**WORKSHEET-4**

**STATISTICS**

**Q1to Q15 are descriptive types.**

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

central limit theorem is a statistical theory stating that given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population.

The central limit theorem is important because it tells us that no matter what the distribution of the population is, the shape of the sampling distribution will approach normality as the sample size (N) increases.

.

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

Sampling is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. The methodology used to sample from a larger population depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling .

There are two types of sampling methods:

**Probability sampling** involves random selection, allowing you to make strong statistical inferences about the whole group.

**Non-probability sampling** involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

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

Type 1 error, in statistical hypothesis testing, is the error caused by rejecting a null hypothesis when it is true. Type II error is the error that occurs when the null hypothesis is accepted when it is not true. Type I error is equivalent to false positive. Type II error is equivalent to a false negative

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

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.

**5.** What is correlation and covariance in statistics?

Correlation measures both the strength and direction of the linear relationship between two variables.

Covariance indicates the direction of the linear relationship between variables.

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

Univariate analysis is the simplest form of data analysis where the data being analyzed contains only one variable. Since it's a single variable it doesn’t deal with causes or relationships.  The main purpose of univariate analysis is to describe the data and find patterns that exist within it.

Bivariate analysis is used to find out if there is a relationship between two different variables.

Multivariate statistics compare more than two variables.

7. Whatdo you understand by sensitivity and how would youcalculate it?

Sensitivity informs us about the proportion of actual positive cases that have gotten predicted as positive by our model. It is also knows as the true positive rate. It is also known as recall

**Sensitivity = number of true positives / (number of true positives + number of false negatives)**

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

Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter.

The sample is selected using one of the various sampling methods, probabilistic or non-probabilistic. H0 is the notation for null hypothesis whereas H1 is the notation for alternate hypothesis. For a two tailed test, the null hypothesis (H0) should be rejected when the test value is in either of two critical regions on either side of the distribution of the test value and vice versa for alternate hypothesis.

9. What is quantitative data and qualitative data?

Quantitative data can be counted, measured, and expressed using numbers. Qualitative data is descriptive and conceptual.

10. How to calculate range and interquartile range?

Range = Max value – Min value

IQR = Q3-Q1……………………(3rd quartile – 1st quartile)

11. What do you understand by bell curve distribution?

A bell curve is a common type of distribution for a variable, also known as the normal distribution. The width of the bell curve is described by its standard deviation.

12. Mention one method to find outliers.

The most effective way to find outliers is by using the interquartile range (IQR).

13. What is p-value in hypothesis testing?

The P value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H0) of a study question is true

14. What is the Binomial Probability Formula?

Binomial [probability](https://www.varsitytutors.com/hotmath/hotmath_help/topics/probability.html)refers to the probability of exactly x successes on n repeated trials in an experiment which has two possible outcomes (commonly called a binomial experiment).

If the probability of success on an individual trial is p , then the binomial probability is

nCx⋅ p^x⋅ (1−p)^(n – x) .

Here nCx  indicates the number of different [combinations](https://www.varsitytutors.com/hotmath/hotmath_help/topics/combinations.html)of x objects selected from a set of n objects.

15. Explain ANOVA and it’s applications.

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.

There are two main types of ANOVA: one-way (or unidirectional) and two-way. There also variations of ANOVA.

Applications of ANOVA :

* Understanding the impact of different catalysts on chemical reaction rates
* Understanding the performance, quality or speed of manufacturing processes based on number of cells or steps they’re divided into