**STATISTICS**

**Statistics Intro:**

It is a huge field where we are going to deal with data.

We are going to deal with data in four ways. They are:

1. Collection of data

2. Analyzing the data

3. Interpret the data

4. Structure to the data

The types of data (variable) will fall into two categories:

1. Quantitative data

2. Qualitative data

**Qualitative data:** It means whenever we are collecting the data in that data whenever I cannot measure what quantity it is, but I can measure only quality then it is called Qualitative data. It is also called as categorical data.

**Quantitative data:** It means whenever we are having the data and if we are able to measure the quantity of the data then it is called Quantitative data. It is also called as numeric data.

It is further classified into two types. They are:

1. Discrete

2. Continuous

**Discrete Data:** W**hen values in a data set are countable and can only take certain values, it is called discrete data**. We can easily count the variables in a discrete data.

For example, number of students in a class, number of players required in a team, etc.

**Continuous Data:** This data has values that are not fixed and have an non-finite number of possible values. These measurements can also be broken down into smaller individual parts.

For Example, the height or weight of a person, the daily temperature in your city, the amount of time needed to complete the task, etc.

**Levels of Measurement**

It will help us to collect and measure the data precisely. Each level will tell us how to actually collect the data and what type of descriptive statistics we can use whenever we are having the below types of data.

1. Nominal
2. Ordinal
3. Interval
4. Ratio

**Nominal:** The nominal data can only be categorized and it will not have rank/order and no intervals. The nominal data belongs to qualitative data

Based on the nominal data the mathematical operations we can do they are:

== (equal to)

!= (not equal to)

**Ordinal:** The ordinal data can be categorized and orderd. The nominal data belongs to qualitative data

Based on the ordinal data the mathematical operations we can do they are:

== (equal to)

!= (not equal to)

<, >(Less than, greater than)

<=,>= (Less than equal to, greater than equal to)

**Interval:** The interval data can be categorized, orderd, can given scale and no true zero. It belongs to continuous data and sometimes it can be discrete data

Based on the Interval data the mathematical operations we can do they are:

== (equal to)

!= (not equal to)

<, >(Less than, greater than)

<=,>= (Less than equal to, greater than equal to)

+,- (Addition, Subtraction)

**Ratio:** The interval data can be categorized, orderd, can given scale and true zero.

Based on the Ratio data the mathematical operations we can do they are:

== (equal to)

!= (not equal to)

<, >(Less than, greater than)

<=,>= (Less than equal to, greater than equal to)

+,\*,% ,-(Addition, Multiplication, Division, Subtraction)

Analysis on data:

Based on analysis of data the statistics is actually divided into two types. They are

1. Descriptive Statistics

2. Inferential Statistics

Descriptive Statistics: It is used to summarize the data and describe the main features of a dataset. Whenever we are having population we can simply go with this statistics.

Inferential Statistics: It uses sample data to make inferences, draw conclusions and make predictions about a larger population. Whenever we are having sample but if we want to know about the population then we can go with this statistics

Population: Whenever you are having problem statement and you are having all observations about the data then it is known as population.

Sample: Whenever you are having problem statement and you are randomly collecting the data of some observations then it is known as sample. It is subset of the population

Sample Size: The total number of observations collected.

The descriptive statistics is divided into three parts. They are:

1. Measures of central tendency

2. Measures of variability/dispersion

3. Distribution

Measures of central tendency: Whenever we are having data and we want to know central or average information about the data then we will use measures of central tendency in three ways . They are:

* Mean
* Median
* Mode

Measures of variability/dispersion: Whenever we are having data and if we want to know how our data is distributed or spread then we will use measures of dispersion in four ways. They are:

* Range
* Variance
* Standard deviation
* Quantile or 5 number summary

Distribution: It describes how values are distributed for a field.