**STATISTICS WORKSHEET-4 ASSIGNMENT**

**Q-1:-** 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. Regardless of whether the population has a normal, Poisson, binomial, or any other distribution, the sampling distribution of the mean will be normal.

**Formula for Calculation:-** \begin{equation*}\bar{X}\sim N (\mu,\dfrac{\sigma}{\sqrt{n}})\end{equation*}

**Q-2:-** Sampling is a process in statistical analysis where researchers take a predetermined number of observations from a larger population.

The method of sampling depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling also.

**Methods of Sampling includes:-**

**1. Simple Random Sampling.**

**2. Systematic Sampling.**

**3. Stratified Sampling.**

**4. Cluster Sampling.**

**Q-3:-** 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.

**Q-4:-** Normal distribution, also known as the Gaussian distribution, is a type of Probability distribution that is symmetric about the mean, showing that data near the mean and are more frequent in occurrence than data far from the mean.

The standard normal distribution has two parameters: the mean and the standard deviation.

**Q-5:- Correlation:**- A systematic relationship between two variables where a change in one variable is caused due to a change in another variable.

**Covariance:-** A systematic relationship between two variables where a change in one variable reciprocated by an equal change in another variable.

**Q-6:-**

**Univariate**:- It summarises only 1 variable at a time. The main purpose of the analysis is to describe the data and find patterns that exist within it.

**Bivariate:**- It is used for comparing 2 variables at a time.

**Multivariate:-** It compares more than two variables**.**

**Q-7:- Sensitivity** is the percentage of true positives, For ex:- 90% sensitivity = 90% of people who have the target disease will test positive).

The **sensitivity** of a test is the proportion of people who **test positive** among all those who actually **have the disease.**

**Formula:- True Positive**

**True Positive+ False Negative.**

**Q-8:-** Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by them depends on the nature of the data used and the reason for the analysis.

The null hypothesis is generally denoted as H0. It states the exact opposite of what an experimenter predicts or expects. It basically defines the statement which states that there is no exact or actual relationship between the variables. The alternative hypothesis is generally denoted as H1.

H1 also denotes a statistical proposition stating that there is a significant difference between a hypothesized value of a population parameter and its estimated value.

For Two-Tail test, the generic Null and Alternative hypotheses are :-

**Null:** The effect equals zero.

**Alternative**: The effect does not equal zero.

**Q-9:-** Quantitative data are measures of values or counts that are expressed in numbers. Quantitative data are data about numeric variables (e.g. how many, how much). Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.

**Q-10:-** The Range is the distance from the highest value to the lowest value.

The Inter-Quartile Range is just the range of the quartiles: the distance from the largest quartile to the smallest quartile, which is as follows:-

**IQR=Q3-Q1**.

**Q-11:-** A Bell Curve is a common type of distribution for a variable, also known as the normal distribution. The term Bell Curve originates from the fact that the graph used to depict a normal distribution consists of a symmetrical bell-shaped curve.

The top of the curve shows the Mean, Median, and Mode of the data collected.

**Q-12:-**

Basically, there are 2 mostly used methods to find outliers:-

1. Zscore.

2. Using IQR method.

**Q-13:-** It is also called as Probability value which tells how likely your data could have occurred under the null hypothesis. It does this by calculating the likelihood of your test statistic, which is the number calculated by a statistical test using your data.

**Q-14:-** The binomial distribution formula allows us to compute the probability of observing a specified number of successes when the process is repeated a specific number of times, For ex. in a set of patients and the outcome for a given patient is either a success or a failure.

**Formula for calculating is:- Px = (nx)px qn-x.**

**Q-15:-** ANOVA, is a statistical method that separates observed variance data into different components to use in additional tests.

**Formula is:-** ​ F=MSE

MST