STATISTICS WORKSHEET-10

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

1. Rejection of the null hypothesis is a conclusive proof that the alternative hypothesis is a. True b. False c. Neither

Ans: c. Neither

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

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

Ans: 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

Ans: 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

Ans: 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

Ans: 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

Ans: b. c. We accept a null hypothesis when it is not 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.

Ans: 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. .013 b. .160 c. .640 d. .800

Ans: a. .013

The standard error of p = $\sqrt{[p (1-p) / n)]}$

p = 0.8

1 - p = 1 - 0.8 = 0.2

n = 1000

=> The standard error of p = $\sqrt{[0.8 (0.2) / 1000)]}$

=> The standard error of p ≈ 0.01265

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

Ans: c. 1.667

Given information

Sample size(n)=9

Satndard deviation(s)=5

The standard error of the sample mean is calculated as follow.

$SE=s/\sqrt{n}=5/\sqrt{9}=1.667$

The standard error of the sample mean is **1.667**

10. Assume the cholesterol levels in a certain population have mean $\mu = 200$ and standard deviation $\sigma = 24$. 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 b. –2.50 c. −0.83 d. 2.50

Ans: c. −0.83

Solution:

A Z-score is defined as the fractional representation of **data point** to the mean using **standard deviations**.

Given the **mean** $\mu = 200$ and **standard deviation** $\sigma = 24$. The cholesterol levels for a random sample of $n = 9$ individuals are measured **sample mean** is 180.

To determine the **Z-score** for the given **sample mean**

⇒ z-score = (X-ц)/σ

Here,

ц = 200

σ = 24

X = 180

Substitute the **values** in the above formula,

⇒ z-score = (180 - 200)/24

⇒ z-score = -20 /24

⇒ z-score = -0.833

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.

Ans: c. There is a difference between the proportions of American men and American women who belong 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.

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

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

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

Ans: The test statistic for a two-sample independent *t*-test is calculated by taking the difference in the two sample means and dividing by either the pooled or unpooled estimated standard error. The estimated standard error is an aggregate measure of the amount of variation in both groups.

Relevant Equations:

<u>Degrees of freedom</u>: Varies by conditions, but the basic rule of thumb for hand calculations is the smaller of $n_1 - 1$ and $n_2 - 1$, where n is the sample size for each group.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SE}$$

Generally, the test statistic is calculated as **the pattern in your data (i.e. the correlation between variables or difference between groups) divided by the variance in the data (i.e. the standard deviation)**.

14. How do you find the sample mean difference?

1. Ans: The expected value of the difference between all possible sample means is equal to the difference between population means. Thus, ...
2. The standard deviation of the difference between sample means (σ_d) is approximately equal to: $\sigma_d = \text{sqrt}(\sigma_1^2 / n_1 + \sigma_2^2 / n_2)$

15. What is a two sample t test example?

Ans: The two-sample *t*-test (also known as the independent samples *t*-test) is a method used to test whether the unknown population means of two groups are equal or not.

One way to measure a person's fitness is to measure their body fat percentage. Average body fat percentages vary by age, but according to some guidelines, the normal range for men is 15-20% body fat, and the normal range for women is 20-25% body fat.

Our sample data is from a group of men and women who did workouts at a gym three times a week for a year. Then, their trainer measured the body fat. The table below shows the data.

Table 1: Body fat percentage data grouped by gender

Group	Body Fat Percentages				
Men	13.3	6.0	20.0	8.0	14.0
	19.0	18.0	25.0	16.0	24.0
	15.0	1.0	15.0		
Women	22.0	16.0	21.7	21.0	30.0
	26.0	12.0	23.2	28.0	23.0

You can clearly see some overlap in the body fat measurements for the men and women in our sample, but also some differences. Just by looking at the data, it's hard to draw any solid conclusions about whether the underlying populations of men and women at the gym have the same mean body fat. That is the value of statistical tests – they provide a common, statistically valid way to make decisions, so that everyone makes the same decision on the same set of data values.