STATISTICS WORKSHEET-4

1. What is central limit theorem and why is it important?

Ans: If you take a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from that population will be roughly equal to the population mean.

The Central Limit Theorem (CLT) is a mainstay of statistics and probability. The theorem expresses that as the size of the sample expands, the distribution of the mean among multiple samples will be like a Gaussian distribution.This theorem is very important for testing hypotheses in statistical analysis.

2. What is sampling? How many sampling methods do you know?

Ans: Sampling is a process in statistical analysis where researchers take a predetermined number of observations from a larger population.

There are 4 types of Sampling,

1) Simple random sampling. In a simple random sample, every member of the population has an equal chance of being selected
2) Systematic sampling
3) Stratified sampling
4) Cluster sampling

3. What is the difference between type1 and typeII error?

Ans: A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population;

a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population.

4. What do you understand by the term Normal distribution?

Ans: A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme.

5. What is correlation and covariance in statistics?

Ans: Correlation - However, in statistical terms we use correlation to denote association between two quantitative variables. We also assume that the association is linear, that one variable increases or decreases a fixed amount for a unit increase or decrease in the other.

Covariance - Covariance is a measure of the relationship between two random variables and to what extent, they change together. Or we can say, in other words, it defines the changes between the two variables, such that change in one variable is equal to change in another variable.

6. Differentiate between univariate ,Biavariate,and multivariate analysis.

Ans: Univariate statistics summarize only one variable at a time. Bivariate statistics compare two variables. Multivariate statistics compare more than two variables.

7. What do you understand by sensitivity and how would you calculate it?

Ans: Sensitivity is the percentage of true positives (e.g. 90% sensitivity = 90% of people who have the target disease will test positive). Specificity is the percentage of true negatives (e.g. 90% specificity = 90% of people who do not have the target disease will test negative.

The sensitivity is calculated by dividing the percentage change in output by the percentage change in input.

8. What is hypothesis testing? What is H0 and H1? What is H0 and H1 for two-tail test?

Ans: Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution.

In hypothesis testing there are two mutually exclusive hypotheses; the Null Hypothesis (H0) and the Alternative Hypothesis (H1). One of these is the claim to

be tested and based on the sampling results (which infers a similar measurement in the population), the claim will either be supported or not. The claim might be that the population proportion (or mean) has increased, decreased, stayed the same, or that it has changed. According to the words used in the problem, the claim will be either H0 or H1.
Null hypothesis (H0): The null hypothesis here is what currently stated to be true about the population. In our case it will be the average height of students in the batch is 100. Alternate hypothesis (H1): The alternate hypothesis is always what is being claimed.

9. What is quantitative data and qualitative data?

Ans: Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often). Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.

10. How to calculate range and interquartile range?

Ans: The IQR describes the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1.

11. What do you understand by bell curve distribution ?

Ans: A bell curve is a type of graph that is used to visualize the distribution of a set of chosen values across a specified group that tend to have a central, normal values, as peak with low and high extremes tapering off relatively symmetrically on either side.

12. Mention one method to find outliers.

Ans: There are four ways to identify outliers:

1. Sorting method.

2. Data visualization method.
3. Statistical tests (z scores)
4. Interquartile range method.

13. What is p-value in hypothesis testing?

Ans: The p-value is defined as the probability of obtaining the result at least as extreme as the observed result of a statistical hypothesis test, assuming that the null hypothesis is true.

14. What is the Binomial Probability Formula?

Ans: Binomial probability distribution is $P(r) = {}^{n}C^{r} \cdot p^{r} (1 - p)^{n-r}$.

15. Explain ANOVA and it's applications.

Ans: Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.

ANOVA is helpful for testing three or more variables.