**Statistics and Probability Theory Assignment**

**II. Theoretical Questions:**

**1. Explain the difference between descriptive and inferential statistics. Provide examples of each.**

|  |  |  |
| --- | --- | --- |
| **Sr.No:-** | **Descriptive Statistics** | **Inferential Statistics** |
| 1 | It gives information about raw data which describes the data in some manner. | It makes inference about population using data drawn from the population. |
| 2 | It helps in organizing, analyzing and to present data in a meaningful manner. | It allows us to compare data, make hypothesis and predictions. |
| 3 | It is used to describe a situation | It is used to explain the chance of occurrence of an event. |
| 4 | It explain already known data and limited to a sample or population having small size. | It attempts to reach the conclusion about the population. |
| 5 | It can be achieved with the help of charts, graphs, tables etc. | It can be achieved by probability. |

Example:-

Mode of the following data using descriptive statistics.  
5, 6, 2, 7, 6, 5,1, 9, 5, 8, 5, 4, 3, 12, 11, 17, 5, 5

**Solution:** Mode is the most frequently occurring observation. Thus, the mode is 5

Population mean 100, sample mean 120, population variance 49

**Solution:** Inferential statistics is used to find the z score of the data. The formula is given as follows:

z = (𝑥−𝜇)/𝜎

Standard deviation = √49 = 7

z = (120 - 100) / 7

= 20 / 7 = 2.86

**Answer:** Z score = 2.86= 99% from Table

**2. Define the Central Limit Theorem and discuss its significance in statistical inference.**

The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough.

**3. Discuss the concept of sampling and its role in statistical analysis.**

Samples are used in statistical testing when population sizes are too large for the test to include all possible members or observations. A sample should represent the population as a whole and not reflect any bias toward a specific attribute.

**4. Explain the process of hypothesis testing and the key components involved.**

Hypothesis testing is a fundamental method in statistics used to determine if there is enough evidence in a sample of data to support or refute a particular hypothesis about a population parameter

Here’s a breakdown of the process and its key components:

1. **Define Hypotheses-Null or Alternative**
2. **Calculate the stat score**
3. **Define alpha -**Significance Level (α)
4. Determine the p-value
5. Calculate P Value
6. **If p-value ≤ α**: Reject the null hypothesis (H₀). There is sufficient evidence to support the alternative hypothesis (Ha).

**5. Describe the T-distribution and how it differs from the normal distribution.**

Differences Between T-Distribution and Normal Distribution

1. Shape of the Distribution:
   * Normal Distribution: Has a fixed shape, with its tails being asymptotic but not heavy.
   * T-Distribution: Has heavier tails, meaning it has more probability in the tails and is more spread out. This reflects the increased uncertainty due to smaller sample sizes.
2. Dependence on Sample Size:
   * Normal Distribution: The shape does not depend on sample size. It remains the same regardless of how much data you have.
   * T-Distribution: The shape depends on the degrees of freedom. As sample size increases (and thus degrees of freedom increase), the t-distribution becomes more like the normal distribution.
3. Use in Hypothesis Testing:
   * Normal Distribution: Used for hypothesis testing when sample sizes are large (typically n > 30) and the population standard deviation is known or when the sample size is sufficiently large for the sample standard deviation to estimate the population standard deviation accurately.
   * T-Distribution: Used when sample sizes are small or when the population standard deviation is unknown. The t-distribution provides a more accurate estimate of the standard error in these cases.

**III. Applied Questions:**

**6. Calculate the mean, median, and standard deviation for the following dataset: [10, 15, 20, 25,**

**30].**

Mean = (10 + 15 + 20 + 25 + 30) / 5 = 100 / 5 = 20

Median = 20

Variance =

[(10 - 20)^2 + (15 - 20)^2 + (20 - 20)^2 + (25 - 20)^2 + (30 - 20)^2] / 5 = [(100) + (25) + (0) + (25) + (100)] / 5

= 250 / 5

Variance = 50

Standard Deviation = √50 ≈ 7.07

**7. A researcher wants to estimate the average height of students in a university. She samples**

**50 students and finds the mean height to be 65 inches with a standard deviation of 3 inches.**

**Construct a 95% confidence interval for the population mean height.**

CI = X̄ ± Z×s/√n

=65 ± 1.9600×(3/√50)

=65 ± 0.832

=64.168 – 65.832

8. A manufacturer claims that the average lifespan of its light bulbs is 1000 hours. A random

sample of 50 light bulbs has a mean lifespan of 980 hours with a standard deviation of 50 hours.

Test the manufacturer's claim at a significance level of 0.05 using a right-tailed hypothesis test.

1. State the Hypotheses

* Null Hypothesis (H₀): The average lifespan of the light bulbs is 1000 hours. (H₀) μ=1000 hours
* Alternative Hypothesis (H₁): The average lifespan of the light bulbs is less than 1000 hours. (H₁​)μ <1000

1. Calculate the Test Statistics

* Z = (x- μ)/( 𝜎/√n)

𝑥= sample mean = 980 hours

𝜇 = hypothesized population mean = 1000 hours

𝜎 = sample standard deviation = 50 hours

𝑛= sample size = 50

So Z=(980-1000)/(50/√50)= -2.828427125 ≈−2.83

1. Right-Tailed Test: The critical value z​ is found by looking up 1−α

For α=0.05, this is 1−0.05=0.95.

So the critical value for z (from z-tables) is approximately 1.645

* For a right-tailed test, you compare positive z-values. Since −2.83is not positive, it does not fall into the critical region (which is positive in a right-tailed test).
* In this context, the appropriate test is a left-tailed test because we are testing if the average lifespan is less than 1000 hours.

1. Correct Test Type

* Given the situation, a left-tailed test is more appropriate based on the alternative hypothesis. If you are conducting a left-tailed test:
* Critical value for a left-tailed test at α=0.05 = −1.645
* Since −2.83<−1.645, we would reject the null hypothesis.

1. Final Conclusion

* In the context of the left-tailed test (which is the correct test for the given situation), you would reject the null hypothesis, concluding that there is sufficient evidence at the 0.05 significance level to support the claim that the average lifespan of the light bulbs is less than 1000 hours.

9. A pharmaceutical company is testing a new drug for lowering blood pressure. They want to

determine if the drug is effective in reducing blood pressure levels. State the null and alternative

hypotheses for this study.

When testing whether a new drug is effective in lowering blood pressure, the hypotheses should be set up to reflect the goal of determining if there is a statistically significant effect from the drug. Here’s how to state the null and alternative hypotheses:

Null Hypothesis (H₀​):

The null hypothesis typically represents the status quo or the absence of an effect. For this study, it states that the drug does not have an effect on blood pressure compared to no drug or a placebo.

H₀: Mean of drug= Mean of placebo

or equivalently:

H₀: Mean of drug-Mean of placebo =0

where mean of drug ​ is the mean blood pressure of the group taking the drug, and mean of placebo ​ is the mean blood pressure of the group taking a placebo.

Alternative Hypothesis (H₁​):

The alternative hypothesis represents the effect we are testing for — in this case, that the drug is effective in lowering blood pressure. This would mean the mean blood pressure of the drug group is less than that of the placebo group.

H₁: Mean of drug< Mean of placebo ​

or equivalently:

H₁: Mean of drug-Mean of placebo<0

Summary

Null Hypothesis (H₀​): The new drug does not reduce blood pressure more than the placebo (i.e., the mean blood pressure with the drug is the same as or higher than the mean blood pressure with the placebo).

Alternative Hypothesis (H₁​): The new drug does reduce blood pressure compared to the placebo (i.e., the mean blood pressure with the drug is lower than the mean blood pressure with the placebo).

10. A quality control manager at a factory wants to ensure that the average weight of products

coming off the production line is 500 grams. She takes a random sample of 30 products and

finds the mean weight to be 495 grams with a standard deviation of 10 grams. Test the

manager's claim at a significance level of 0.01 using a left-tailed hypothesis test.

1. State the Hypotheses

* Null Hypothesis (H₀): The average weight of products coming off the production line is 500 grams. (H₀) μ=500 grams
* Alternative Hypothesis (H₁): The average weight of products coming off the production line is less than 500 grams. (H₁​)μ <500

1. Calculate the Test Statistics

Since the sample size is 30 (which is relatively small) and the population standard deviation is not known, we use the t-test for the hypothesis test.

* T = (x- μ)/( 𝜎/√n)

𝑥= sample mean = 495 grams

𝜇 = hypothesized population mean = 500 grams

𝜎 = sample standard deviation = 10 grams

𝑛= sample size = 30

So T=(495-500)/(10/√30)= --2.738612788 ≈−2.74

1. For a left-tailed test with α=0.01 and degrees of freedom (df) equal to n−1=30−1=29, we look up the critical value from the t-distribution table.

The critical t-value for a left-tailed test at α=0.01 with 29 degrees of freedom is approximately −2.462

Since −2.74 is less than −2.462, we would reject the null hypothesis.

1. Final Conclusion

At the 0.01 significance level, there is sufficient evidence to reject the null hypothesis. This suggests that the average weight of the products is significantly less than 500 grams.