

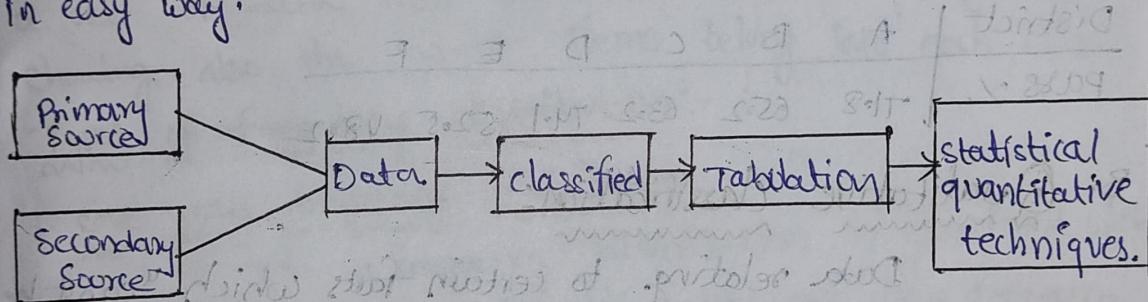
## Basics of Statistics

### Basics of statistics

Statistical data need to be collected to provide vital base for inferences. Data collected from various sources is mass & complex. Raw and complex data is simplified in such a way that the analyst can understand the nature and scope of data to make efficient quantitative analysis.

The first in the process of simplification is splitting the data in groups and sub groups called classification. These groups are classified under specific heads and captions called tabulation.

After tabulation quantitative analysis can be done in easy way.



Classification: Classification is the process of arranging data into sequences and groups according to their common characteristics.

Methods of classification: Depending on the nature of data classification is different types. The following are some important types of classification.

1. Chronological classification: Data relating to a point of time (or) a period of time is chronological data & such data is

classified on the basis of time are called a period data relating to growth of population, exports & imports, sales, profits etc. are classified under this

## Category.

Eg:- performance of Aditya company over

Year	Sales (Rs)	Expenditure	Profits (Rs)
2019	555	10	85
2020	602	12	102
2021	628	15	115
2022	675	16	142
2023	710	18	162

2. Geographical Classification:-

If the type is classified on the basis of area such as region state country. It is geographical classification.

Eg:- Results of intermediate examination Andhra pradesh March 2023 (or) 2021.

District	A	B	C	D	E	F
Poss. %.	71.8	65.2	63.2	74.1	52.5	48.2

3. Qualitative Classification:-

Data relating to certain facts which "can't be measured directly". In qualitative terms are called qualitative data (or) attributes which are present in data.

Eg:- Religion, honesty, beauty etc.

There are two types of classification by attributes

1. two-fold classification.

2. many-fold classification.

1. Two-fold classification:

When the data is classification on the basis of

Present (or) absents of one attribute it is called two-fold classification.

2. Many-fold classification:-

If the data is classified on the basis of more than one attribute it is called many-fold classification.

#### A) Quantitative classification:-

If the data is classified on the basis of measurable quantitative terms it is called quantitative (or) classification by variables.

Eg:- If the data is classified on the basis of marks, then the marks are variables.

#### Tabulation:-

Tabulation is the process of arranging classified data in the form of a table, so that it may be more easily understood & so that any comparisons involved may be more readily made (clear).

#### frequency Distribution:-

A frequency distribution is a table that divides a set of data into suitable no. of classes (categories) showing also the no. of items belong into each class.

#### Types of frequency distribution:-

1) One way frequency distribution:-  
If a frequency distribution is classified with only one frequency, then it is called one way frequency distribution.

Eg:- Height of students.

155	158	162	163	165	169	170
-----	-----	-----	-----	-----	-----	-----

No. of students.

8      12      20      25      13      16      12

Again univariant frequency distribution is divided

into three types.

1) series of individual observation:-

It is a series of items listed single one after the other. After observing the data, it may be a record Series.

Eg:- Income of # members in a shop is given by

member	A	B	C	D
Income (Rs)	2100	1800	2200	3100

## ② Discrete series :-

This is a discrete frequency distribution where observations are independent to each other. In the discrete series data is presented as it is in the Universe along with its frequency of occurrence each universe is different and separate from others.

Eg:- Prepare a discrete series table for the marks of 20 students in a test given by 9, 8, 8, 3, 5, 1, 6, 5, 4, 2, 1, 5, 1, 5, 1, 4, 3, 1, 5, 4, 6, 7, 8.

marks (x)	0	1	2	3	4	5	6	7	8	9
No. of students	1	1	2	3	5	3	1	2	1	1

## 3. Continuous series :-

In a continuous series observations are expressed in small groups called class intervals with certain limits frequency in continuous series refer to the group but not to individual items.

Eg:- The arrangement of 500 visitors of a library with age group from 10 to 90 years in a continuous table is given by.

Age group	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of visitors	25	35	58	62	100	90	30	

### Components of continuous series :-

#### 1. Range :-

Range can be defined as the difference b/w the

(lower limit of the first class) (smallest class) and the upper limit of the highest class (largest class).

#### 2. Class interval :-

In a continuous frequency distribution data is classified into small groups with certain limits & class intervals (or) simply classes

3. class limits :-

Each class interval consist of two limits with in. the size of the class is restricted this are called upper limit & lower limit of a class.

4. class boundaries :-

where there is a gap between the upper limit of the class & lower limit of the first class & lower limit of the next class there is a need to determine the class boundaries without which same measure can not be

compute.

Eg: 9-19, 20-29, 30-39.

for such classes class boundaries can be determined by using the following formulas.

$$\text{lower class boundary} = \text{lower limit} - \frac{d}{2}$$

$$\text{upper class boundary} = \text{upper limit} + \frac{d}{2}$$

$d$  = difference between upper limit of the first class & lower limit of the next class.

Example:-

Construct a class boundaries for a class intervals 50-54, 55-59, 60-64, 65-69, 70-74

Class	class boundaries
50-54	49.5 - 54.5
55-59	54.5 - 59.5
60-64	59.5 - 64.5
65-69	64.5 - 69.5
70-74	69.5 - 74.5

mid value of class interval :-

Once data is classified under class intervals with certain limits individual items of the series loss their identity to represent a class interval in measure in any technique its mid point

is consider.

mid point of a class =  $\frac{\text{lower limit} + \text{upper limit}}{2}$

magnitude of a class interval's -

The difference between upper limit & lower limit of a class is called magnitude or width or size of that class interval. It is denoted by  $i$ .

Class frequency :-

No. of observations falling within a particular class interval is its class frequency. It is denoted by  $f$ .

Construction of class interval:-

H.A. Sturges gave formula to determine no. of class intervals length of each class interval

$$n = 1 + 3.222 (\log N)$$

where  $n$  = no. of class intervals

$N$  = total no. of observations or. total frequency

$$i = \frac{\text{Range}}{N}$$

• where,  $i$  = length of each class interval.

Range = largest observation - smallest observation.

Types of class intervals:-

There are two ways of forming class intervals

1. Inclusive class interval:-

Under this method no item appears twice because the upper limit of first class is not same as

the lower limit of second class.

Eg:- 0-9, 10-19, 20-29, 30-39

gap is larger at higher limit than at lower limit

which is important part in construction of histogram

Q. Exclusive class interval :-  
 under this method the upper limit of first class is same as the lower limit of the next class. It continues till the end.

Eg:- 0-5, 5-10, 10-15, 15-20.  
 Cumulative frequency distribution :-  
 cumulative frequency (C.F) is the running total of frequencies with cumulative totals of every class interval.

It is constructed by adding frequency of first class to the frequency of second class, again this total to the frequency of third class and so on till end.

Eg:-

C.I	f	C.F
0-5	2	2
5-10	5	7
10-15	8	15
15-20	10	25
20-25	12	37

Two way frequency distribution :-

Some times it is desirable to study the relationship of observations with two types of frequencies which have equal impact on the observation.

Eg:-  
 1. Employees can be classified under year of experience and salary.

2. marks of students can be classified on the basis of no. of classes attended & no. of tests written.

Construction of frequency table :-

frequency table is constructed with the help of tally sheet.

\* After determining the observations (in case of

discrete series) or class intervals list them under 'x' or classes column.

- \* observe the information given and for every item of class or observation a line is drawn against the class for fifth item a cross line across 4 lines is drawn. It is known as four and cross method.
- \* At the end all the lines are counted and marked in the frequency column. finally the frequency table with class interval & frequency is to be prepared.

Problems:-

1. The following data is about consumption of petrol by 60 two wheeler owners in a town prepare a discrete series. 9, 2, 13, 11, 4, 5, 12, 6, 2, 3, 8, 9, 7, 16, 5, 4, 11, 3, 2, 7, 5, 4, 3, 5, 4, 3, 18, 7, 15, 4, 13, 6, 5, 3, 4, 28, 6, 8, 5, 4, 7, 5, 4, 15, 6, 9, 7, 11, 8, 5, 1, 12, 19, 18, 17, 8, 4, 6, 5, 1.

A)

Consumption (x)

Tally marks

frequency.

1

||||

01

00-3

2

||||

01

25-5

5

|||| |||

10

6

|||| |

6

7

|||| |

6

9

||| |

6

60

2. The following are the defective pairs identify in shoe factory every day in January. construct Discrete series frequency table.

No. of Defects 9, 0, 1, 13, 6, 4, 5, 7, 8, 6, 5, 4, 3, 5, 4, 1, 2, 3, 5, 8, 7, 4, 5, 3, 2, 6, 5, 1, 9, 0, 6.

A)

Defective pairs (x)

tally marks

frequency

0

• 11 April 11th 2023 by S. S. J. in 9001

1

forwards

III

crosses

3

I-2

2

P

II

III IV

2

P-E-Z-A

3

S

III

III IV

3

Z-N-P-S

4

the

III

III III III III III III

4

A-Z-B-Z-P

5

C

III

III III III III

7

F-A-Z-Z

6

H

III

III III III

4

S-F-R-A

7

F

II

III

2

P-S-E-S-T

8

E

II

III

2

P-O-I-P-Z

81

II

2

9

II

31

3. The following are the monthly savings of 75 families in a colony. prepare the suitable frequency table using Sturges rule.

Savings: 53, 44, 55, 79, 56, 48, 57, 67, 42, 97, 59, 34, 51, 90, 76, 54, 60, 55, 71, 46, 49, 53, 56, 61, 81, 51, 50, 63, 71, 28, 17, 45, 51, 68, 60, 93, 35, 53, 39, 53, 29, 30, 67, 71, 75, 49, 65, 50, 42, 40, 69, 23, 54, 72, 66, 46, 70, 45, 80, 41, 50, 24, 53, 21, 54, 58, 73, 15, 37, 4, 28, 86, 28, 55, 63, 72.

A) Here.  $N=75$

By Sturges Rule

No. of class intervals

$$\begin{aligned}
 n &= 1 + 3.222 (\log N) \\
 &= 1 + 3.222 (\log 75) \\
 &= 1 + 3.222 (1.8751) \\
 &= 1 + 6.0414 \\
 &= 7.0414 \\
 &\approx 7
 \end{aligned}$$

length of the each class interval

$$\begin{aligned}
 i &= \frac{\text{Range}}{N} \\
 &= \frac{97 - 23}{75} \\
 &= 10.5714 \\
 &\approx 11
 \end{aligned}$$

(x) divided

Here we take 7 classes with length 11.

C.I	Tally marks	frequency	
23-34		5	
34-45		8	
45-56		14	
56-67		13	
67-78		14	
78-89		4	
89-100		3	
			75

4. Construct frequency table for the data relating to the pairs of shoes produced by 25 workers in a factory using inclusive method.
- pairs of shoes: 1, 36, 35, 13, 18, 3, 11, 29, 5, 21, 32, 13, 12, 19, 6, 9, 37, 42, 140, 16, 10, 19, 10, 7, 15, 4, 8, 16, 42, 39, 41, 45, 10, 6, 12, 21, 9, 14, 32, 15, 11, 16, 10, 9, 5, 6, 149, 7, 12, 17, 3, 48, 81, 7, 12, 17, 18, 14, 124, 7, 10, 29, 12, 31, 42, 26, 31, 34, 32, 27, 4, 8, 9, 15, 19, 21, 25, 32, 39, 48, 47, 48, 5, 8.

A)  $N = 25$   
By sturges rule  
No. of class intervals

$$\begin{aligned}
 n &= 1 + 3.222 (\log N) \\
 &= 1 + 3.222 (\log 25) \\
 &= 1 + 3.222 (1.9294) \\
 &= 1 + 6.2165 = 7.2165 \approx 7
 \end{aligned}$$

length of each class interval

$$\begin{aligned}
 i &= \frac{\text{Range}}{N} \\
 &= \frac{47 - 1}{25} \\
 &= \frac{46}{25} = 1.84 \\
 &\approx 2
 \end{aligned}$$

C-I

1-7

tally marks

111 111 111 11

frequency.

PE-08

8-14

111 111 111 11

11 111 17

PN-04

15-21

111 111 111

111 15

22-28

111

5

22-28

How to draw frequency table of stock production.

29-35

111 111 111

11 111 15

22-28

36-42

111 111

11 111 15

22-28

43-49

111 111

11 111 15

22-28

50-56

111 111

11 111 15

22-28

57-63

111 111

11 111 15

22-28

64-70

111 111

11 111 15

22-28

71-77

111 111

11 111 15

22-28

78-84

111 111

11 111 15

22-28

85-91

111 111

11 111 15

22-28

92-98

111 111

11 111 15

22-28

99-105

111 111

11 111 15

22-28

106-112

111 111

11 111 15

22-28

113-119

111 111

11 111 15

22-28

120-126

111 111

11 111 15

22-28

127-133

111 111

11 111 15

22-28

134-140

111 111

11 111 15

22-28

141-147

111 111

11 111 15

22-28

148-154

111 111

11 111 15

22-28

155-161

111 111

11 111 15

22-28

162-168

111 111

11 111 15

22-28

169-175

111 111

11 111 15

22-28

176-182

111 111

11 111 15

22-28

183-189

111 111

11 111 15

22-28

190-196

111 111

11 111 15

22-28

197-203

111 111

11 111 15

22-28

204-210

111 111

11 111 15

22-28

211-217

111 111

11 111 15

22-28

218-224

111 111

11 111 15

22-28

225-231

111 111

11 111 15

22-28

232-238

111 111

11 111 15

22-28

239-245

111 111

11 111 15

22-28

246-252

111 111

11 111 15

22-28

253-259

111 111

11 111 15

22-28

260-266

111 111

11 111 15

22-28

267-273

111 111

11 111 15

22-28

274-280

111 111

11 111 15

22-28

281-287

111 111

11 111 15

22-28

288-294

111 111

11 111 15

22-28

295-301

111 111

11 111 15

22-28

302-308

111 111

11 111 15

22-28

309-315

111 111

11 111 15

22-28

316-322

111 111

11 111 15

22-28

323-329

111 111

11 111 15

22-28

330-336

111 111

11 111 15

22-28

337-343

111 111

11 111 15

22-28

344-350

111 111

11 111 15

22-28

351-357

111 111

11 111 15

22-28

358-364

111 111

11 111 15

22-28

365-371

111 111

11 111 15

22-28

372-378

111 111

11 111 15

22-28

379-385

111 111

11 111 15

22-28

386-392

111 111

11 111 15

22-28

393-399

111 111

11 111 15

22-28

400-406

111 111

11 111 15

22-28

407-413

111 111

11 111 15

22-28

414-420

111 111

11 111 15

22-28

421-427

111 111

11 111 15

22-28

428-434

111 111

11 111 15

22-28

435-441

111 111

11 111 15

22-28

442-448

111 111

11 111 15

22-28

449-455

111 111

11 111 15

22-28

456-462

111 111

11 111 15

22-28

463-469

111 111

11 111 15

22-28

470-476

111 111

11 111 15

22-28

477-483

111 111

11 111 15

22-28

484-490

111 111

11 111 15

22-28

491-497

111 111

11 111 15

22-28

498-504

111 111

11 111 15

22-28

505-511

111 111

11 111 15

22-28

512-518

111 111

11 111 15

22-28

519-525

111 111

11 111 15

22-28

526-532

111 111

11 111 15

22-28

533-539

111 111

11 111 15

22-28

540-546

111 111

11 111 15

22-28

547-553

111 111

11 111 15

22-28

554-560

111 111

11 111 15

22-28

561-567

111 111

11 111 15

22-28

568-574

111 111

11 111 15

22-28

575-581

111 111

11 111 15

22-28

582-588

111 111

11 111 15

22-28



7. Construct a suitable frequency table for the following data.

1.0, 1.04, 1.08, 1.05, 1.07, 1.05, 1.09, 1.05, 1.04, 1.0, 1.08, 1.07  
 1.04, 1.01, 1.05, 1.06, 1.04, 1.07, 1.05, 1.04, 1.01, 1.03, 1.06, 1.02  
 1.01, 1.04, 1.02, 1.04, 1.02, 1.05, 1.01, 1.04, 1.01, 1.03, 1.06, 1.04  
 1.06, 1.0, 1.07, 1.01, 1.05, 1.00, 1.06, 1.08, 1.09, 1.05, 1.09, 1.03  
 1.0, 1.09, 1.04, 1.08, 1.04, 1.04, 1.05, 1.06, 1.03, 1.02, 1.07, 1.01

A) Here  $N = 65$

By Sturges rule

$$k = 1 + 3.222(\log N)$$

$$= 1 + 3.222(\log 65)$$

$$= 6.8411$$

$\approx 7$

$$i = \frac{\text{Range}}{n}$$

$$= \frac{1.09 - 1.0}{7}$$

$$= 0.1285$$

$$\approx 0.01$$

1.08 - 0.05

1.08 - 0.02

1.08 - 0.01

Here we take 5 classes of size 0.2

C.I. : tally marks of frequency

1.0 - 1.2	III	III	II							
1.2 - 1.4	II	III	II							
1.4 - 1.6	III									
1.6 - 1.8	III									
1.8 - 2.0	III									

65

88 / 21

8. From the following data construct a frequency table by inclusive class intervals there should be 8 classes with a class magnitude of 4.

22, 15, 20, 16, 19, 24, 27, 28, 29, 21, 23, 19, 12, 18, 38, 31, 33, 28, 30, 29, 11, 20, 17, 18, 22, 14, 36, 16, 13, 18, 26, 17, 10, 24, 32, 15, 34, 16, 25, 15, 30, 34, 38, 41, 37, 41, 37, 38, 29, 25, 21, 17, 13, 10, 14, 18, 23, 27, 31, 35.

	C.I	Tally marks	frequency
5-9	5-9		5
10-14	10-14		5
15-19	15-19		5
20-24	20-24		4
25-29	25-29		3
30-34	30-34		4
35-39	35-39		4
40-44	40-44		5
			48
			48

9. The following data gives the age of recently married husbands and wives, in years. Construct a 2 way frequency distribution tables.

Age of husband	27	25	28	26	29	27	26	25	25
Age of wife	21	20	20	21	23	22	20	20	19

A) Let  $x$  denotes the age of husband and  $y$  denotes the age of wife now, the two way frequency table is given by.

x/y	19	20	21	22	23	$\Sigma f_x$
25	II	III	—	—	II	6
26	IV	I	III	—	—	5
27	—	II	I	III	I	5
28	—	II	III	IV	III	3
29	—	—	—	IV	II	2
$\Sigma f_y$	38	5	4	4	5	21

10. The following data gives age in years and blood pressure for a group of 24 persons. Prepare 2 way frequency distribution taking class intervals for ages as 35-45, 45-55, 55-65, 65 ... & BP as 115-130, 130-145, 145-160.

Age	55	36	72	88	65	46	58	50	38	42	41	47	69
BP	131	140	160	124	148	130	132	149	115	145	163	161	159
	60	58	57	43	57	59	44	35	62	67	70		
	161	131	136	141	164	16	128	118	142	157	162		

Let  $x$  denotes the age and  $y$  denotes BP of 24 persons. Then the 2 way frequency table for the above table.

x/y	115-130	130-145	145-160	160-175	$\Sigma f_x$
35-45	III	II	I	I	8
45-55	—	I	I	I	3
55-65	—	III	II	III	8
65-75	—	—	III	II	5
$\Sigma f_y$	4	6	7	7	24

11. Construct Bi-variate frequency distribution of ten marks obtained by 16 students in mathematics and statics by taking class intervals 20-30, 30-40 etc. for both.

marks in maths	37	20	46	28	35	26	41	48	32	23	10
marks in statistics	30	32	41	33	29	43	30	21	44	38	47
	39	44	33	27	26						42
	24	32	21	20	21					38	41

A) Let  $x$  denotes the marks in mathematics and  $y$  denotes the marks in statistics of a student then the tabular way of frequency table for the above data given by.

$2/4$	20-30	30-40	40-50	fx
20-30	11	111	111	7
30-40	111	111	111	5
40-50	111	111	111	111
$f_y$	6	6	6	6

x					180-182	182-184	184-186
s				II			
c					III		
s	III	II	III				
r	II	III		-	-	22-23	21-22
H <sub>2</sub>	F	F	d		H		

Sets:

1. In a group of 120 people, 54 like cococola & 84 like pepsi & each person likes atleast one of the 2 beverages. How many like both cococola & pepsi.

A)

formula

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

2. In a class of 40 student 22 play Hockey, 26 play basketball and 14 play both. How many students play don't either of the games.

A) Let A be the Hockey players

$$n(A) = 22$$

Let B be the basketball players.

$$n(B) = 26$$

The no. of students both Hockey & basketball player  
 $= n(A \cap B) = 14$

Total no. of students = 40

We know that  $n(A \cup B) = n(A) + n(B) - n(A \cap B)$

$$= 22 + 26 - 14$$

$$= 48 - 14 = 34$$

The no. of students either Hockey or basketball players

$$= 34.$$

The no. of students don't play either of the games

$$= 40 - 34 = \underline{\underline{6}}$$

$$(A \cap B) n - (B) n + (A) n = (A \cup B) n$$