

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.

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 I and type II 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

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

Independence: Observations are independent of each other.

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