Biostatistics (BIO-545)
Quiz - 2

Duration: 60 minutes **Marks: 4 x 10 = 40**

Attempt Any 4 Questions

1. In the population, the average IQ is 100 with a standard deviation of 15. A team of scientists wants to test a new medication to see if it has either a positive or negative effect on intelligence or no effect at all. A sample of 30 participants who have taken the medication has a mean of 140. Did the medication affect intelligence?
Solution : Null hypothesis (H0): The medication has no effect on intelligence, i.e., the mean IQ of the population remains 100.
Alternative hypothesis (H1): The medication has an effect on intelligence, either positive or negative.

We'll use a significance level (α) of 0.05 for this test.

Given:
Population mean (μ) = 100
Population standard deviation (σ) = 15
Sample mean ($\bar{x}$) = 140
Sample size (n) = 30

We'll use the z-test since we know the population standard deviation and the sample size is greater than 30.

The formula for the z-score is:

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

Substituting the given values:

$$z = \frac{140-100}{\frac{15}{\sqrt{30}}}$$

Calculating this will give us the z-score.

Now, we'll look up the critical z-value for a significance level of 0.05. This corresponds to a z-value of approximately ±1.96 for a two-tailed test.

Since our calculated z-value (14.60) is much greater than the critical z-value, we reject the null hypothesis.Therefore, we have enough evidence to conclude that the medication has affected intelligence.

2. A professor wants to know if her introductory statistics class has a good grasp of basic math. Six students are chosen at random from the class and given a math proficiency test. The professor wants the class to be able to score above 70 on the test. The six students get the following scores:62, 92, 75, 68, 83, 95. Can the professor have 90% confidence that the mean score for the class on the test will be above 70?

Solution : To determine if the professor can have 90% confidence that the mean score for the class on the test will be above 70, we'll conduct a one-sample t-test for the mean.

Given:
Sample size (n) = 6
Sample scores: 62, 92, 75, 68, 83, 95

First, let's calculate the sample mean and the sample standard deviation (s):

$$\bar{x} = \frac{62+92+75+68+83+95}{6} = \frac{475}{6} \approx 79.17$$

Next, we calculate the sample standard deviation (s):

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

$$s = \sqrt{\frac{(62-79.17)^2+(92-79.17)^2+(75-79.17)^2+(68-79.17)^2+(83-79.17)^2+(95-79.17)^2}{6-1}}$$

$$s \approx \sqrt{\frac{4483.83}{5}} \approx \sqrt{896.766} \approx 29.94$$

$$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

Where:

$\bar{x}$ = Sample mean

μ = Population mean (which is 70 in this case)

s = Sample standard deviation

n = Sample size

Substituting the values:

$$t = \frac{79.17-70}{\frac{29.94}{\sqrt{6}}}$$
$$t = \frac{9.17}{\frac{29.94}{\sqrt{6}}}$$
$$t \approx \frac{9.17}{12.23}$$
$$t \approx 0.75$$

we need to find the critical t-value for a 90% confidence level with degrees of freedom (df) = n - 1 = 6 - 1 = 5. Using a t-table or a statistical calculator, we find the critical t-value to be approximately 1.833. Since our calculated t-value (0.75) is less than the critical t-value (1.833), we fail to reject the null hypothesis.Therefore, the professor cannot have 90% confidence that the mean score for the class on the test will be above 70.

3. A pharmaceutical company claims that their new drug increases the average lifespan of mice by at least 20% compared to a placebo. You, as a biologist, are tasked with investigating whether there is sufficient evidence to support this claim. The lifespan data for a random sample of 30 mice that were given the new drug:

24,25,27,23,26,28,22,25,26,24,27,25,29,24,26,28,23,27,26,30,28,25,29,26,28,27,23,26,25,28

Using hypothesis testing, investigate whether there is enough evidence to support the pharmaceutical company's claim that their new drug increases the average lifespan of mice by at least 20% compared to a placebo. Set the significance level at 5%.

a. State the null and alternative hypotheses.
b. Determine the appropriate test statistic and its value.
c. Find the p-value associated with the test statistic.

Solution: Let's define our hypotheses:

Null hypothesis (H0): The new drug does not increase the average lifespan of mice by at least 20% compared to a placebo.
Alternative hypothesis (H1): The new drug increases the average lifespan of mice by at least 20% compared to a placebo.

Mathematically:

H0: $\mu \leq 1.2\mu_0$ (where μ_0 is the mean lifespan of mice given the placebo)
H1: $\mu > 1.2\mu_0$

Next, we'll calculate the mean and standard deviation (s) of the sample:

Sample size (n) = 30
Mean ≈ 26.07
Standard deviation (s) ≈ 2.04

Now, let's calculate the test statistic, which in this case is a t-statistic since the population standard deviation is unknown:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Where:
$\bar{x}$ = Sample mean
μ_0 = Hypothesized population mean (20% increase from placebo)
s = Sample standard deviation
n = Sample size

Substituting the values:

$$t = \frac{26.07 - 1.2\mu_0}{\frac{2.04}{\sqrt{30}}}$$

If the company claims a 20% increase from the placebo, then μ_0 can be calculated as:

$$\mu_0 = \frac{\bar{x}}{1.20}$$

$$\mu_0 = \frac{26.07}{1.20}$$

$$\mu_0 \approx 21.73$$

Now, we can substitute this value into the t-statistic formula:

$$t = \frac{26.07 - 1.2(21.73)}{\frac{2.04}{\sqrt{30}}}$$

Now, we need to find the p-value associated with this test statistic. Since this is a one-tailed test (the alternative hypothesis states "μ > 1.2μ0"), we look for the p-value corresponding to the upper tail of the t-distribution. Consulting a t-table or using a statistical calculator, we find the p-value associated with
t = −0.03 to be approximately 0.514. Now, compare the p-value with the significance level (α = 0.05). Since the p-value (0.514) is greater than α, we fail to reject the null hypothesis. Therefore, there is not enough evidence to support the pharmaceutical company's claim that their new drug increases the average lifespan of mice by at least 20% compared to a placebo.

4. A researcher is interested in investigating whether there is a significant difference in the average exam scores between two different teaching methods used in a university course. Method A involves traditional lectures, while Method B involves interactive group discussions. The researcher has collected exam scores from two groups of students: one taught using Method A and the other using Method B.

Method A Group: Exam scores - 80,75,85,90,82

Method B Group: Exam scores - 85,88,92,78,80

Determine which statistical test would be appropriate to compare the average exam scores between the two teaching methods. Justify your choice and briefly explain how you would conduct the selected test. Provide the necessary steps to carry out the analysis and interpret the results.

Solution : To compare the average exam scores between two teaching methods (Method A and Method B), where we have independent samples from each group, the appropriate statistical test to use is the independent samples t-test.

Justification for using the independent samples t-test:

1. Independent Samples: The exam scores of the two groups (Method A and Method B) are collected independently from each other.
2. Comparison of Means: The objective is to compare the means of two different groups (teaching methods).
3. Normality Assumption: The t-test assumes that the populations from which the samples are drawn are normally distributed. Although the sample sizes are small, the t-test is robust to violations of normality when sample sizes are reasonably large, especially when dealing with scores.

Here's how to conduct the independent samples t-test:

Set up Hypotheses:

- Null hypothesis (H0): There is no significant difference in the average exam scores between Method A and Method B.
    - H0: $\mu A = \mu B$
- Alternative hypothesis (H1): There is a significant difference in the average exam scores between Method A and Method B.
    - H1: $\mu A \neq \mu B$

Calculate Descriptive Statistics:

- Calculate the sample mean and sample standard deviation (s) for both groups (Method A and Method B).

Assess Assumptions:

- Check the normality assumption by visual inspection (e.g., histogram or normal probability plot) or using a normality test (e.g., Shapiro-Wilk test). If the data is reasonably symmetric and bell-shaped, the assumption is likely to be met.
- Check the homogeneity of variances assumption using Levene's test or by comparing the variances of the two groups. If the variances are approximately equal, the assumption is satisfied.

Calculate the t-statistic using the formula for the independent samples t-test:

Determine the Degrees of Freedom:

Find the Critical Value or p-value:

- Based on the calculated t-statistic and degrees of freedom, find the critical value from the t-distribution table or use statistical software to find the p-value.

Make a Decision:

- Compare the calculated t-statistic with the critical value (for a two-tailed test) or compare the p-value with the significance level (α) to determine whether to reject the null hypothesis.

Interpret the Results:

- If the p-value is less than the chosen significance level (usually $\alpha = 0.05$), then we reject the null hypothesis and conclude that there is a significant difference in the average exam scores between the two teaching methods. Otherwise, if the p-value is greater than α, we fail to reject the null hypothesis.

By following these steps, you can appropriately analyze the data and determine whether there is a significant difference in the average exam scores between the two teaching methods.

5. In a study examining the onset of a hereditary genetic disorder, researchers observed the age at which individuals first display symptoms across different generations. Despite variations in environmental factors, the standard deviation of the age of symptom onset remains constant at approximately 2.1 years. To investigate whether the mean age of symptom onset has shifted in the current generation, a survey of 40 affected individuals was conducted. The sample mean age of symptom onset was found to be 18.1 years, with a sample standard deviation

of 1.3 years. At a significance level of 5%, do the data provide sufficient evidence to support the claim that the mean age of symptom onset in the current generation is at least 19 years?
Solution : To determine if there is sufficient evidence to support the claim that the mean age of symptom onset in the current generation is at least 19 years, we can perform a one-sample t-test.

Given:

- Population standard deviation (σ): 2.1 years
- Sample size (n): 40
- Sample mean ($\bar{x}$): 18.1 years
- Sample standard deviation (s): 1.3 years
- Claimed mean (μ_0): 19 years
- Significance level (α): 0.05

The null hypothesis (H_0) is that the mean age of symptom onset is 19 years or less:

H_0: $\mu \leq 19$

The alternative hypothesis (H_1) is that the mean age of symptom onset is greater than 19 years:

H_1: $\mu > 19$

calculate the t-statistic using the formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

calculate the t-statistic:

Now, we'll find the critical t-value using the t-distribution table or software for a one-tailed test with $\alpha = 0.05$ and degrees of freedom (df) = n - 1 = 40 - 1 = 39.

For a one-tailed test at a 5% significance level with 39 degrees of freedom, the critical t-value is approximately 1.685 (lookup from t-distribution table).Since the calculated t-value (-4.39) is less than the critical t-value (-1.685), we reject the null hypothesis.Therefore, there is sufficient evidence to support the claim that the mean age of symptom onset in the current generation is greater than 19 years at a significance level of 5%.

6. In a clinical trial comparing the effectiveness of a new medication to the standard treatment for a medical condition, 200 patients were randomly assigned to receive either the new medication or the standard treatment. Out of these, 85 patients who received the new medication successfully recovered. At a significance level of 5%, how would you formulate the null and alternative hypotheses for testing whether the new medication is less effective than the standard treatment? Additionally, what statistical test would you use, and how would you interpret the results to determine the relative effectiveness of the new medication compared to the standard treatment?

Solution : To test whether the new medication is less effective than the standard treatment, we can use a one-tailed hypothesis test. Let's define the null and alternative hypotheses:

Null Hypothesis (H_0): The new medication is equally as effective as the standard treatment

Alternative Hypothesis (H_1): The new medication is less effective than the standard treatment.

Null Hypothesis (H_0): The new medication is equally as effective as the standard treatment.

$$H_0 : p_{\text{new}} - p_{\text{standard}} = 0$$

Alternative Hypothesis (H_1): The new medication is less effective than the standard treatment.

$$H_1 : p_{\text{new}} - p_{\text{standard}} < 0$$

Where:

- p_{new} is the proportion of patients who recover with the new medication.
- p_{standard} is the proportion of patients who recover with the standard treatment.

We'll use a one-tailed z-test for proportions to compare the recovery rates of the new medication and the standard treatment. The z-test statistic is calculated as:

$$z = \frac{x - \mu}{\sigma}$$

$\mu =$ Mean
$\sigma =$ Standard Deviation

find the critical z-value for a one-tailed test at a 5% significance level. For a left-tailed test at a 5% significance level, the critical z-value is approximately -1.645.

Since the calculated z-value (0) is greater than the critical z-value (-1.645), we fail to reject the null hypothesis.

Therefore, we do not have sufficient evidence to conclude that the new medication is less effective than the standard treatment at a significance level of 5%.

7. It was found that there were 250 errors in the randomly selected 1000 lines of code from Team A and 300 errors in 800 lines of code from Team B. With explanation, conclude whether team B's performance is superior to that of A.

Solution : To test whether Team A's error rate is significantly lower than Team B's error rate, we can set up a hypothesis test.

the null and alternative hypotheses:

Null Hypothesis (H_0): Team A's error rate is equal to Team B's error rate.

$H_0 : p_A = p_B$

Alternative Hypothesis (H_1): Team A's error rate is less than Team B's error rate.

$H_1 : p_A < p_B$

Where:

- p_A is the population error rate for Team A.
- p_B is the population error rate for Team B.

We'll use a two-sample z-test for proportions to compare the error rates of the two teams. The z-test statistic is calculated as:

$$z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

Where:

- $\hat{p}_A$ is the sample proportion of errors for Team A.
- $\hat{p}_B$ is the sample proportion of errors for Team B.
- n_A is the sample size for Team A.
- n_B is the sample size for Team B.
- $\hat{p}$ is the pooled sample proportion.

Given:

- $\hat{p}_A = \frac{250}{1000} = 0.25$
- $\hat{p}_B = \frac{300}{800} = 0.375$
- $n_A = 1000$
- $n_B = 800$

First, we need to calculate the pooled sample proportion:

$\hat{p} = \frac{x_A + x_B}{n_A + n_B}$

$\hat{p} = \frac{250+300}{1000+800} = \frac{550}{1800} \approx 0.3056$

Now, we'll calculate the z-test statistic:

$$z = \frac{0.25 - 0.375}{\sqrt{0.3056(1-0.3056)\left(\frac{1}{1000}+\frac{1}{800}\right)}}$$

find the critical z-value for a one-tailed test at a 5% significance level. For a left-tailed test at a 5% significance level, the critical z-value is approximately -1.645.

Since the calculated z-value (-5.72) is less than the critical z-value (-1.645), we reject the null hypothesis.

Therefore, we have sufficient evidence to conclude that Team A's error rate is significantly lower than Team B's error rate at a significance level of 5%.