29 November 2023 08:11 PM

1. Population → whole set of observations

process is called sampling.

2. Sample → representation of the actual data

i.e sample is subset of population.

Types of data: 1. Time series Data → different time period (same place)

2. Cross section Data → different places.

3. Panel Data

(i) West Bengal (Relation Rainfall and crop Produ) during 2010-2020 → Time series.

(ii) Collecting data on Rainfall and crops across different states of India during year 2022-23.

cross-section

(iii) Collection of data on rainfall and crop across different states over the years 2010-2020 → Panel data.

A. Quantive Data → variables

→ attributes

A. Quantitive

B. Qualitative Data → attributes

Variables → Discrete variable (whole numbers)

→ Continuous variable (fraction, decimal etc.)

Method of collection of Data

1. Primary Data

- ⇒ Raw in nature
- ⇒ First time collected
- ⇒ Directly supervised or surveyed by researcher or surveyor.
- ⇒ More reliable because of less chances of discrepancies.

1. Direct observation method.
2. Interview Method
   a. Telephonic
   b. personal
   c. indirect
3. Questionnaire method

2. Secondary Data :

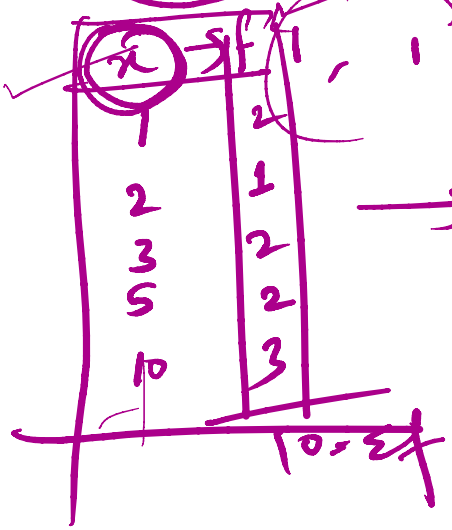
- → collected from sources already published
- → chances of error is high
- → transcription error is very common
- → less reliable than primary data.

→ less reliable primary data.

Frequency Distribution :

$n = 10$

1, 1, 10, 10, 10, 2, 3, 5, 5, 3

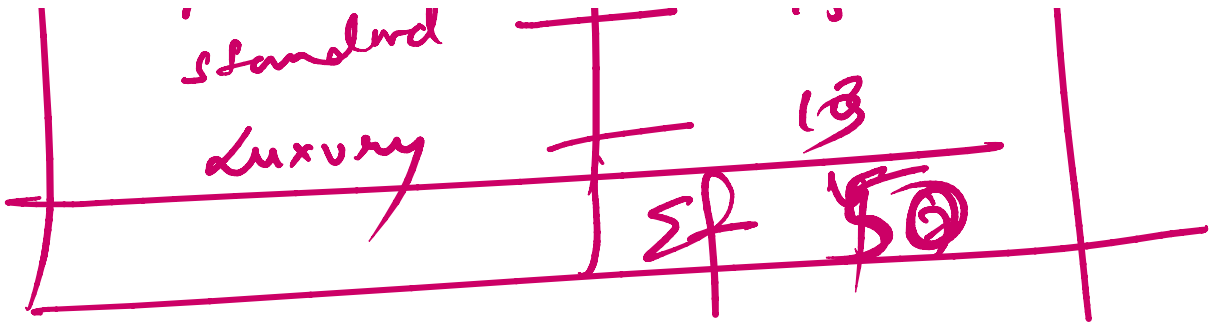


→ frequency distribution.

Ex: Sex of infants born in City hospital:

Sex (attribute)	frequency
Male	18 ✓
female	22 ✓
Total birth	$\Sigma f = 40$

Quality of Rooms	frequency
very poor	10
poor	12
standard	15
	13



Discrete variable

Ex: family size.

family size (discrete)	No. of families (frequency)
2	9
3	20
4	30
5	17
6	10
7	4
	$\Sigma f = 90$

Ex: Height of students collected (100 students) — 10 class

166.5 160.5 155 152.5

Ex: Marks of 50 students.

5 classes

10

31, 32, 80, 75, 70,

50,

50 students

Marks (class intervals)	Number of students (frequency)

(class interval)	(frequency)
31 - 40	6
41 - 50	14
51 - 60	20
61 - 70	7
71 - 80	3
	$\Sigma f = 50$

class limits

class boundaries $d = \frac{41 - 40}{2} = \frac{1}{2} = 0.5$

Class boundary = lower limit − 0.5 → lower bound
upper limit + 0.5 → upper bound

Class boundary	frequency (f)	Relative frequency	Frequency Density
30.5 - 40.5	6	6/50 = 0.12	6/10 = 0.6
40.5 - 50.5	14	14/50 = 0.28	14/10 = 1.4
50.5 - 60.5	20	20/50 = 0.4	20/10 = 2
60.5 - 70.5	7	7/50 = 0.14	7/10 = 0.7
70.5 - 80.5	3	3/50 = 0.06	3/10 = 0.3
	$\Sigma f = 50$		

✓ Relative Frequency = $\frac{\text{class frequency}}{\text{Total frequency}}$

✓ Frequency Density = $\frac{\text{class frequency}}{\text{class size or class width}}$

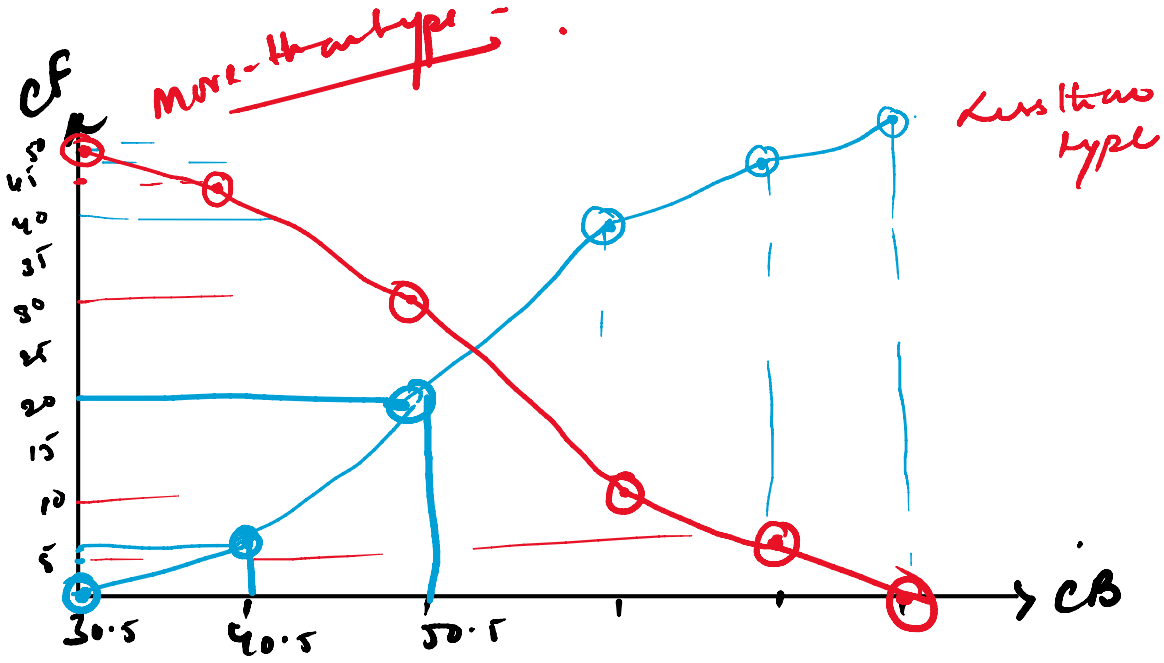
✓ Class mark or mid value of class = $\frac{\text{upper limit/boundary} + \text{lower limit/boundary}}{2}$

Cummulative Frequency Distribution

Class Boundaries	Frequency (f)	Cummulative Frequency: Less than type	Cummulative Frequency: More than type
30.5 – 40.5	6	6	50
40.5 – 50.5	14	6+14 = 20	44
50.5 – 60.5	20	40	30
60.5 – 70.5	7	47	10
70.5 – 80.5	3	50 = N	2
	$\Sigma f = 50 = N$		

(Marks) Class Boundaries	Cummulative Frequencies: Less than	More than
Less than 30.5	0	50
Less than 40.5	6	44
" 50.5	20	30
" 60.5	40	10
" 70.5	47	3
" 80.5	50	0
"		

Ogive: Graphical Representation of Cummulative frequency Distribution.


CF
More-Hazard type
Less Hazard type
CB
30.5
40.5
50.5