STATISTICS WORK SHEET10- SET 03

<u>Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.</u>

1. Rejection of the null hypothesis is a conclusive proof that the alternative hypothesis is

a. True b. False c. Neither

Answer: b. False

2. Parametric test, unlike the non-parametric tests, make certain assumptions about

a. The population size b. The underlying distribution c. The sample size

Answer: b. The underlying distribution

3. The level of significance can be viewed as the amount of risk that an analyst will accept when making a decision

a. True b. False

Answer: a. True

4. By taking a level of significance of 5% it is the same as saying

a. We are 5% confident the results have not occurred by chance

b. We are 95% confident that the results have not occurred by chance

c. We are 95% confident that the results have occurred by chance

Answer: b. We are 95% confident that the results have not occurred by chance

5. One or two tail test will determine

a. If the two extreme values (min or max) of the sample need to be rejected

b. If the hypothesis has one or possible two conclusions

c. If the region of rejection is located in one or two tails of the distribution

Answer: c. If the region of rejection is located in one or two tails of the distribution

6. Two types of errors associated with hypothesis testing are Type I and Type II. Type II error is committed when

a. We reject the null hypothesis whilst the alternative hypothesis is true

b. We reject a null hypothesis when it is true

c. We accept a null hypothesis when it is not true

Answer: b. We reject a null hypothesis when it is true

7. A randomly selected sample of 1,000 college students was asked whether they had ever used the drug Ecstasy. Sixteen percent (16% or 0.16) of the 1,000 students surveyed said they had. Which one of the following statements about the number 0.16 is correct?

a. It is a sample proportion. b. It is a population proportion.

c. It is a margin of error. d. It is a randomly chosen number.

Answer: a. It is a sample proportion.

8. In a random sample of 1000 students, pˆ = 0.80 (or 80%) were in favour of longer hours at the school library. The standard error of pˆ (the sample proportion) is

a. 0.013 b. 0.160 c. 0.640 d. 0.800

Answer: a. 0.013

The standard error of the sample proportion is calculated as:

SE = sqrt[(p * (1 - p)) / n]

p is the sample proportion, n is the sample size

SE = sqrt[(0.80 * (1 - 0.80)) / 1000] = 0.013

9. For a random sample of 9 women, the average resting pulse rate is x = 76 beats per minute, and the sample standard deviation is s = 5. The standard error of the sample mean is

a. 0.557 b. 0.745 c. 1.667 d. 2.778

Answer: c. 1.667

SE = s / sqrt(n)

where s is the sample standard deviation and n is the sample size.

SE = 5 / sqrt(9) = 1.667

10. Assume the cholesterol levels in a certain population have mean μ= 200 and standard deviation σ = 24. The cholesterol levels for a random sample of n = 9 individuals are measured and the sample mean x is determined. What is the z-score for a sample mean x = 180?

a. –3.75 c. –2.50 c. –0.83 d. 2.50

Answer: c. –0.83

z = (x - μ) / (σ / sqrt(n))

where x is the sample mean, μ is the population mean, σ is the population standard deviation, and n is the sample size.

z = (180 - 200) / (24 / sqrt(9)) = -5 / 8 = -0.625 closest to option (c)

11. In a past General Social Survey, a random sample of men and women answered the question “Are you a member of any sports clubs?” Based on the sample data, 95% confidence intervals for the

population proportion who would answer "yes" are .13 to .19 for women and .247 to .33 for men. Based on these results, you can reasonably conclude that

a. At least 25% of American men and American women belong to sports clubs.

b. At least 16% of American women belong to sports clubs.

c. There is a difference between the proportions of American men and American women who belong to sports clubs.

d. There is no conclusive evidence of a gender difference in the proportion belonging to sports clubs.

Answer: d. There is no conclusive evidence of a gender difference in the proportion belonging to sports clubs

12. Suppose a 95% confidence interval for the proportion of Americans who exercise regularly is 0.29 to 0.37. Which one of the following statements is FALSE?

a. It is reasonable to say that more than 25% of Americans exercise regularly.

b. It is reasonable to say that more than 40% of Americans exercise regularly.

c. The hypothesis that 33% of Americans exercise regularly cannot be rejected.

d. It is reasonable to say that fewer than 40% of Americans exercise regularly.

Answer: b. It is reasonable to say that more than 40% of Americans exercise regularly.

<u>Q13 to Q15 are subjective answers type questions. Answers them in their own words briefly.</u>

13. How do you find the test statistic for two samples?

Answer:

The test statistic for comparing two samples depends on the type of hypothesis test being performed. Here are some commonly used test statistics for two-sample hypothesis tests:

1. For a two-sample t-test, the test statistic is calculated as follows:

   t = (x1 - x2) / SE

   where x1 and x2 are the sample means, and SE is the standard error of the difference between the means. The formula for SE depends on whether the samples are assumed to have equal variances or not. If they are assumed to have equal variances, then SE is calculated as:

   SE = sqrt((s1^2 / n1) + (s2^2 / n2))

   If they are not assumed to have equal variances, then SE is calculated as:

   SE = sqrt((s1^2 / n1) + (s2^2 / n2))

2. For a two-sample z-test, the test statistic is calculated as follows:

   z = (x1 - x2) / SE

where x1 and x2 are the sample means, and SE is the standard error of the difference between the means. The formula for SE is the same as for the t-test above.

3. For a chi-square test of independence, the test statistic is calculated by comparing the observed and expected frequencies in a contingency table. The formula for the chi-square statistic is:

   chi-square = sum((O - E)^2 / E)

   where O is the observed frequency and E is the expected frequency for each cell in the contingency table.

   In all cases, the calculated test statistic is then compared to a critical value from a distribution (t-distribution, standard normal distribution, or chi-square distribution) to determine whether to reject or fail to reject the null hypothesis.

14. How do you find the sample mean difference?

Answer:

The sample mean difference is the difference between the means of two samples. To find the sample mean difference, follow these steps:

1. Take a random sample from the first population of interest and record the sample mean (x1).
2. Take a random sample from the second population of interest and record the sample mean (x2).
3. Calculate the difference between the two sample means:

Sample mean difference = x1 - x2

For example, suppose you want to compare the heights of male and female college students. You take a random sample of 50 male students and record their average height as 70 inches. Then, you take a random sample of 50 female students and record their average height as 65 inches. The sample mean difference is:

sample mean difference = x1 - x2 = 70 - 65 = 5 inches

This means that, on average, male students in this sample are 5 inches taller than female students. Note that the sample mean difference can be positive or negative, depending on which group has the higher mean.

15. What is a two sample t test example?

Answer:

A two-sample t-test is used to compare the means of two independent samples. Here is an example:

Suppose a company is testing a new manufacturing process for producing a certain product. The company produces two samples of the product using the old process and the new process, respectively. The samples are of equal size, say n1 = n2 = 25. The company wants to know if the new process produces a significantly different mean product weight than the old process.

The null hypothesis is that there is no difference between the means of the two populations. The alternative hypothesis is that the means are different.

To perform a two-sample t-test, we need to calculate the t-statistic, which is given by:

$$t = (x1 - x2) / (s1^2/n1 + s2^2/n2)^{0.5}$$

where x1 and x2 are the sample means, s1 and s2 are the sample standard deviations, and n1 and n2 are the sample sizes.

Suppose the sample means and standard deviations are as follows:

Old process: x1 = 10.2, s1 = 0.8 New process: x2 = 10.5, s2 = 1.2

The t-statistic is then calculated as:

$$t = (10.5 - 10.2) / ((0.8^2/25) + (1.2^2/25))^{0.5} = 2.2$$

The degrees of freedom for the test are df = n1 + n2 - 2 = 48. Using a t-distribution table with df = 48 and a significance level of 0.05, we find the critical value to be 2.01.

Since the calculated t-value (2.2) is greater than the critical value (2.01), we reject the null hypothesis and conclude that there is a statistically significant difference between the means of the two populations. This means that the new manufacturing process does produce a significantly different mean product weight than the old process.