 68 books with the following columns: prices of the book from UCLA and prices of the book from Amazon. Describe how you will calculate the sample statistic.

c. State the null and alternative hypotheses. Be sure to use the proper notation and define your parameter.

d. Check the necessary conditions for this test.

e. Suppose the observed mean is 3.5832 with standard deviation of 13.423. Calculate the test statistic.

f. Write the Python code to calculate the p-value for this test.

g. The p-value for the test is 0.031. What is the conclusion for our hypothesis test at 5% significance level?

h. Would your conclusion change if instead, you want to run a one-sided alternative?

5. Consider the following two sets of numbers:

$$X = \{1, 2, 4, 4, 6, 5, 3\}$$

$$Y = \{3, 4, 1, 4, 4, 2, 1\}$$

Compute the correlation between x and y by hand. You may use Python to check your work but you must show your work.