Frequency distribution:-

The frequency of a variable is the number of times it occur in a dataset.

A frequency distribution is defined whenever
(i) the values which the variable takes
(ii) the frequency of each variable value
are both specified.

Types of frequency distributions:-

1. Ungrouped frequency distributions
2. Grouped frequency distributions

Raw or ungrouped data:- the data in original form is called raw or ungrouped data.
consider the daily wages (in Rs) of 30 labourers in a factory.

800, 700, 550, 500, 600, 650, 400, 300, 800, 900
750, 450, 350, 650, 700, 800, 820, 550, 650, 800
600, 800, 600, 550, 380, 650, 750, 850, 900, 650

when raw data is arranged in ascending or descending order of magnitude, the data are said to be arranged in an array.

Discrete (ungrouped) frequency distribution :-

In this form of distribution, the frequency refers to discrete value. In this case the data is presented a way that exact measurement of units are cleary visible.

Ex :- In a survey of 40 families, the no. of children per family was recorded and following data obtained.

1	2	3	2	2
0	1	2	2	3
3	0	1	3	4
2	3	5	0	4
1	4	3	2	1
5	2	3	1	2
6	1	2	4	4
2	6	4	5	5

above raw data can be arrange in an array and discrete frequency distribution of above data is given as

Table-1 Discrete freq. distribution

No. of children	frequency
0	3
1	7
2	10
3	8
4	6
5	4
6	2
	Total - 40

classmate

Date

Page

The frequency table presents a clearer picture about the data but to get still better picture about the structure of frequency distribution, we re-classify the data into group. ~~as follows~~.

Grouped frequency distribution or Continuous frequency distribution:-

we arrange the data for the class intervals. This type of representation of frequencies is called a grouped frequency distribution.

Ex:- consider the marks obtained by a class of 60 students in stats paper.

22, 47, 9, 42, 31, 17, 13, 15, 18, 13, 2, 21, 27, 38, 15, 0, 33, 10, 34, 29, 26, 16, 25, 33, 36, 10, 24, 22, 26, 19, 14, 36, 18, 25, 21, 33, 35, 25, 18, 28, 25, 17, 38, 10, 3, 31, 24, 3, 12, 16, 33, 18, 26, 29, 29, 27, 29, 28, 35, 26, 27.

In Group frequency distribution, we classify the data into groups:-

Table-2

Marks	NO. of students	Commutative freq.
0-5	4	4
6-10	1	5
11-15	7	12
16-20	11	23
21-25	6	29
26-30	16	45
31-35	7	52
36-40	6	58
41-45	1	59
46-50	1	60

the class interval 0-5, 6-10 etc are inclusive since they include upper limit of class.

Here 60 students have been divided into 10 groups, 0-5, 6-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50. The groups are also called classes

for any class say

6 — 10

lower limit ↑ upper limit ↑

interval or width of the class = upper limit − lower limit

The value which lies b/w the lower and upper limits of a class is called midvalue or central value.

eg:- for class 21 - 25
mid value is 23.

The commulative frequency of a class is the total no. of observative less than or equal to the upper limit. ie Total no. of all the frequencies up and including that class.

In case of continuous variable, it is not possible to classify the data into groups of the type 0-5, 6-10 etc
For example, the distribution of ages in years of a student in a college.
clearly we can not arrange the data in the age group 16-20 or 21-25 as there can be students having ages b/w 20 and 21. ie In case of continuous variable the data should be classified as
16 - 20, 20 — 24, 24 — 28 etc.
where it being understood that

the students with ages less than 20 are included in the class 16 - 20.

Note:- class of type $a \le x < b$ are ~~class~~ called exclusive, since these exclude upper limit of class.

Ex Wages of 100 employees.

Table-3

Weekly wages	No. of employees
50 - 100	4
100 - 150	12
150 - 200	22
200 - 250	33
250 - 300	16
300 - 350	8
Total	100

continuous freq. distribution

Special case in Grouped frequency distribution :-

In order to ensure the continuity of class limits, we can converts inclusive class intervals into exclusive class intervals in grouped freq. distribution.

Let d = upper limit of any class - lower limit of succeeding class

then class limits for any class are given by:-

upper limit = upper limit + d/2

lower limit = lower limit - d/2.

In order
For example :- given grouped frequency distribution

x	f
15 - 19	9
20 - 24	11
25 - 29	10
30 - 34	30
35 - 39	40

⇓ $d = 20 - 19 = 1$

x	f
14.5 - 19.5	9
19.5 - 24.5	11
24.5 - 29.5	10
29.5 - 34.5	30
34.5 - 39.5	40

Measures of Central Tendency

Sometime a diagram or a graph may fail to convey the clear picture for which it is intended. It is difficult to describe that what graph conveys. Therefore one of the most important objective of statistical analysis to get the single measurement / value that describes the main characterstic of the data. These measures are generally in the central portions of the distributions and known as measures of central Tendency

(1) Arithmatic Mean

(2) Median

(3) Mode

Arithmatic Mean :-

For ungrouped data / Individual observations :-

(1) $$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Q1 The monthly income of 10 families is given below. Calculate the Arithmetic mean of Incomes.

4780, 5760, 4690, 4750, 4840, 4920, 4100, 4850, 5050, 5950

Sol^n $\Sigma X = 49,690$

$N = 10$

$\bar{X} = \frac{\Sigma X}{N}$

$\boxed{\bar{X} = 4969}$

If the values of x are large, then calculation of mean by above method is quite tedious. In this case step deviation (shortcut method) method is very useful to compute the mean.

step deviation method :-

Let A be the assumed mean (arbitrary origin). Then deviations of item are take from the assumed mean

$d = X - A$

$$\text{Mean} = A + \frac{\Sigma d}{N}$$

$$\boxed{\bar{X} = A + \frac{\Sigma d}{N}}$$

S.NO	Monthly income X	deviation $d = X - A$
1	4780	-20
2	5760	960
3	4690	-110
4	4750	-50
5	4840	40
6	4920	120
7	4100	-700
8	4850	50
9	5050	250
10	5950.	1150.
		$\Sigma d = 1690$

Let assumed mean $A = 4800$

$$\bar{X} = A + \frac{\Sigma d}{N}$$

$$\bar{x} = 4800 + \frac{1690}{10}$$

$$\boxed{\bar{X} = 4969}$$ Ans.

Arithmatic Mean

For grouped data :- Discrete data / continuous data

(1) For Discrete series the Arithmetic mean may be computed by applying

(i) Direct method :-

If x_i / f_i, $i = 1, 2, \ldots N$ is freq' distribution then.

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{N} f_i x_i}{N}}$$

where $N = \sum_{i=1}^{N} f_i$

(ii) short cut method.

$$\boxed{\bar{x} = A + \frac{\sum f d}{N}}$$

A = assumed mean

$d = x - A$

$N = \sum f$ = total observation

Q2 From the following data of the marks obtained by 60 students of a class, calculation arithmetic mean.

Marks :	20	30	40	50	60	70
No. of Students :	8	12	20	10	6	4

classmate

Date
Page

Direct method:-

Marks x	No. of students f	fx
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
	$\Sigma f = 60$	$\Sigma fx = 2460$

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{2460}{60} = 41$$

$$\boxed{\bar{x} = 41}$$

Shortcut method:-

Let assumed mean $= 40$

x	f	$d = x - A$	fd
20	8	-20	-160
30	12	-10	-120
40 - A	20	0	0
50	10	10	100
60	6	20	120
70	4	30	120
	$\Sigma f = 60$		$\Sigma fd = 60$

$$\bar{X} = A + \frac{\Sigma fd}{N}$$

$$\bar{X} = 40 + \frac{60}{60}$$

$$\boxed{\bar{X} = 41}$$

(2) For continuous series the arithmatic mean may be computed using
(a) Direct method (2) Shortcut method

Direct method:-
For continuous series, we calculate mid value of each class, denoted by x
then

$$\boxed{\text{Arithmetic mean} = \frac{\Sigma fx}{N}}$$

Q8 **Shortcut method:-**
for continuous series, with class interval of equal magnitude, the calculations are further simplified by taking

$$d = \frac{x - A}{h},$$

where x = mid value of class interval
h = class interval

$$\text{Mean} \; \boxed{\bar{X} = A + \frac{\Sigma fd}{N} \times h}$$

classmate
Date
Page

Q3 from the following data compute the arithmetic mean by direct method.

Marks :-	0 - 10	10 - 20	20-30	30-40	40-50	50-60
No of Students :-	5	10	25	30	20	10

Sol^n by direct method

Marks	No. of Students f	Mid-point x	fx
0 - 10	5	5	25
10 - 20	10	15	150
20 - 30	25	25	625
30 - 40	30	35	1050
40 - 50	20	45	900
50 - 60	10	55	550
	$\Sigma f = 100$		$\Sigma fx = 3,300$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

$$\bar{x} = \frac{3300}{100}$$

$$\boxed{\bar{x} = 33}$$

(b) by step deviation method

Marks	No. of students f	Mid point x	$d = \frac{x-A}{h}$	$f.d$
0 - 10	5	5	-3	-15
10 - 20	10	15	-2	-20
20 - 30	25	25	-1	-25
30 - 40	30	35 - A	0	0
40 - 50	20	45	1	20
50 - 60	10	55	2	20
	$\Sigma f = 100$			$\Sigma fd = -20$

$A = 35$ (assumed mean)

$h = 10$ class interval

$$d = \frac{x - 35}{10}$$

$$\bar{x} = A + \frac{\Sigma fd}{N} \times h$$

$$\bar{x} = 35 + \frac{(-20)}{100} \times 10$$

$$\boxed{\bar{x} = 33}$$ Ans

Q.4 compute the arithmatic mean for following frequency distribution :-

class :-	50-59	60-69	70-79	80-89	90-99	100-109	110-119
frequency :	1	3	8	17	35	4	2

class	cont. class	Midvalue(x)	$d = \frac{x-A}{h}$	f	fd
50-59	49.5 - 59.5	54.5	-3	1	-3
60-69	59.5 - 69.5	64.5	-2	3	-6
70-79	69.5 - 79.5	74.5	-1	8	-8
80-89	79.5 - 89.5	84.5	0	17	0
90-99	89.5 - 99.5	94.5	1	35	35
100-109	99.5 - 109.5	104.5	2	4	8
110-119	109.5 - 119.5	114.5	3	2	6
				$\Sigma f = 70$	$\Sigma fd = 32$

class interval $h = 10$

Let assumed mean = 84.5 = A

$$d = \frac{x - 84.5}{10}$$

$$\bar{x} = A + \frac{\Sigma fd}{N} \times h$$

$$\bar{x} = 84.5 + \frac{32}{70} \times 10$$

$$\boxed{\bar{x} = 89.0714}$$ Ans

Note If class intervals are unequal, then in step deviation method, d is calculated as

$$d = \frac{x - A}{c}$$

where c is suitable divisor of $x - A$

Q:- Marks:- 0-10 10-30 30-60 60-100
freq:- 5 12 25 8

Marks	Midvalue x	freq. f	$x - A$	$d = \frac{x-A}{c}$	fd
0-10	5	5	-40	-8	-40
10-30	20	12	-25	-5	-60
30-60	45 → A	25	0	0	0
60-100	80	8	35	7	56
		N = 50			$\Sigma fd = -44$

Let $c = 5$

$$\bar{x} = A + \frac{\Sigma fd}{N} \times c$$

$$\bar{x} = 45 + \frac{(-49)}{50} \times 5$$

$$\boxed{\bar{x} = 40.6}$$

Prob on finding missing frequencies:

Q1 - For a certain frequency table, which has been partially reproduced here, the mean was found to be 1.46

No. of Accidents	0	1	2	3	4	5	Total
Frequencies	46	?	?	25	10	5	200

x	f	fx
0	46	0
1	f_1	f_1
2	f_2	$2f_2$
3	25	75
4	10	40
5	5	25
	$\Sigma f = 86 + f_1 + f_2$	$\Sigma fx = 140 + f_1 + 2f_2$

Mean $\bar{x} = \frac{\Sigma fx}{\Sigma f} \Rightarrow \Sigma fx = \bar{x} \times \Sigma f$

$\cancel{1.46} = \frac{\cancel{140 + f_1 + 2f_2}}{\cancel{86 + f_1 + f_2}}$

from the table we have

$N = \Sigma f = 86 + f_1 + f_2$

ie $86 + f_1 + f_2 = 200$

$f_1 + f_2 = 114 \quad \text{—} (1)$

also $\Sigma fx = 140 + f_1 + 2f_2$

$\bar{x} \times \Sigma f = 140 + f_1 + 2f_2$

$140 + f_1 + 2f_2 = 1.46 \times 200$

$f_1 + 2f_2 = 292 - 140$

$f_1 + 2f_2 = 152$ —— ②

solving ① & ②

$f_1 = 76, \quad f_2 = 38$

Q2 Find the missing frequency from the following data, given that average marks is 16.82

Marks	freq.	Midvalue(x)	fx
0-5	10	2.5	25
5-10	12	7.5	90
10-15	16	12.5	200
15-20	f_4	17.5	$17.5 f_4$
20-25	14	22.5	315
25-30	10	27.5	275
30-35	8	32.5	260

$\Sigma f = 70 + f_4$

$\Sigma fx = 1165 + 17.5 f_4$

$\bar{x} = \frac{\Sigma fx}{\Sigma f}$

$16.82 = \frac{1165 + 17.5 f_4}{70 + f_4}$

$1177.4 + 16.82 f_4 = 1165 + 17.5 f_4$

$12.4 = .68 f_4$

$\boxed{f_4 = 18.23}$