What is statistics?

It is science of collecting organizing ,analyzing , data for better decision making.

What is data?

Facts or piece of information that can be measured.

Types of stat?

There are 2 types of stat – DESCRIPTIVE STAT, INFERENTIAL STAT

DESCRIPTIVE STAT- Summarize and describe the features of a data set .that provide simple summaries about the data. Mean ,median ,mode range, variance, standard davitation.

INFERENTIAL STAT- Make inference and predication about a population based on a sample data. Confidence interval hypothesis testing.

SAMPLE AND POPULATION-

SAMPLE- Sample is a subset of the population selected to represent the population in a study.

POPULATION- Population includes all members of a defined group that we are studying or collecting information on for data decision .

SAMPLING TECHNIQUE-

SIMPLE RANDOM- Equal chance for all members.

SYSTEMATIC SAMPLING- Every 5th member.

STARTIFIED SAMPLING- Sampling from homogeneous subgroup.

CONVIENCE SAMPLING- Samling from easily accessible group.

VARIABLES- It is a property that can held,store,take any value.

2 TYPES OF VARIABLES.

QULITATIVE VARIABLES- Variables that represent categories or group . ex- gender,brand

QUANTITATIVE VARIABLES- Variables that represent numerical values. Ex- height, weight.

VARIABLES MEASUREMENT SCALE

ORDIANAL- Ordered

NOMINAL – Categorical Value

RATIO- Zero means nothing

INTERVAL- Represent order categories with equal interval

FREQUENCY- The number of times a specific value or event occurs in a dataset.

HISTOGRAM- A histogram is a graphical representation of the distribution of numerical data.it

provides a way to see the frequency distribution (shape) of a set of data.

BINS- The range of value is divided into interval called bins.

BARS- Eacch bin is represent by a bar the height of the bar of the number of data point.

OUTLIER- A datapoint who dosent follow pattern or trend of the data set then it is a consider as outlier.

VARIANCE- It mean how far the number in a dataset are from the mean.

STANDARD DEVIATION- Is a statistics that measures the dispregen or spred of a data. It indicates how much the value in a dataset different from the mean value of data.

SQUARE ROOT OF VARIANCE- It gives measure of spred that is in the same units as the original data.

FIVE NUMBER SUMMARY-

MINIMUM (Q)- The smallest value in the dataset.

FIRST QUARTILE (Q1)- The median of the lowest half of the dataset.

MEDIAN(Q2)- The middle value of the dataset.

Q3- The median of the upper half of the dataset.

MAXIMUM- The largest value in the dataset.

BOX PLOT- Is a graphical representation of the five number summery of a dataset.

DATA DISTRIBUTION- Refer to how the value in a dataset are spread out a clustred together. It show how often different value occour in data set and describes the overall pattern of the data.

NORMAL DISTRIBUTION- Often referred to as a bell curve is a type of continuous probability distribution for a real valued random variable.

Z score- Also known as standard score.

STANDARD NORMAL DISTRIBUTION- Is a special case of the normal distribution mean is 0 and standard deviation is 1.

NORMALIZATION- Refer to the process of adjusting value measured on different scale.

POSITIVELY SKEWED DISTRIBUTION-

A positively skewed distribution also known as right skewed distribution is a type of distribution where the majority of the data points are clustered on the left side. With tail extending to the right.

Mean is greter than median, median is greater than mode.

NEGATIVELY SKEWED DISTRIBUTION-

A negatively skewed distribution also known as left skewed distribution is a type of distribution where the majority of the data points are clustered on the right side. With tail extending to the left.

Mean is less than median ,median is less than mode.

BERNOALILY DISTRIBUTION-

It model single experiment with two possible outcomes.

BINOMICAL DISTRIBUTION-

Binonimal distribution extends bernomali to independent trails.

EXPONTIAL DISTRIBUTION-

It describes the time below events in a process where events occour indepentily and a constant rats.

UNIFORM DISTRIBUTION- Is a type of probability distribution in which all outcomes are equally likely.

DISCRETE UNIFORM- In a discrete uniform distribution there is a finite number of outcomes each with equal probability.

CONTINUONS UNIFORM- Any value within a given range is equally likely.

CONFIDENCE INTERVAL- It is a range of value within which we expect a particular population parameter to fall.

CONFIDENCE LEVAL- A confidence leval refer to the percentage leval refer to the percentage of all possible sample that can be expected to include the true population parameter.

HYPOTHESIS TASSTING-

It is statistical method that is used to decide whether there is enonums eviendance to reject a null hypothesis based on sample data.

NULL HYPOTHESIS- A statement suggest that there is no effect , no difference, or no change in the population.

ALTERNATIVE HYPOTHESIS- A statement that contradict the null hypothesis it suggest that there is an effect a difference or a change in the population.

REJECTION REGION METHOD- Wheather to reject the null hypothesis.

P VALUE- It is measure of the strength of the evidence against the null hypothesis.

T test- It's a statistical method that helps you determine if there's a significant difference between the means of two groups. Essentially, it tests whether the difference in sample averages (means) likely reflects a true difference in the populations, or if it's just due to random chance.

INDEPENDENT T TEST- Used when you have two different groups of subjects (e.g., comparing the test scores of two different classes.

PAIRED T TEST- Used when you have two sets of related measurements (e.g., before-and-after measurements on the same subjects.

ONE SAMPLE T TEST- It's handy when you want to compare the average of your sample to a specific value and see if there's a meaningful difference.

CHI-SQUARE TEST- Chi-square test is a statistical test used to determine whether there is a significant association between two categorical variables or whether an observed frequency distribution matches an expected distribution.

KURTOSIS- Is a statistical measure that describes the shape of a distribution’s tails in relation to its overall shape. It specifically measures the extremity of data points (outliers) in the tails compared to a normal distribution.

THERE ARE 3 TYPES OF KURTOSIS

MESOKURTIC- A distribution with kurtosis similar to that of a normal distribution. The kurtosis value is approximately 3.

LEPTOKURTIC- A distribution with heavy tails and sharp peak. The kurtosis value is greater than 3. Indicates more outliers than the normal distribution.

PLAYKURTIC- A distribution with light tails and a flatter peak. The kurtosis value is less than 3. Indicates fewer outliers than the normal distribution.

HIGH KURTOSIS- Indicates data with heavy tails or outliers. High probability of extreme values.

LOW KURTOSIS- Indicates data with light tails. Lower probability of extreme values.