Mathematics for Data Science 2023/2024

Week 4- Inferential Statistics

Answer the following questions

Q1. Some of the following statements refer to the null hypothesis, some to the alternate hypothesis.

State the null hypothesis, $H0$ and the alternative hypothesis, Ha , in terms of the appropriate parameter (μ or p).

a. The Mean number of cars a person owns in her lifetime is not more than ten.
b. About half of Americans prefer to live away from cities, given the choice.
c. Europeans have a mean paid vacation each year of six weeks.
d. Private universities' mean tuition cost is more than $20,000 per year.

Q2. When using the *p*-value to evaluate a hypothesis test, it is sometimes useful to use the following memory device

If the *p*-value is low, the null must go.

If the *p*-value is high, the null must fly.

This memory aid relates a *p*-value less than the established alpha (the *p* is low) as rejecting the null hypothesis and, likewise, relates a *p*-value higher than the established alpha (the *p* is high) as not rejecting the null hypothesis.

Problem

Fill in the blanks.

Reject the null hypothesis when _____.

The results of the sample data _____.

Do not reject the null when hypothesis when _____.

The results of the sample data _____.

Q3. It's a Boy Genetics Labs claim their procedures improve the chances of a boy being born. The results for a test of a single population proportion are as follows:

H_0: $p = 0.50$, H_a: $p > 0.50$

$\alpha = 0.01$

p-value = 0.025

Interpret the results and state a conclusion in simple, non-technical terms.

Q4. The cost of a daily newspaper varies from city to city. However, the variation among prices remains steady with a standard deviation of 20¢. A study was done to test the claim that the mean cost of a daily newspaper is $1.00. Twelve costs yield a mean cost of 95¢ with a standard deviation of 18¢. Do the data support the claim at the 1% level?

Q5. The National Institute of Mental Health published an article stating that in any one-year period, approximately 9.5 percent of American adults suffer from depression or a depressive illness. Suppose that in a survey of 100 people in a certain town, seven of them suffered from depression or a depressive illness. Conduct a hypothesis test to determine if the true proportion of people in that town suffering from depression or a depressive illness is lower than the percent in the general adult American population.

a. Is this a test of one mean or proportion?
b. State the null and alternative hypotheses.
   H_0: _____ H_a: _____
c. Is this a right-tailed, left-tailed, or two-tailed test?
d. What symbol represents the random variable for this test?
e. In words, define the random variable for this test.
f. Calculate the following:
   i. x = _____
   ii. n = _____
   iii. p′ = _____
g. Calculate σ_x = _____. Show the formula set-up.
h. State the distribution to use for the hypothesis test.
i. Find the p-value.
j. At a pre-conceived $\alpha = 0.05$, what is your:
   i. Decision:
   ii. Reason for the decision:
   iii. Conclusion (write out in a complete sentence):

Q6. The standard deviation of the weights of elephants is known to be approximately 15 pounds. We wish to construct a 95% confidence interval for the mean weight of newborn elephant calves. Fifty newborn elephants are weighed. The sample mean is 244 pounds. The sample standard deviation is 11 pounds.

1. Identify the following:
    a. $\bar{x}$ = _____
    b. σ = _____
    c. n = _____

2. In words, define the random variables X and $\bar{X}$.

3. Which distribution should you use for this problem?

Q7. The U.S. Census Bureau conducts a study to determine the time needed to complete the short form. The Bureau surveys 200 people. The sample mean is 8.2 minutes. There is a known standard deviation of 2.2 minutes. The population distribution is assumed to be normal.

Identify the following:

a. $\bar{x}$ = _____
b. σ = _____
c. n = _____

Construct a 90% confidence interval for the population mean time to complete the forms. State the confidence interval, _____

Q8. A hospital is trying to cut down on emergency room wait times. It is interested in the amount of time patients must wait before being called back to be examined. An investigation committee randomly surveyed 70 patients. The sample mean was 1.5 hours with a sample standard deviation of 0.5 hours.

Identify the following:

a. $\bar{x}$ =_____
b. sx=_____
c. n =_____
d. $n - 1$ =_____

Construct a 95% confidence interval for the population mean time spent waiting. State the confidence interval _____