

Statistics - Statistics is the science of collecting, organizing and Analysing the Data.

Data :- facts or pieces of information.

Eg Ages of Student in classroom

{24, 25, 35, 29, 28} \Rightarrow Mean, Median, Mode
Standard deviation

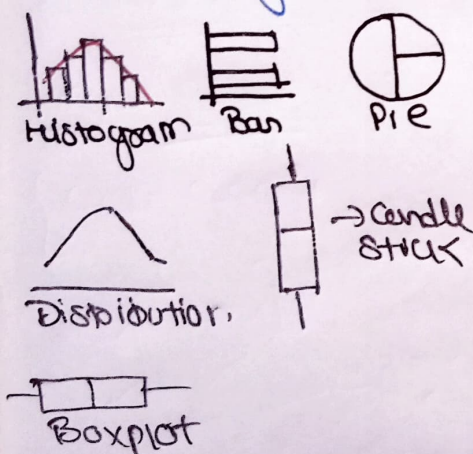
Eg Weights of Students in classroom.

Types of Statistics

Statistics

Descriptive Stat [EDA + FE]

① It consist of organizing & summarizing the data



Inferential Stat

It consist of collecting sample data & making conclusion about population data using comp Experiments

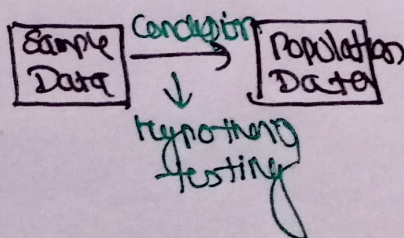
Done by hypothesis testing

Eg

University \rightarrow 500 people

Class A \rightarrow 60 people

Sample Data \rightarrow Age \rightarrow Avg age of entire University



Ex Let say there are 20 classrooms in a university and you have collected the age of student in one classroom.

Age = {17, 18, 25, 30, 31}

weight = {2, 2, 2, 2, 2}

Descriptive Stat - What is the avg age of student in the class?

- Relation b/w Age & Gender

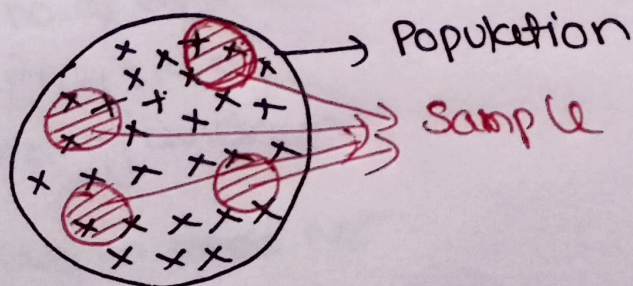
Inferential Stat - Are the avg age of the student in the classrooms less than the avg. age of student in the university?
($<$ or \geq)

- from thousands student from class A the avg marks of 50 girls & 50 boys are 85% & 92%, so can we say girls performed well

Sample (n) & Population (N)

Population (N) is the entire group ~~of~~ that you want to draw conclusion about.

Sample (n) is specific group that you will collect data from



Sampling techniques

① Simple Random Sampling:-

Every member of the population (N) has an equal chance of being selected for your sample (n)

Ex: Lottery, Exit poll, Criminal survey, Movie reviewing

② Stratified Sampling

Strata \rightarrow layers \rightarrow categories \rightarrow groups

Gender \rightarrow Male
 \rightarrow female

Education
layer \rightarrow High school
 \rightarrow Master
 \rightarrow PhD
 \rightarrow Bachelor

} layers

Blood group \rightarrow A
 \rightarrow B
 \rightarrow AB
 \rightarrow O

} layers

③ Systematic Sampling \rightarrow selecting every n^{th} individual out of population.

Ex: Airport

{ credit card }

Person \rightarrow approach
giving
reason
for credit
card

Person \rightarrow approach
giving
reason
for credit
card

④ Convenience Sampling - only those who are interested in the survey will only participate.

Ex: Survey req new technology

- ① Survey regarding new technology \Rightarrow Convenience Sampling
- ② RBT Survey \Rightarrow Married women \rightarrow Stratified + Random
- ③ Credit card calls \rightarrow Stratified + Random

① Variable: A variable is a property that can take any values

Eg age = 14
age = 25
age = 100

Variable
Ages = [24, 25, 26, 27, 28, 29]

- Types of variable

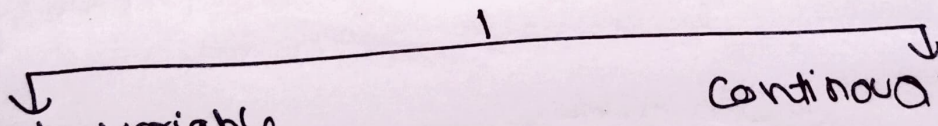
① Quantitative Variable \rightarrow measured numerically
Mathematical operation

Eg Age, wt, height, rainfall (cm), temp, distance

② Qualitative Variable \rightarrow categorical variable
Eg Gender, type of flowers, types of movies

{ Based on some characteristics they are grouped together }

Quantitative Variable



Discrete variable

Eg whole numbers

Eg no. of Bank account
{1, 2, 3, 4, 5}

Eg no. of children

based on whole no

Continuous variable

Eg continuous

Eg Height, wt, Age, Rainfall
Speed,

Assesment

- ① what kind of variable is Marital status → Categorical
- ② " " " " " Orange River length → Continuous
- ③ " " " " " Movie duration → Continuous
- ④ " " " " " Pincode → Discrete
- ⑤ " " " " " IQ → Continuous

$\mu = 11$
 $\sigma = 5$
 $\sigma^2 = 100$

Types of variable

- ① Quantitative variable → measured numerically
 ex: Age, wt, height, temperature, time, distance, money, etc.
- ② Qualitative variable → categorical variable
 ex: sex, color, religion, etc.

