STATISTICS

[DAY - 1]

* **Simple Definition of Statistics :**

Statistics is the science of collecting, organizing and analyzing data.

* TYPES OF STATISTICS

Descriptive statistics

↳ consists of organizing and summarizing of data.

↳ Example :
Histogram, Barchart, PDF, pie chart, Box plot, candle Stick

Inferential statistics

↳ It consists of a technique to form some conclusions.

↳ Example :
z-test, p-value, t-test, χ^2 test, ANOVA

* SAMPLING TECHNIQUES :

(1) Simple Random Sampling :

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

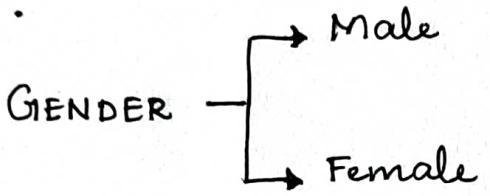
(2) Stratified Sampling :

Strata → layers → clusters

⇓

non-overlaping

EXAMPLE :



(3) Systematic Sampling

⇒ Select every nth individual

(4) Convenience Sampling

⇒ Only those people who are interested will only be picked.

Questions Based on Sampling

Which sampling should be used ?

(1) Salesman sells credit card → use Systematic or Random sampling

(2) Survey regarding new Technology → convenience sampling

(3) (age > 18) EXIT POLL → Stratified and random sampling

* VARIABLE

A variable is a property that can take on any value

* TWO KINDS OF VARIABLE

(1) Quantitative → Measure numerically

(2) Qualitative → Categorical variable

(1) Quantitative variable

- Discrete variable
  - EXAMPLE :
  - (i) No. of Bank Accounts
  - (ii) No. of children in a family
  - (ie) Whole numbers (no decimals)
- Continuous variable
  - EXAMPLE :
  - (i) Height = {172.5, 150.56, ...}
  - (ii) Rainfall = 1.35cm
  - (iii) Weight (iv) Temperature

(2) Qualitative variable

Based on some characteristics we can group it

EXAMPLE : Gender, Types of flowers, Types of movies, etc.,

* HOW TO CONSTRUCT A HISTOGRAM

Suppose the given data is

Ages = { 10, 13, 22, 18, 32, 27, 38, 40, 51, 45, 56, 57, 88, 90, 99, 92, 94 }

SOLUTION:

[STEPS: (1) Sort the numbers into ascending order

(2) Bins → (no. of groups)

(3) Bin size → (Size of bins)

If suppose the range is between [1-100]

Taking bins = 5 then the bin size = $\frac{100}{5} = 20$]

⇒ Ages = { 10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99 }

Let bins = 5, bin size = 20