STATISTICS WORKSHEET-3

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

Ans- a. True

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

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

Ans- a. True

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

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

5. One or two tail test will determine

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

Ans- 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?

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

Ans- a. .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

Ans- c. 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 $x = 180$?

Ans- c. −0.83

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

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?

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

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

Ans- A two sample t-test is used to determine whether or not two population means are equal.

Two Sample t-test: Formula

A two-sample t-test always uses the following null hypothesis:

H0: $\mu 1 = \mu 2$ (the two population means are equal)

The alternative hypothesis can be either two-tailed, left-tailed, or right-tailed:

H1 (two-tailed): $\mu 1 \neq \mu 2$ (the two population means are not equal)

H1 (left-tailed): $\mu 1 < \mu 2$ (population 1 mean is less than population 2 mean)

H1 (right-tailed): $\mu 1 > \mu 2$ (population 1 mean is greater than population 2 mean)

We use the following formula to calculate the test statistic t:

Test statistic: (x1 – x2) / sp(√1/n1 + 1/n2)

Where, x1 and x2 are the sample means, n1 and n2 are the sample sizes, and where sp is calculated as:

sp = √ (n1-1)s12 + (n2-1)s22 / (n1+n2-2)

Where, s12 and s22 are the sample variances.

Syntax:

import scipy

```
from scipy import stats
stats.ttest_ind(sample1 mean, sample2 mean)
```

If the p-value that corresponds to the test statistic t with (n1+n2-1) degrees of freedom is less than your chosen significance level (common choices are 0.10, 0.05, and 0.01) then you can reject the null hypothesis.-

14. How do you find the sample mean difference?

Ans- The mean difference, or difference in means, measures the absolute difference between the mean value in two different groups. In clinical trials, it gives you an idea of how much difference there is between the averages of the experimental group and control groups.

Formula: $x = \Sigma x_i / n$
where:
Σ: means "sum"
xi: The value of the ith observation in the dataset
n: The sample size

For example: suppose we collect a sample of 10 turtles with the following weights (in pounds):
70, 80, 80, 85, 90, 95, 110, 120, 140, 150
The sample mean would be calculated as:
x = (70+ 80+80+85+90+95+110+120+140+150) / 10 = 102

15. What is a two sample t test example

Ans- Suppose we want to know whether or not the mean weight between two different species of turtles is equal. To test this, will perform a two sample t-test at significance level α = 0.05 using the following steps:

Step 1: Gather the sample data.

Suppose we collect a random sample of turtles from each population with the following information:

<u>Sample 1:</u>
Sample size n1 = 40
Sample mean weight x1 = 300
Sample standard deviation s1 = 18.5

<u>Sample 2:</u>
Sample size n2 = 38
Sample mean weight x2 = 305
Sample standard deviation s2 = 16.7

Step 2: Define the hypotheses.

We will perform the two sample t-test with the following hypotheses:

$H0: \mu 1 = \mu 2$ (the two population means are equal)

$H1: \mu 1 \neq \mu 2$ (the two population means are not equal)

Step 3: Calculate the test statistic t.

First, we will calculate the pooled standard deviation sp:

$sp = \surd\ (n1-1)s12 + (n2-1)s22 / (n1+n2-2) = \surd\ (40-1)18.52 + (38-1)16.72 / (40+38-2) = 17.647$

Next, we will calculate the test statistic t:

$t = (x1 – x2) / sp(\surd 1/n1 + 1/n2) = (300-305) / 17.647(\surd 1/40 + 1/38) = -1.2508$

Step 4: Calculate the p-value of the test statistic t.

According to the T Score to P Value Calculator, the p-value associated with $t = -1.2508$ and degrees of freedom $= n1+n2-2 = 40+38-2 = 76$ is 0.21484.

Step 5: Draw a conclusion.

Since this p-value is not less than our significance level $\alpha = 0.05$, we fail to reject the null hypothesis. We do not have sufficient evidence to say that the mean weight of turtles between these two populations is different.