Statistics

Definition - collection, organizing, visualization, summerizing and analyzing of the data is called statistics.

* It has two types

1. Descriptive statistics
2. Inferential statistics

1. Descriptive statistics -

In it we collect, summerize and visualization of the data

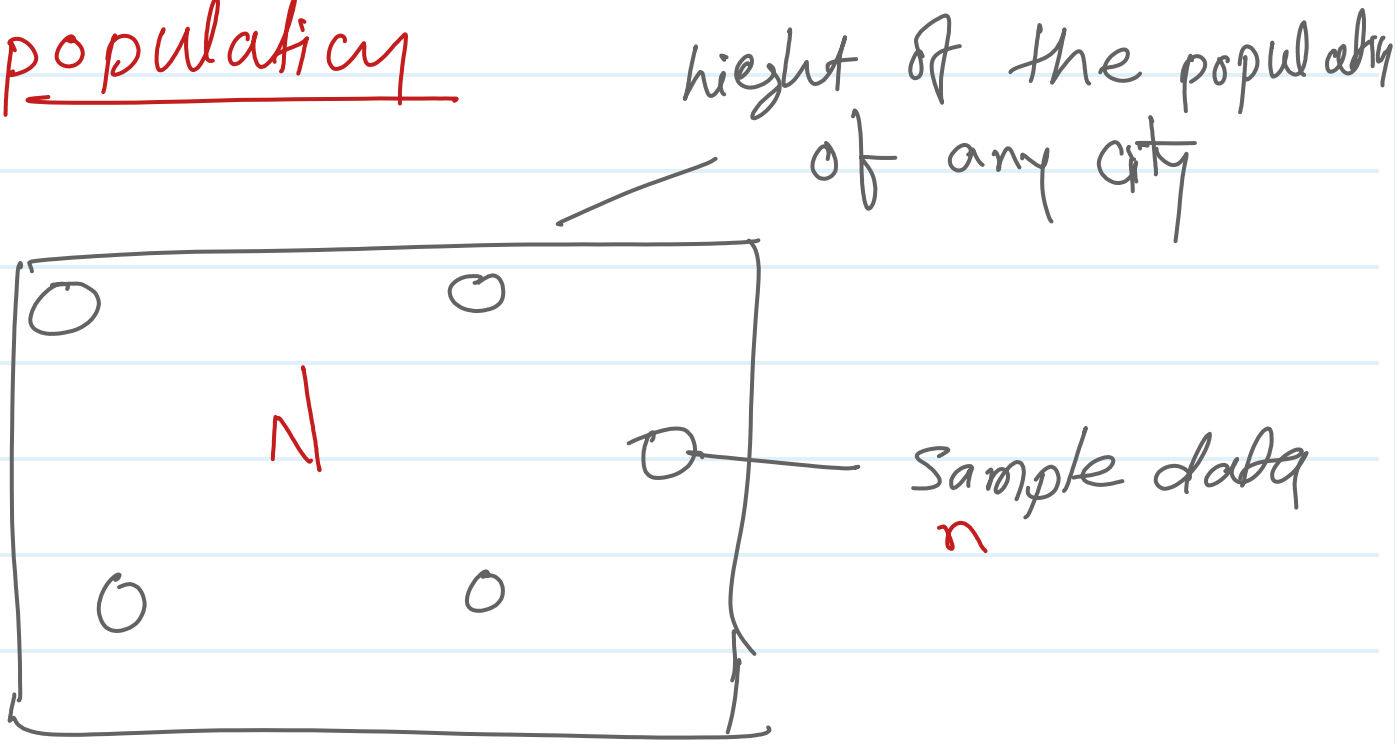
1. measure of centres tendency
2. measure of dispersion
3. measure of shape
4. measure of positing

2. Inferential statistics -

In It we analys the data by perform statistical test like z-test, T-test, chi-squre test so on. to conclude result.

Hypothesis testing → z
probability t
F
chi-

✱ populatiay



population is entire data which is available.

Sample is a small dataset collected from population data.

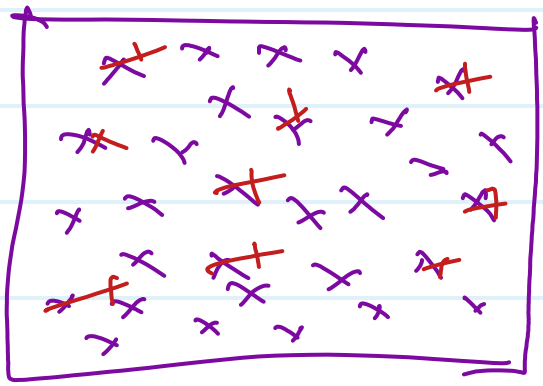
population — N

Sample — n

Type of sampling method.

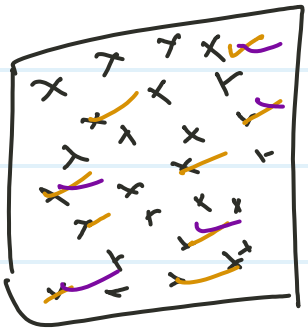
1. Simple Random sampling.
2. stratified sampling.
3. systematic sampling
4. Convience sampling

1 SRS —



$n = 6$

② stratified sampling



N = 500

= m = 280

f = 220

collect sample age > 18 and educatn
= Graduate.

③ Systematic sampling –

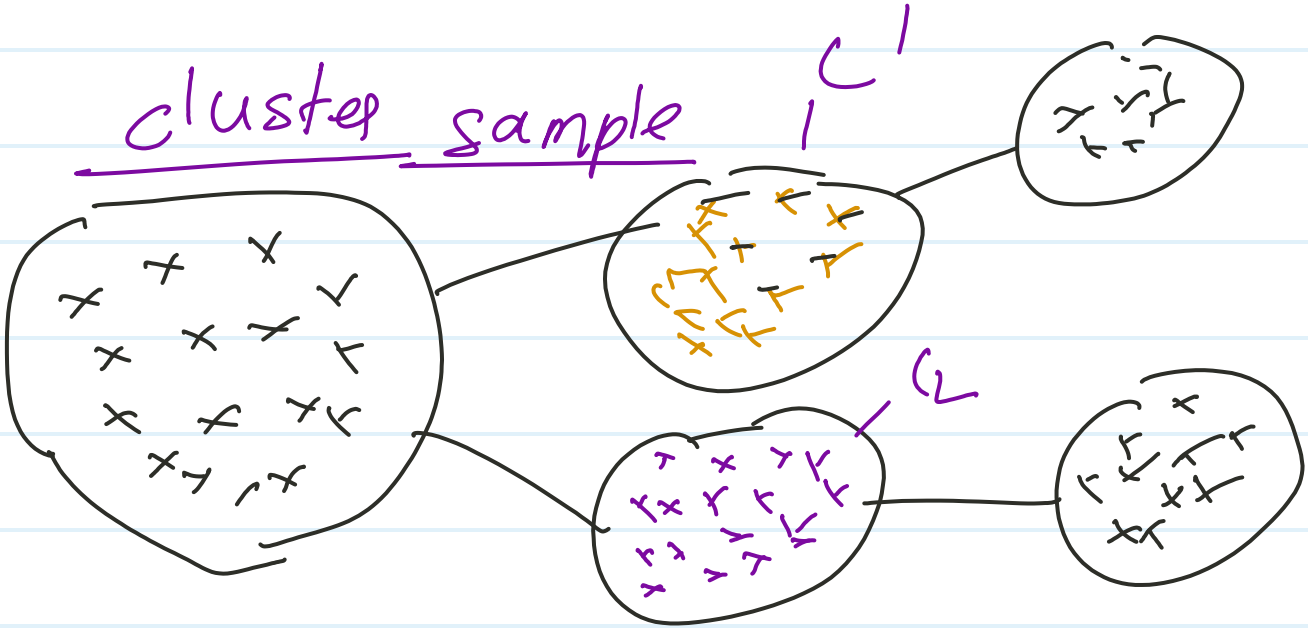
|\|+|||+|||+|||+|||+|||+|||+|||+|

④ convience sampling :-

Eg: corona vaccine trials

volintiers –

cluster sample



population given.

⇒ How to find sample size.

Cochran formula to calculate Sample size.

$$n_0 = \frac{z^2 pq}{e^2}$$

e = margin of error (5-1)

p = population proportion - 50%.

q = 1 − p

z = value from z-table

C.I.		Z value (Z-table)
90%	Z	1.645
~~95%~~		1.960 ✓
99%		2.576

e margin of error 5%

$\frac{5}{100} \Rightarrow 0.05$

$P = 0.5$ (Assumed)

$$n_0 = \frac{(1.960)^2 \times 0.5 \times (1-0.5)}{(0.05)^2}$$

$\Rightarrow 384.16$

Sample size for infinity population.

Sample size for required population

$$N = 100000$$

$$S = \frac{384.16}{1 + \left[\frac{384.16 - 1}{100000}\right]}$$

$$S = \frac{384.16}{1.0038316}$$

$$S = 382.69$$

$$S = 382 \quad \text{(Approximate)}$$

other formula for the small population

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

✡ Descreptive statistics

✡ Measure of centres tendency

(i) mean -

Dataset = [2, 6, 9, 7, 3, 5, 4]

$$\text{mean} = \frac{2+6+9+7+3+5+4}{7}$$

$$= 5.14$$

population mean - μ ✓

Sample mean - $\bar{x}$ ✓

② median

[2, 6, 9, 7, 3, 5, 4]

shorting the data [2, 3, 4, 5, 6, 7, 9]

median = 5

If dataset is even

[2, 3, 4, 5, 6, 7, 8, 9]

median = $\frac{5+6}{2}$ ⇒ 5.5

③ mode

[2, 3, 4, 5, 5, 6, 7, 8, 8, 8]

highest frequency of any number.

mode = 8