

Introduction to Statistics

- 1. Types of Statistics:
 - Discriptive and Inferential
- 2. population & sample

Descriptive:
It is used to describe the dataset.

For example:
In politics,there are no. of parties in it like AAP,BJP,Con so we can predict like which party gone win the election.
we can detemine/guess that which party will win our area as well.
Anything which we can describe called as 'descriptive'.
This the way like ppl talk , we heard so we can describe who is gonna win, this is called as discriptive staisctics.

Inferential:
if we are unable to define a subset from set n then concept of inferential comes into pitchure.
e.g. an election,there is no of states in india,if wants to find the winner from each state,its difficult by an individual,then we take samples from each state and then we tend to get result like getting A from ABC called as 'inferential' statistics.

If we are not able to describe dataset then we get some sample and inference called as inferential statistics.

Example: SONY TV set-
if they manufacture 10,000 tv set, then we hace to check quality.at that time,we'll check 4-5 tv's amongs the 1 set each,lets say we've 100 sets to be delivered. once found out 4-5 are in good condition then 100 ll be good then we can say all 10,000 may be in good condition where we are able describe the sets called as descriptive & inferential statistics.

- Statistics is all about the working behind the scene. majority theory.
- 1. Analytics methology and how industry use statistics:
 - * weather forecasting
 - * giving insurance
 - * stock market
 - * drug effectiveness before releasing to public
 - * diseased survival probability
 - * election winning and exit poll prediction
 - * loan approval and fraud detection
 - * netflix/amazon recommendation
 - * new campaign effectiveness

- 1. Parameter and statistics (mean,median,mode):
 - mean - comparision of sample statistics and population parameter
sample sats represent as x-bar which gives the mean of sample
mean of population data, if i got 1000 then add 1000 and then 1000 ll definately get some data which represents by μ .
 - there is some data lets say 1000 so +/-10 which represent by s amd sigma-standard deviation.
Variance- is define as square of standard deviation

For Example:find out mean, median & mode of series 1,2,2,3,3,5,4,6,7,7,7,8,9,8
mean=avaaerage of number is 5.14
median=is the center no,bt if there are more than 1 then take mean of those no=5.5
mode=no which occur max time is 7

```
In [1]: import statistics as st

x=[1,2,2,3,3,5,4,6,7,7,7,8,9,8]

st.mean(x)

Out[1]: 5.142857142857143

In [3]: #We can find out by pandas as well
import pandas as pd

x=pd.DataFrame(x)
print(x_.mean())

0    5.142857
dtype: float64

In [4]: st.median(x)

Out[4]: 5.5

In [6]: st.mode(x)

Out[6]: 7
```

- 1. Uses of variable: dependent and independent variable
 - for example:year of experience and salary of emmployee,so experience is independent becoz its keeps incresing but salary is based on year of experience so salary is the dependent variable.

- 1. Types of variable: Continuos and categorical variable
 - Continuos variable(numerical data): continuos and discrete
 - continuos is continuos in data e.g.age-is something continuos,there is nothing u can change
 - discrete-e.g ranks,shoe size is what we get as 6,6.5,7,7.5 like these there is nothing between 6 to 6.5. its called discrete.
 - Catogerical(String):most of the time it is in string.
 - ordinal & nominal:
 - ordinal-those which are in order for example:the seneority level jr.eng,eng,sr.eng,mgr or edu level 1st,2nd....10th,+2,grad....
 - nominal-has types:e.g gender F,M or brands:lifestyle,colorbar brands are to identify products comes in nominal.

```
In [ ]:
```