Subject: Data Science

MON TUE WED THR FRI SAT SUN

3rd September, 2022

Statistics Note (Day-1)

By Krish Naik Sir.

★ What is statistics ?

→ Statistics is the science of collecting, organising and analyzing the data.

★ What is Data ?

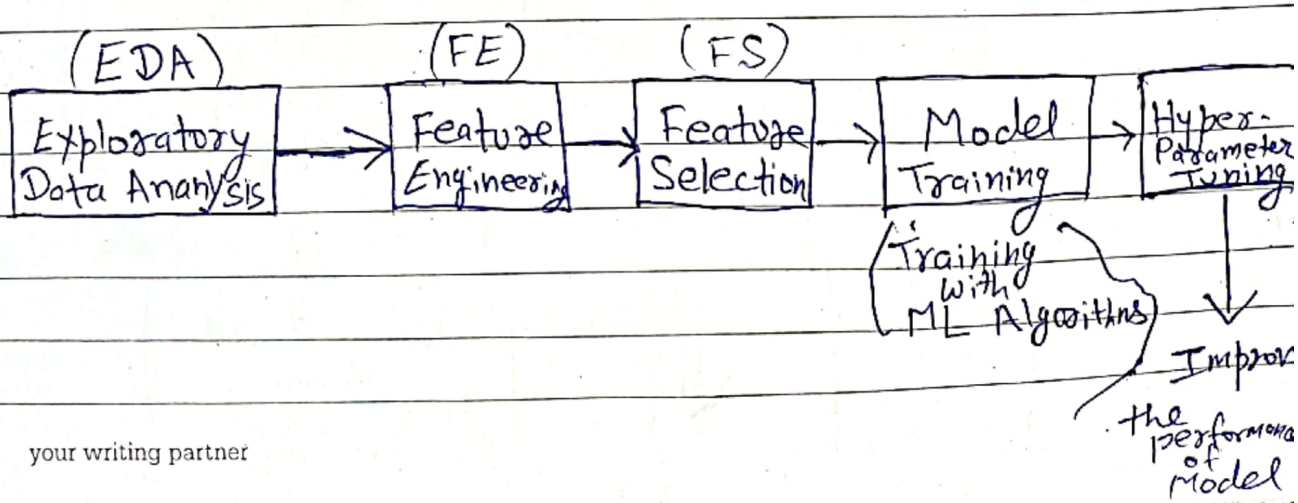
→ "facts or Pieces of information".

Eg:- Ages of Students in classroom.

(a) {24, 25, 32, 29, 28} ⇒ Mean, Median, Mode, Standard Deviation

(b) Weights of students in classroom.

★ Life Cycle of Data Science Projects :-



your writing partner

Subject _____

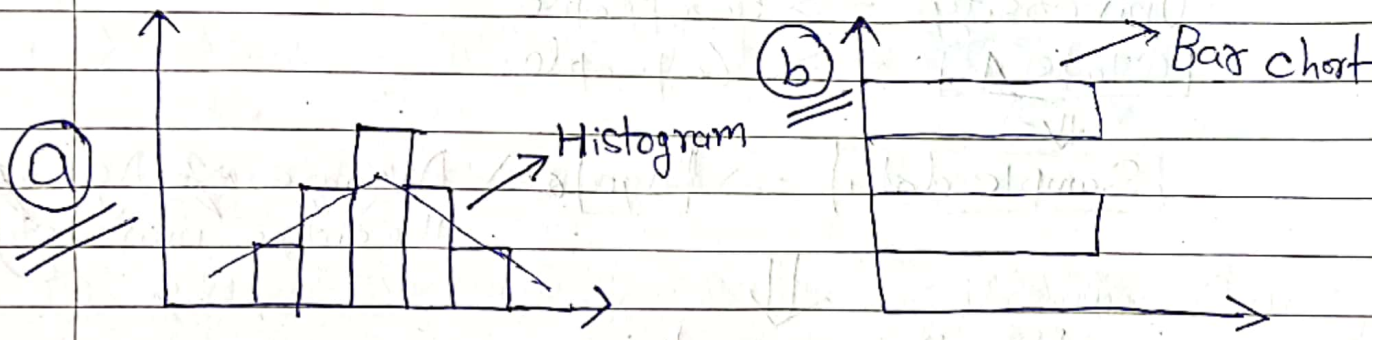
MON TUE WED THR FRI SAT SUN

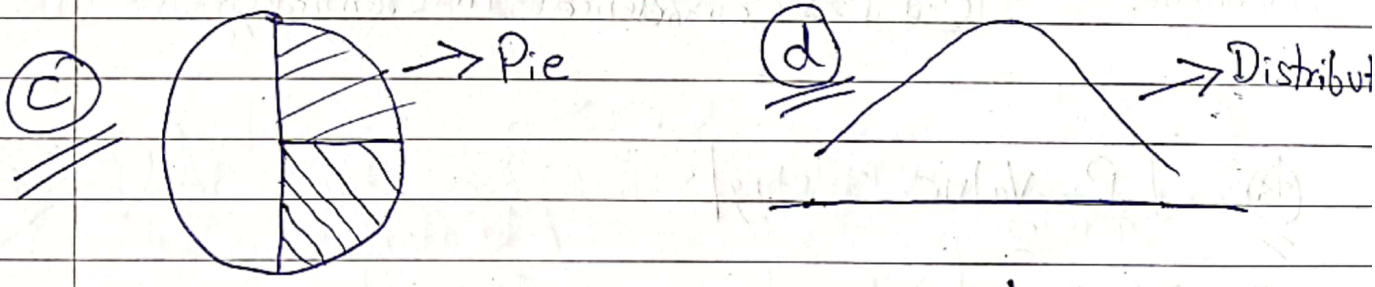
Type of Statistics

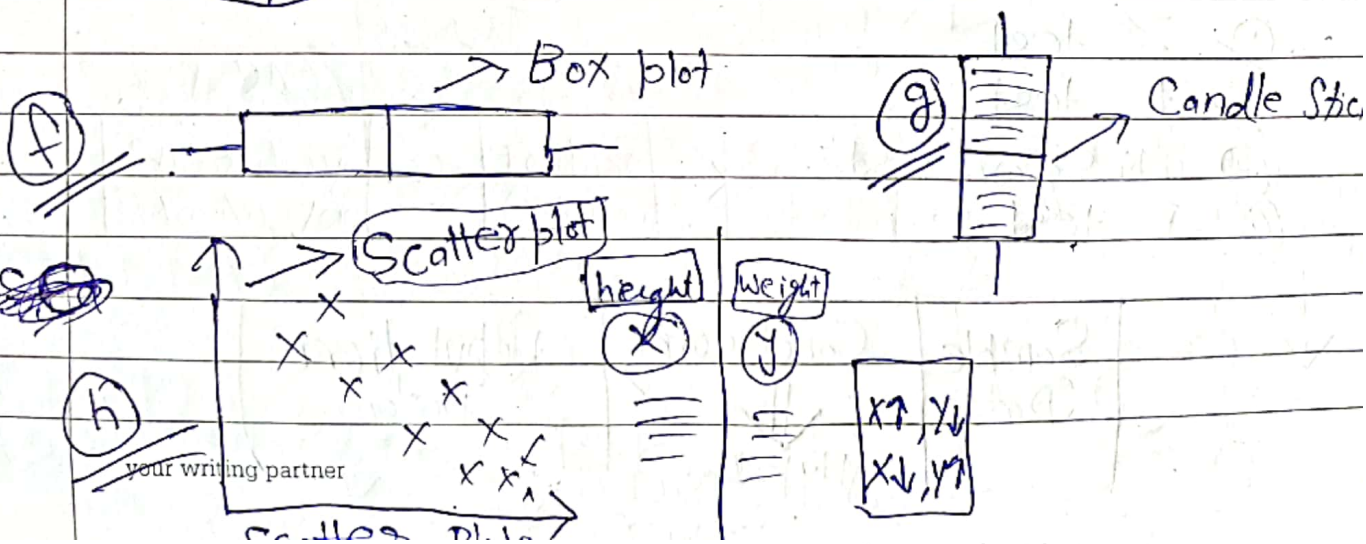
- Descriptive Stats (EDA + FE)
- Inferential Stats

1. Descriptive Stats.

→ It consists of organising and Summarizing the Data.







your writing partner

Subject _____

MON TUE WED THR FRI SAT SUN

② Inferential Stats :-

→ It consists of collecting Sample Data and making conclusion about populations data using some experiments.

(A) Making Conclusion ⇒ by Hypothesis Testing.

Example:

University → 500 people
Class A → 60 people

⇓

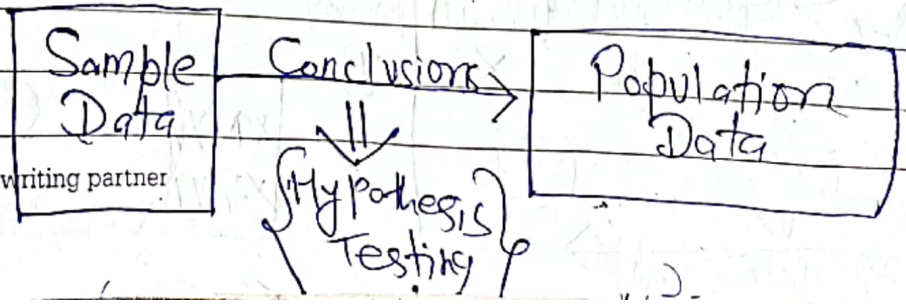
Sample data → Age ⇒ Average Age of the entire University

⇓

Hypothesis Testing.
C.I ⇒ Confidence Inferential,

(A) P- Value Testing

1. Z test
2. t test
3. Chi Square Test
4. F test



Subject _____ MON TUE WED THR FRI SAT SUN

Sample Data Vs Population Data

10 Crs
↓
Population Data

Exit poll
↑
Party A will Win
Party B will Loose

Sample Size = 1000

Eg:- Let's say there are 20 Classrooms in a university and you have collected the age of Students in one classroom.

Ages = { 21, 20, 18, 34, 17, 22, 24, 25, 26, 23, 22 }

Weights = { - - - - - - - - - - - }

Descriptive Stats :- What is the average age of Students in the classroom ?

or

Relationship between Age & Genders ?

Inferential Stats :- Are the average age of the Students in the classroom less (significantly more) than the average age of the students in the university

Hypothesis Testing ← 1000 Students (your writing partner)

Class A ⇒ 50 girls 95% | 50 Boys 92 %

Subject _____

MON TUE WED THR FRI SAT SUN

population (N)
Sample (n)

(A) Sampling Techniques

(1) Simple Random Samplings :-

→ Every member of the population (N) has an equal change of being selected for your Sample(n).

n = 1000

<18 | >18 → Random Sampling

(2) Stratisfied Samplings :-

→ Strata → Layers → Clusters → Groups.

(eg)

Gender → Males
→ Female

Blood group

Education Degress → High School
→ Master
→ Phd

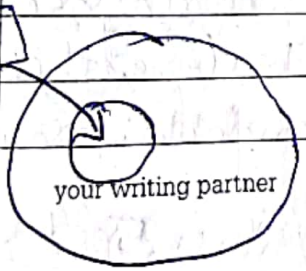
Population {Exit Poll}

<18 | >18

<18 → Not Vote

>18 → Random Sampling

<18

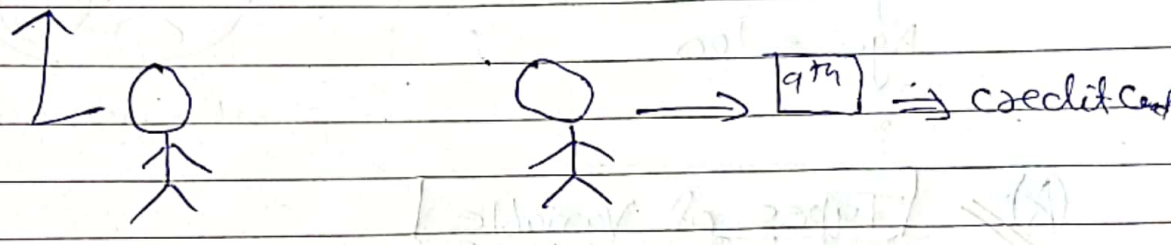


Subject _____ MON TUE WED THR FRI SAT SUN

③ Systematic Sampling :-

Eg {Credit Card}

Select every n^{th} individual out 5th person of population (N)



④ Convenience Sampling :-

→ Only those who are interested in the survey will only participate.

Eg :- {Data Science Survey → General AI Survey}

{fill the Form} ⇒ iNeuron job for Specific Role.

Assessment

(a) Survey Regarding New Technology ⇒ Convenience Sampling

(b) RBI Survey (Women) ⇒ Stratified + Random Sampling + Married Women

(c) Credit Card ⇒ Stratified + Random Sampling.

Subject _____

MON TUE WED THR FRI SAT SUN

⭐ Variables

→ A variable is a property that can take any values.

for Eg:-

Variable

Age = 14
Age = 25
Age = 100

Variables

Ages = [24, 25, 26, 27, 28] → Collection

⭐ Types of Variable

- Quantitative Variable
  - Discrete Variable
  - Continuous Variables
- Qualitative Variable

① Quantitative Variable :-

→ Measured Numerically {Mathematical operation}

for Eg:-

Age, Weight, Height, rainfall (cm), temp., Distance

② Qualitative Variable :-

→ Categorical Variable {Based on some Characteristic they are grouped together}

for Eg:- Gender, Type of flowers

Subject: Data Science

MON TUE WED THR FRI SAT SUN

Quantitative Variable

Discrete Variable

Eg:- Whole Number → fixed

Eg:- No. of Bank Accounts
{1, 2, 3, 4, 5} → 2.5 ✗

Eg:- No. of children → Whole Number

Eg:- Pincode = fixed,

Continuous Variable

Eg:- Continuous → Decimal value

Eg:- Height, Weight, age, Rainfall, Speed

Eg:- Marital [Married, Not Married] → Categorical Variable

Assessment

(a) What kind of variable is Marital Status ? ⇒ Categorical Variables

(b) What kind of variable is Ganga River ~~Range~~ Length ? ⇒ Continuous Variable

(c) What kind of variable is Movie Duration? ⇒ Continous variable

(d) What kind of variable is Pincode ⇒ Discrete

(e) What kind of variable is IQ ⇒ Discrete