

INTRODUCTION TO STATS

7 days Session

7pm - 8pm → 8:30 pm

① Basics To Advance

{DATA Scientist, Data Analyst, BUSINESS
INTELLIGENCE TOOLS}

2 days Basics

① DESCRIPTIVE STATS



{① Measure of Central Tendency }
{② Measure of Dispersion }

Summarizing the data.

Histograms, Pdf, Cdf, Probability,

Permutation, Mean, Median, Mode,

Variance, Standard deviation

i) Gaussian Distribution

② LogNormal Distr

③ Binomial Distr

④ Bernoulli's Distr

⑤ Poach Distr {Power law}

⑥ Standard Normal Distr → python

⑦ Transformation and Standardization

⑧ Q-Q plot

⑨

② Inferential Stats

⑧ Z test → python

t test → python

ANOVA → F test

CHISQUARE:

HYPOTHESIS TESTING {P values}

Confidence Intervals

Z table, t table

What is Statistics?

Statistics is the science of collecting, organizing and analyzing data. { Better Decision Making }

Dyn Data? 2

Data : Facts or pieces of information that can be measured

Eg : The IQ of a class

{ 98, 97, 60, 55, 75, 65 }

Ages of students of a class

{ 30, 25, 24, 23, 27, 28 } → DATA

Types of Statistics

① DESCRIPTIVE STATS

It consists of organizing and summarizing data

② Inferential Stats

Technique where in we used the data that we have measured to form

Conclusions

① Classroom of Maths student (20)

Marks of the 1st Sem

84, 86, 78, 72, 75, 65, 80, 81, 92, 95, 96, 97, - - -

Eg : Descriptive Stats

① What is the average marks of the students in the class?

Inferential Stats

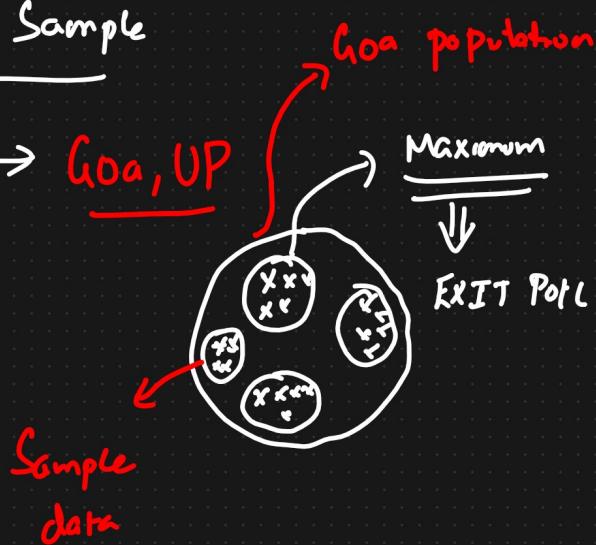
Sample 1 → Population.

Eg : Are the marks of the students of this classroom similar to the age of the Maths classroom in the college?

Population And Sample

Elections

EXIT Poll



Population (N) Sample (n)

Sampling Techniques

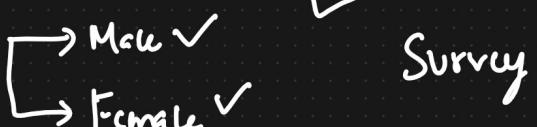
① Simple Random Sampling : Every member of the population (N) has an equal chance of being selected for your sample (n)



② Stratified Sampling : Where the population (N) is

Split into non-overlapping groups (strata)

Eg: Gender

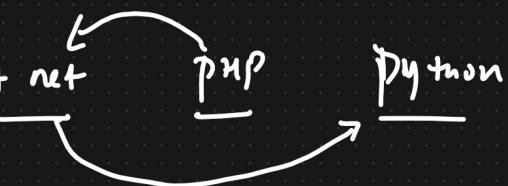


Survey

Age group

(0-10) (10-20) (20-40) (40-100)

Eg: Profession : Dot net PHP Python



Doctors, Engineer

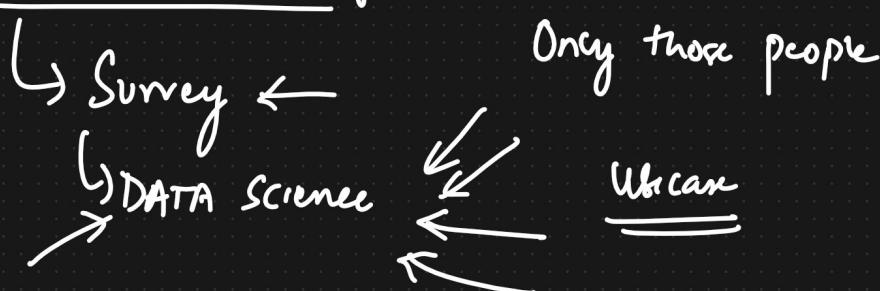
③ Systematic Sampling Thanos

(N) \rightarrow n^{th} individual

Eg: Mail \rightarrow Survey (Covid)

\downarrow 8th person \rightarrow Survey

④ Convenience Sampling



Eg: EXIT POLL

{Random Sampling}

RBI \rightarrow Household Survey



Survey \rightarrow Women

Eg: Drug \rightarrow Tested \Rightarrow
 \Downarrow

What Kind of Sample ??

Variables :

A variable is a property that can take on

any value $\{182, 178, 168, 150, 160, 170\}$

Eg: Height =
Weight $\{78, 99, 100, 60, 50, \dots\}$

Two Kinds of Variables \rightarrow Eg: Age
Weight
Height

- ① Quantitative Variable \rightarrow Measured Numerically, {Add, Subtract, multiply, divide}
- ② Qualitative / Categorical Variables

Eg: Gender $\begin{cases} M \\ F \end{cases}$ {Based on some characteristics we can define Categorical Variables}

Eg: IQ

$\frac{0-10}{\Downarrow}$	$\frac{10-50}{\Downarrow}$	$\frac{50-100}{\Downarrow}$
Low IQ	Medium IQ	Good IQ

<u>Blood group</u>	<u>Shirt size</u>
A	L
B	XL
O	M
AB	S

Quantitative

Discrete Variable

Eg: Whole number
No. of Bank Accounts

Continuous Variables

Eg: Height = 172.5, 162.5 cm, 163.5 cm,
Weight = 100kg, 99.5, 99.75

Eg: 2, 3, 4, 5, 6, 7,

Rainfall = 1.1, 1.25, 1.35 - - -

② Total of children in a family

Eg: 2, 3, 4, 5,

Eg: What kind of variable Gender IS? Categorical

② What " " " Marital Status? . "

③ River Length? Continuous

④ population of the state is? Discrete

⑤ Song length? Continuous

Blood pressure? Continuous.

PIN CODE ? $\left\{ \begin{array}{l} \text{Discrete or categorical} \\ \hline \end{array} \right\}$

⑥ Variable Measurement Scales

4 types of Measured Variable

Colors, Gender, Type of flower

① Nominal data { Categorical data } → Classes

② Ordinal → Order of the data matters, value does not

③ Interval → Order matters, values also matter, natural zero is not present

④ Ratio.

Eg:

Students (Marks)	Rank	
100	1	
95	2	
57	4	
85	3	

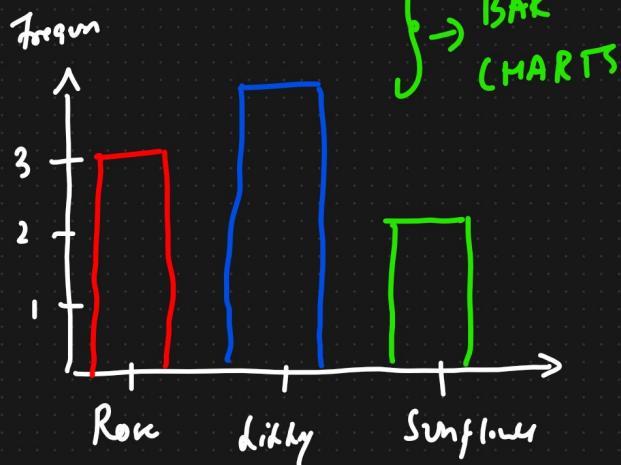


TemperaturesFahrenheit70 - 8080 - 9090 - 100100] Fahrenheit(F) Ratio data {Assignment}Frequency Distribution

Sample character : Rose, lilly, Sunflower, Rose, lilly, Sunflower,
 Rose, lilly, lilly

Flower	Frequency	Cumulative Frequency
Rose	3	3
lilly	4	7
Sunflower	2	9

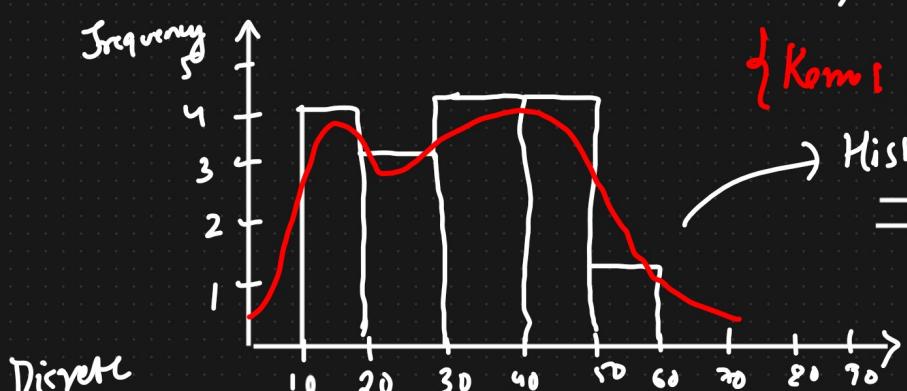
(I) BAR GRAPH



② Histograms \div Continuous

Ages = { 10, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41, 42, 43, 50, 51 }

Bins = 10



BAR VS Histogram

pdf: probability density function