Assignment 2- Introduction to statistics.

1. Differentiate between inferential statistics descriptive statistics?

   Ans:

   - **Inferential statistics** is a branch of statistics that makes the use of various analytical tools to draw inferences about the population data from sample data.
   - we use inferential statistics to explain the probability of occurrence of an event.
   - It helps to compare data, make hypotheses and predictions.
   - Main aim is to draw inferences or conclusions about a whole population.
   - We use probability methods for inferential statistics.
   - In this statistics, we need to establish a relationship between variables in an entire population.
   - **Descriptive statistics** gives information about raw data regarding its description or features.
   - We use descriptive statistics to describe a situation
   - It helps to organize, analyze and present data in a meaningful manner.
   - Descriptive statistics explains already known data related to a particular sample or population of a small size.
   - We use charts, graphs, and tables to represent descriptive statistics,
   - It is simpler to perform a study using descriptive statistics rather than inferential statistics.
   - Population refers to the collection of all elements possessing common characteristics, that comprises universe.

2. Differentiate between population and sample

   Ans:

   - **Population** refers to the collection of all elements possessing common characteristics that comprises universe.
   - Measures obtained from population are called population parameter.
   - For example population mean is a parameter
   - A **sample** is the specific group that you will collect data from. The size of the sample is always less than the total size of the population.
   - Measures obtained from sample are called sample statistics
   - For example sample mean is a statistic

3. What is a hypothesis? Differentiate between null hypothesis and alternative hypothesis?

   Ans: A statistical hypothesis defined as tentative conclusion logically drawn concerning the parameter or the form of distribution of the population. For example, “the sample drawn from a normal population with mean=40 and standard deviation= 10” is a hypothesis.

- Null and alternative hypotheses are used in statistical hypothesis testing. Null hypothesis is original hypothesis. Any hypothesis other than null hypothesis is called alternative hypothesis. The null hypothesis of a test always predicts no effect or no relationship between variables, while the alternative hypothesis states your research prediction of a effect or relationship. So when the null hypothesis is rejected we accept the other hypothesis known as alternative hypothesis. Null hypothesis is denoted by H_0 and alternative hypothesis is denoted by H_1.

4. What is the central limit theorem?

   Ans: Central limit theorem is a statistical theory which states that when the large sample size has a finite variance, the samples will be normally distributed and the mean of samples will be approximately equal to the mean of the whole population.

   Conditions of this theorem are:

   - Variable must be independent
   - All variables should have common mean and common standard deviation.
   - All variables should have same distribution
   - 'n' is very large

5. Differentiate between type 1 and type 11 error

   Ans: A type I error appears when the null hypothesis (H_0) of an experiment is true, but still, it is rejected. It is stating something which is not present or a false hit. A type I error is often called a false positive. Type II error appears when the null hypothesis is false but mistakenly fails to be refused. It is losing to state what is present and a miss. A type II error is also known as false negative.

6. What is linear regression?

   Ans: If the regression curve is a straight line, we can say that there is linear regression between the variable under study. The equation of such a curve is the first degree equation in the variables x and y. It can be expressed in the form of y= a + bx.

   In this, the change in the dependent variable is proportionate to the change in independent variable.

7. What are the assumptions required for linear regression?

   Ans:

   Linearity: The relationship between X and the mean of Y is linear.

   Homoscedasticity: The variance of residual is the same for any value of X.

   Independence: Observations are independent of each other.

   Normality: For any fixed value of X, Y is normally distributed.

8. How is the statistical significance of an insight assessed?

   Ans: Statistical significance can be accessed using hypothesis testing. Statistical significance is often calculated with statistical hypothesis testing, which tests the validity of a hypothesis by figuring out the probability that your results have happened

by chance. Here, a hypothesis is an assumption or belief about the relationship between the datasets.

9. What is mean?
   Ans: mean is one of the measures of central tendency. Mean is the average of the given number and it is calculated by dividing the sum of given numbers by the total number of numbers.
   Mean=sum of all the observation /no.of observations

10. What is the meaning of standard deviation?
    Ans: It is the measure of the dispersion of statistical data. Standard deviation is a measure which shows how much variation from the mean exists. It is a popular measure of variability because it returns to the original units of measure of the data set.

11. What is correlation?
    Ans: Correlation is a statistical measure that expresses the extent to which two variables are linearly related. It's a common tool for describing simple relationships without making a statement about cause and effect.

12. What is the meaning of covariance?
    Ans: covariance defines the changes between the two variables, such that change in one variable is equal to change in another variable.

13. Where is inferential statistics used?
    Ans: It is majorly used in the future prediction for various observations in different fields. It helps us to make inference about the data. The statistical inference has a wide range of application in different fields, such as:
    - Business analysis
    - Artificial intelligence
    - Financial Analysis
    - Fraud Detection
    - Machine Learning
    - Share Market
    - Pharmaceutical Sector

14. What is one sample t test?
    Ans: The one-sample t test is a statistical procedure used to compare a mean value measured in a sample to a known value in the population. It is specifically used to test hypotheses concerning the mean in a single population with an unknown variance.

15. What is the relationship between standard deviation and standard variance?

Ans: Variance and Standard Deviation are the two important measurements in statistics. Variance is a measure of how data points vary from the mean, whereas standard deviation is the measure of the distribution of statistical data. The basic difference between both is standard deviation is represented in the same units as the mean of data, while the variance is represented in squared units. Variance of the data set is the average square distance between the mean value and each data value. And standard deviation defines the spread of data values around the mean.

16. What is one way Anova test?

    Ans: The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent groups.