MS 351	Probability and Statistics	3-1-0-4
1. Basic Statistics: Mean, mode, median for grouped and ungrouped data, variance, standard deviation, dispersion and measures of dispersion, skewness and measures of skewness, moments.		[10]
2. ProbabilityTheory: Recapitulation of basic probability theory, law of addition and law of multiplication, conditional probability, total probability, Bayes' theorem and its applications.		[08]
3. Random Variables and Distribution: Discrete and continuous random variables, types of distribution, probability mass and density functions,mean variance and standard deviation of distribution,discrete distributions (binomial and Poisson), continuous distributions (normal and uniform).		[13]
4. Hypothesis and Correlation-Regression: Simple and composite hypothesis, type I and type II errors, tests of significance, simple correlations, linear regression.		[08]
Text Books: 1. Elements of Probability and Statistics by A.P. Baisnab and Manoranjan Jas, TMH 2003. 2. Fundamental of Mathematical Statistics by Gupta and Kapoor, SCS Publisher, 2002. 3. Introduction to Probability and Statistics by J. Susan Milton and Jesse C. Arnold, Fourth Edition, McGraw Hill Education, 2016.		
Reference Books: 1. Probability, Random Variables and Stochastic Processes by Papoulis, TMH. 2. Probability and Statistics (Schaum's outline Series) by Spiegel, TMH. 3. Mathematical Statistics with Applications by Miller and Freund's, 7th Edition, Pearson.		

Scanned with CamScanner

Kinds of average or measures of Central tendency

(i) Arithmetic Mean / Arithmetic average / Simple mean
Symbol (A.M. or M).
(ii) Median Symbol "M_d"
(iii) Mode Symbol "M_o"
(iv) Geometric Mean Symbol "G.M." or G.
(v) Harmonic Mean Symbol "H.M." or H.

Note :- Arithmetic Mean, Geometric Means and Harmonic Means are Called Mathematical average.
Mode and Median are Called positional averages

There are three methods to find the mean method
(i) Direct Method (ii) short cut method (iii) Step deviation

(i) Direct Method :- Let $x_1, x_2 \cdots x_n$ are the variates. then the arithmetic mean.

$$A.M. = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$= \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\Sigma x}{n}.$$

Scanned with CamScanner

Scanned with CamScanner

(2) In a Discrete Series.

when frequency is given.

$x_1 \quad x_2 \quad \cdots \quad x_n$

$f_1 \quad f_2 \quad \cdots \quad f_n$

$$A.M = \frac{x_1f_1 + x_2f_2 + \cdots + x_nf_n}{\Sigma f_n = N} = \frac{\sum_{i=1}^{n} x_i f_i}{N}.$$

(3) In a grouped or Continuous Series :-

when the class interval is given (30-40)

then we find the mid value of x.

$$x = \frac{1}{2}(30-40) = 35.$$

(2) Short Cut Method :-

$$A.M. = A + \frac{\Sigma fu}{N} \qquad (u = x - A.)$$

where

$N = f_1 + f_2 + \cdots + f_n.$

(3) Step deviation Method :-

$$A.M = A + i.\frac{\Sigma fu}{N}$$

where :-

$A \rightarrow$ assumed mean.

$i \rightarrow$ interval.

$u \rightarrow \frac{x-A}{i}.$

Scanned with CamScanner

$$= \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n} = \frac{1}{N} \sum_{i=1}^{n} f_i x_i \qquad \ldots (3)$$

or
$$M = \frac{\Sigma fx}{N}$$

(3) **In a Grouped or Continuous Series.** In this case the arithmetic mean is given by the above formula (3) where x's denote mid-value of the class intervals. For example, for the class 30—40, the value of $x = \frac{1}{2}(30 + 40)$ *i.e.*, 35.

Illustrative Examples

Example 1 (a). *The marks obtained by nine students in a paper of Statistics are as follows :*

52, 75, 40, 70, 43, 40, 65, 35, 48.

Calculate the mean.

Sol. By direct method A. M. (*i.e.* M) is given by

$$M = \frac{\Sigma fx}{n} = \frac{52 + 75 + 40 + 70 + 43 + 40 + 65 + 35 + 48}{9} \qquad \textbf{Ans.}$$
$$= 468/9 = 52$$

Example 1 (b). *Find the arithmetic mean (A. M.) of first n natural numbers.*

[Vikram 1990; Sagar 94; Ravishankar 94 S Indore 96]

Sol. Here x : 1 2 3 ..... n.

∴ By direct method, the required arithmetic mean (M) is given by

$$M = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1 + 2 + 3 + \ldots + n}{n} = \frac{\frac{1}{2} n(n+1)}{n} = \frac{1}{2}(n+1) \qquad \textbf{Ans.}$$

Example 2 (a). *Find the mean height of the students from the following frequency distribution :*

Height (in inches.)	64	65	66	67	68	69	70	71	72	73
No. of Students	1	6	10	22	21	17	14	5	3	1

Scanned with CamScanner

Sol. By direct method.

Height (in inches) x	No. of Students f	fx
64	1	64
65	6	390
66	10	660
67	22	1474
68	21	1428
69	17	1173
70	14	980
71	5	355
72	3	216
73	1	73
	$\Sigma f = N = 100$	$\Sigma fx = 6813$

∴ Arithmetic mean height (M) is given by

$$M = \frac{1}{N}\Sigma f x = \frac{6813}{100} = 68{\cdot}13 \text{ inches.}$$ **Ans.**

Example 2 (b). *Compute the arithmetic mean of the marks from the following table :*

Marks	0—10	10—20	20—30	30—40	40—50	50—60
No. of students	12	18	27	20	17	6

Sol. By directed method.

Marks	Mid Value x	No. of students f	fx
0—10	5	12	60
10—20	15	18	270
20—30	25	27	675
30—40	35	20	700
40—50	45	17	765
50—60	55	6	330
Total		$N = 100$	$\Sigma fx = 2800$

Required A. M. (M) is given by

$$M = \frac{1}{N}\Sigma fx = \frac{2800}{100} = 28.$$ **Ans.**

Scanned with CamScanner

Q.1 Compute the arithmetic mean (A.M.) of the following by both direct and Short Cut methods.

Height in CM. :	219	216	213	210	207	204	201	198	195
f :	2	4	6	10	11	7	5	4	1.

Sol:-

Height in CM. x	f	fx	$u = x - A$	fu
219	2	438	12	24
216	4	864	9	36
213	6	1278	6	36
210	10	2100	3	30
$\boxed{207}$	11	2277	0	0
204	7	1428	−3	−21
201	5	1005	−6	−30
198	4	792	−9	−36
195	1	195	−12	−12
	$N=50$	$\Sigma fx = 10377$		$\Sigma fu = 27$

① By direct method

$$M = \frac{\Sigma fx}{N} = \frac{10377}{50} = 207.54 \text{ cm.}$$

② By Short Cut-method. Let assume mean $A = 207$

$$M = A + \frac{\Sigma fu}{\Sigma f} = 207 + \frac{27}{50} = 207 + 0.54 = 207.54 \text{ CM.}$$

Scanned with CamScanner

Ex:- Compute the mean of the following by both direct and short cut method

Class :-	20-30	30-40	40-50	50-60	60-70
Frequency :-	8	26	30	20	16 .

Sol :-

Class	Mid Value x	f	fx.	$u = x - A$. ($A = 45$)	fu.
20-30	25	8	200	-20	-160
30-40	35	26	910	-10	-260
40-50	45	30	1350	0	0
50-60	55	20	1100	10	200
60-70	65	16	1040	20	320
Total		$N=100$	$\Sigma fx = 4600$		$\Sigma fu = 100$.

By direct method

$$M = \frac{\Sigma fx}{\Sigma f} = \frac{4600}{100} = 45.$$

By Short cut method. Let assumed mean $A = 45$

$$M = A + \frac{\Sigma fu}{\Sigma f} = 45 + \frac{100}{100} = 45.$$

Scanned with CamScanner

Step deviation method

Ex:- Compute the mean of the following frequency distribution

Class :-	0–10	11–22	22–33	33–40	44–55	55–66
Frequency :-	9	17	28	26	15	8

Sol :-

Class	Mid-value x	f	$(A = 38.5)$ $u = x - A$	$u = (x-A)/i$ $i = 11$	fu
0–11	5.5	9	−33	−3	−27
11–22	16.5	17	−22	−2	−34
22–33	27.5	28	−11	1	−28
33–44	38.5	26	0	0	0
44–55	49.5	15	11	1	15
55–66	60.5	8	22	2	16
		$N = 103$			$\Sigma fu = -58$

Let the assumed mean. $A = 38.5$ then.

$$M = A + i\frac{\Sigma fu}{N} = 38.5 + \frac{11(-58)}{103}$$

$$= 38.5 - \frac{638}{103} = 38.5 - 6.194 = 32.306.$$

Ex:- Find the average marks of the students from the following table.

Marks : ~~No. of student~~	No. of student	Marks.	No. of student.
Above 0	80	Above 60	28
Above 10	77	Above 70	16
Above 20	72	Above 80	10
Above 30		Above 90	8
Above 40	65	Above 100	0.
Above 50	55		

Scanned with CamScanner

Sol :- Here we are given the Cummulative freq of distribution. Therefore we shall it into Simple freq. distribution. Thus.

Marks.	No. of students.
0-10	3 → (80-77)
10-20	5 → (77-72)
20-30	7 → (72-65)
30-40	10 → (65-55)
40-50	12 → Similarly find remaining fre
50-60	
60-70	
70-80	
80-90	
90-100	

Hence the frequency table can be arranged as follows.

Let A = assumed mean = 55, i = class interval = 10.

Marks	Mid-term.	f.	$u = x-55$	$u = x-55/10$	fu.
0-10	5	3	-50	-5	-15
10-20	15	5	-40	-4	-20
20-30	25	7	-30	-3	-21
30-40	35	10	-20	-2	-20
40-50	45	12	-10	-1	-12
50-60	55	15	0	0	0
60-70	65	12	10	1	12
70-80	75	6	20	2	12
80-90	85	2	30	3	6
90-100	95	8	40	4	32
		N=80			$\Sigma fu = -26$

Scanned with CamScanner

By step deviation method the average marks (A.M.) are given by.

$$M = A + i\,\frac{\Sigma f u}{N} = 55 + 10\left(\frac{-26}{80}\right)$$

$$= 55 - 3.25 = 51.75.$$

Scanned with CamScanner

07/08/20

Median :— The median is defined as the measure of the Central term, when the given terms are arranged in the ascending or descending order of magnitudes.

Note :— The median is that value of the variable which divides the group into two equal parts one part comprising all values greater and the other all values less than the median.

Computation of median :—

(a) Median in Individual Series :—

Let 'n' be the total number of terms.

Two Cases arises :—

Case I :— In n is odd then the value of $\frac{1}{2}(n+1)^{th}$ term gives the median.

Case II :— In n is even then there are two Central terms i.e. $\left(\frac{n}{2}\right)^{th}$ & $\left(\frac{n}{2}+1\right)^{th}$.

The mean of these two values gives median.

$$M_d = \frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2}$$

Scanned with CamScanner

(b) Median in discrete series :- First of all we write the values of the variates in ascending or descending order of magnitudes, then find Cummulative frequencies. Now we find median as in Case I or II of (a) above.

(c) Median in Continuous Series (or grouped series) :- In this Case, the median (M_d) is Computed by the following formula.

$$M_d = l + \frac{\frac{N}{2} - F}{f} \times i .$$

Where :- l = lower limit of median class.
N = total frequency.
F = total of all frequencies before median class.
f = frequency of median class.
i = Class width of median class.

Scanned with CamScanner

Q.1 According to the census of 1991, following are the population figure in thousands of 10 cities
1400, 1250, 1670, 1800, 700, 650, 570, 488, 2100, 1700
find the median.

Sol :- Arranging the terms in ascending order
488, 570, 650, 700, 1250, 1400, 1670, 1700, 1800, 2100
Here $n=10$, therefore the median is the mean of the measure of the 5^{th} and 6^{th} terms
Here 5^{th} term is 1250 and 6^{th} term is 1400

$$M_d = \frac{1250+1400}{2} \text{ Thousands.}$$

$$\boxed{M_d = 1325}$$

Q.2 Below are given the marks obtained by a batch of 15 Students in a certain test in Mathematics and english

Students. x	
1	In which Subject is the level of knowledge of the Students higher.
2	
3	
4	
5	
6	
7	
8	
9	
10	

Scanned with CamScanner

student.	Marks in Maths.	Marks in English
1	46	42
2	20	24
3	41	38
4	43	35
5	25	30
6	54	45
7	47	58
8	36	50
9	30	40
10	61	42
11	50	55
12	63	54
13	45	52
14	56	47
15	58	43

arrange in ascending

x	M. in Maths.	M. in E.
1	20	24
2	25	30
3	30	35
4	36	38
5	41	40
6	43	42
7	45	43
8	46	45
9	47	47
10	50	50
11	54	54
12	56	56
13	58	58
14	61	61
15	63	63

Median Marks in Maths = measure of $\left(\frac{15+1}{2}\right)^{th}$ term

= Measure of 8^{th} term = 46

Median Marks in English = Measure of $\left(\frac{15+1}{2}\right)^{th}$ term

= Measure of 8^{th} term = 45

Since the median marks in Maths are more than those in English hence the level of knowledge in Maths is higher.

Scanned with CamScanner

Q.3. Find the median for the following distribution

x	f	Cummulative Frequencies.
0-10	22	22
10-20	38	$F = \boxed{60} = (38+22)$
$l =$ 20-30	$f = 46$	$106 = (22+38+46)$
30-40	35	$141 = (22+38+46+35)$
40-50	20	$161 = (22+38+46+35+20)$
	$N = \boxed{161}$	

$i = 10$

Here $\boxed{N = \Sigma f_i = 161}$

Now $\left(\frac{161+1}{2}\right)^{th}$ term $= 81^{th}$ term. (this number is Situated in the class $\boxed{20-30}$

Hence (20-30) is the median class.

$$\boxed{M_d = l + \frac{N/2 - F}{f} \times i.}$$

where:- $N = 161$

$F = 60$ &

$f = 46$

$i = 10$

$l = 20$

$$= 20 + \frac{\frac{161}{2} - 60}{46} \times 10$$

$$M_d = 20 + \frac{205}{46} = 20 + 4.46 = 24.46$$

$$\boxed{M_d = 24.46}$$

Scanned with CamScanner

Date 08/08/20

Q.4 Calculate the median for the following data

class	frequency
0-4	4
4-6	6
6-8	8
8-12	12
12-18	7
18-20	2

Sol:- Here the class intervals are unequal and therefore arranging the frequencies as follows.

class	f	c.f.
0-4	4	4
4-8	6+8=14	18 → F
$\boxed{8-12}$	f= $\boxed{12}$	$\boxed{30}$
12-16	5	35
16-20	4	39
	n=39	

Here $\Sigma f_i = N = 39$, which is odd

then Median $= \frac{1}{2}(N+1)^{th}$ term.

$= \frac{1}{2}(39+1) = 20^{th}$ term.

Now the 20^{th} term lies in the class interval (8-12)

$\therefore$ $l = 8$, $F = 18$, $f = 12$, $i = 4-0 = 4$.

Median $(M_d) = l + \frac{\frac{N}{2} - F}{f} \times i$

$= 8 + \frac{\frac{39}{2} - 18}{12} \times 4$

$= 8 + 0.5 = \boxed{8.5 \text{ Ans}}$

Scanned with CamScanner

Date
Page

Q.1 (a) Find the median of the following
20, 18, 22, 27, 25, 12, 5.

Ans: 20

Q.2. Find the median of the following
65, 67, 69, 61, 60, 65, 66, 70, 71, 62, 72.

Ans 66

Q.3. The daily wages of 10 workers in a factory are 4, 6, 9, 12, 11, 8, 5, 10, 11, 5. Find the median

Ans. 8.5

Q.4 Find Median.

Measure:	3	5	7	9	11	13	15
Frequency:	7	3	12	28	10	9	6

Ans [illegible]

Q.5. Find Median.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Students	2	18	30	45	35	22	6	3

Ans :- 36.66

Q.6. Calculate median.

Wages in RS	0-10	10-20	20-30	30-40	40-50
No of workers	22	38	46	35	25

Ans:- (24.57)

Scanned with CamScanner

Mode :- The mode is the value of the variate for which the frequency is maximum.

(i) When the values of all the terms are given. In this case the mode is the value of the term which occurs most frequently.

Ex :- Find the mode from the following sizes of shoes
Size of shoes : 3, 4, 2, 1, 7, 6, 6, 7, 5, 6, 8, 9, 5.

Sol :- Firstly writing the individual observations in a discrete series as follows.

Size of Shoes :	1	2	3	4	5	$\boxed{6}$	7	8	9
Frequency :	1	1	1	1	2	3	2	1	1

Max

Here maximum frequency is 3 whose term value is '6'. Hence the mode is modal size number '6'.

(ii) For a frequency distribution :

$$\text{Mode } (M_o) = l + \frac{f - f_1}{2f - f_{-1} - f_1} \times i.$$

$l \rightarrow$ lower limit of modal class.
$f \rightarrow$ frequency of modal class.
$f_{-1} \rightarrow$ frequency of the class just preceding to the modal class.
$f_1 \rightarrow$ frequency of the class just following of the modal class.
$i \rightarrow$ Class interval.

Scanned with CamScanner

Date
Page

Q.2. Find the mode of the following distribution.

class	0-7	7-14	14-22	21-28	28-35	35-42	42-49
Frequency	19	25	36	$\boxed{72}$	51	43	28
			f_{-1}	Max f	f_1		

Sol:- Here maximum frequency 72 lies in the class 21-28
Thus 21-28 is the modal class

$l = 21, \; f = 72, \; f_{-1} = 36, \; f_1 = 51, \; i = 7.$

$$\text{Mode } (M_o) = l + \frac{f - f_1}{2f - f_{-1} - f_1} \times i.$$

$$= 21 + \frac{72 - 36}{144 - 36 - 51} \times 7.$$

$$= 21 + \frac{84}{19} = 21 + 4.4 = 25.4.$$

Scanned with CamScanner

Date
Page

(iii) Mode by method of grouping :- This method is usually applied in the cases when there are two maximum frequencies against two different size of items.

* Firstly the items are arranged in ascending or descending order and corresponding frequencies are written aginst them.
* The frequencies are then grouped in two and then in threes and then in fours ---

Ex:- Compute the mode from the following.

Size of Item :	4	5	6	7	8	9	10	11	12	13
Frequency ;	2	5	8	9	12	14	14	15	11	13

Sol:- From the given data we observe that size 11 has the maximum frequency 15, but it is possible that the effect of neighbouring frequencies on the size of the item may greater.

Thus it may happen that the frequencies of size 10 or 12 may be greater and 11 may not remain mode. we shall apply the method of grouping

Scanned with CamScanner

Date ____
Page ____

Size of Items	I	II	III	IV	V	VI
4	2	7		2		
5	5		5+8=13	5 + = 15	5	
6	8	17		8	8 + = 22	
7	9				9	8
8	12	26	9+12=21	9	12	9 = 29
9	14			12 + = 35	14 + = 40	12
10	14	29	14+14=28	14	14	14
11	15			14		14 = 43
12	11	24	15+11=26	15 + = 40	15	15
13	13			11	11 + = 39	
					13	

I	II	III	IV	V		
2	7	2		2	2	2
5		5		5	5	5
8	17	8		8	8	8
9		9		9	9	9
12	26	12		12	12	12
10 14		14		14	14	14
11 15	29	14		14	14	14
11	24	15		15	15	15
13		11		11	11	11
		13		13	13	13

Scanned with CamScanner

$n_2 M_2 + \ldots + n_k M_k$ Number of Plants 1 4 5 7 11 10 6 4 66 67 68 69 70 71 72 73

Now we shall find the size of the item containing maximum frequency :-

Column	Size of item having maximum frequency
I	11
II	10, 11
III	9, 11
IV	10, 11, 12
V	8, 9, 10
VI	9, 10, 11

Here size 8 occurs 1 time, 9 occures 3 times 10 occures 5 times, 11 occures 4 times, 12 occurs 1 time.

Since 10 occurs maximum number of times (5 times) Hence the required mode is size 10.

Scanned with CamScanner

Find mode.

Q.1. 15 25 23 27 40 25 23 25 20 21 25

Ans 25

Q.2. Find mode.

x	2	3	4	5	6	7	8	9	10	11	12	13
f	3	8	10	12	16	14	10	8	17	5	4	1

$\boxed{\text{Ans} \to 16}$

Q.3 Find mode.

x	f I	II	III	IV	V	VI
10	8	8+12	leave Ist one (12+36)	8+12+36	leave Ist one. 12+36+35	leave Ist (2) 36+35+28 =99
15	12					
20	36					
25	35					
30	28					
35	18					
40	9					

$\boxed{\text{Ans} = 35}$

Q.4 Find mode.

x :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f :	6	10	10	16	12	8	7

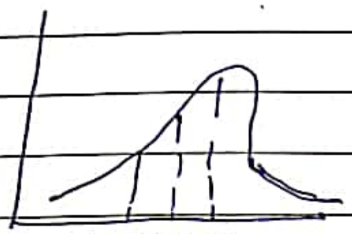
Ans → 35

Scanned with CamScanner

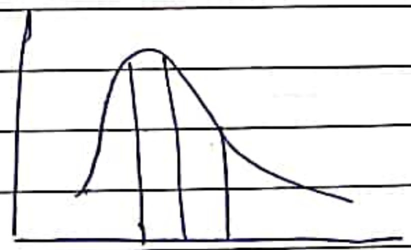
Relation between Median & Mode.

$$\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$$ for asymmetrical distribution.

① mean < median < mode.
$(M < M_d < M_o)$



$M < M_d < M_o$



$M_o < M_d < M$.

For symmetrical distribution.
Mean, Median & Mode coincide.

Scanned with CamScanner

Illustrations

Illustration 1. Obtain the median for the following frequency distribution.

x :	3	8	10	12	16	18	21	25	28
f :	8	10	11	16	20	25	15	9	6

Solution. First we prepare cumulative frequency table.

x	f	c.f.	
3	8	8	
8	10	18	
10	11	29	
12	16	45	
16	20	65	> Just greater than N/2.
18	25	90	
21	15	105	
25	9	114	
28	6	120	
	N= 120		

Here, $N = 120 \quad > \quad \frac{N}{2} = 60$

We find the cumulative frequency just greater than N/2 i.e. 60 is 65 and the value of x corresponding to 65 is 16.

Therefore, Median = 16.

Scanned with CamScanner

4.0 MODE OF A GROUPED DATA

Mode: Mode is that value among the observations which occurs most often i.e. the value of the observation having the maximum frequency.

In a grouped frequency distribution, it is not possible to determine the mode by looking at the frequencies.

Modal Class: The class of a frequency distribution having maximum frequency is called modal class of a frequency distribution.

The mode is a value inside the modal class and is calculated by using the formula.

$$\text{Mode} = \ell + \left\{\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right\} \times h$$

Where ℓ = Lower limit of the modal class.

h = Size of class interval

f_1 = Frequency of modal class

f_0 = Frequency of the class preceding the modal class

f_2 = Frequency of the class succeeding the modal class.

Scanned with CamScanner

Illustration 4. Find the mean marks from the following data :

Marks	No. of Students
Below 10	5
Below 20	9
Below 30	18
Below 40	29
Below 50	45
Below 60	60
Below 70	70
Below 80	78
Below 90	83
Below 100	85

Solution We may prepare the table as given below :

Marks	No. of students	Class Interval	f_i	Class mark (x_i)	$f_i x_i$
Below 10	5	0-10	5	5	25
Below 20	9	10-20	4	15	60
Below 30	18	20-30	9	25	225
Below 40	29	30-40	11	35	385
Below 50	45	40-50	16	45	720
Below 60	60	50-60	15	55	825
Below 70	70	60-70	10	65	650
Below 80	78	70-80	8	75	600
Below 90	83	80-90	5	85	425
Below 100	85	90-100	2	95	190
			N = 85		$\Sigma f_i x_i = 4105$

$\therefore$ Mean , $\bar{x} = \frac{\sum f_i x_i}{N} = \frac{4105}{85} = 48.29$

Scanned with CamScanner

Illustration 5. Find the mean marks of students from the adjoining frequency distribution table

Marks	No. of Students
Above 0	80
Above 10	77
Above 20	72
Above 30	65
Above 40	55
Above 50	43
Above 60	23
Above 70	16
Above 80	10
Above 90	8
Above 100	0

Solution We may prepare the table as given below :

Marks	No. Of students	Class Interval	f_i	Class mark (x_i)	f_ix_i
Above 0	80	0-10	3	5	15
Above 10	77	10-20	5	15	75
Above 20	72	20-30	7	25	175
Above 30	65	30-40	10	35	350
Above 40	55	40-50	12	45	540
Above 50	43	50-60	20	55	1100
Above 60	23	60-70	7	65	455
Above 70	16	70-80	6	75	450
Above 80	10	80-90	2	85	170
Above 90	8	90-100	8	95	760
Above 100	0	100-110	0	105	0
			N = 80		Σf_ix_i = 4090

$\therefore$ Mean, $\bar{x} = \frac{\sum f_i x_i}{N} = \frac{4090}{80} = 51.125 = 51.1$ (approx)

Scanned with CamScanner

3.1 Merits and Demerits of median

The following are some merits and demerits of median :

Merits

(i) It is easy to compute and understand.

(ii) It is well defined an ideal average should be.

(iii) It can also be computed in case of frequency distribution with open ended classes.

(iv) It is not affected by extreme values.

(v) It can be determined graphically.

(vi) It is proper average for qualitative data where items are not measured but are scored.

Demerits

(i) For computing median data needs to be arranged in ascending or descending order.

(ii) It is not based on all the observations of the data.

(iii) It cannot be given further algebraic treatment.

(iv) It is affected by fluctuations of sampling.

(v) It is not accurate when the data is not large.

(vi) In some cases median is determined approximately as the mid-point of two observations whereas for mean this does not happen.

Scanned with CamScanner

Illustration 3. A life insurance agent found the following data for distribution of ages of 100 policy holders. Calculate the median age, if policies are given only to persons having age 18 years onwards but less than 60 years.

Age (in years)	No. of policy holders
Below 20	2
Below 25	6
Below 30	24
Below 35	45
Below 40	78
Below 45	89
Below 50	92
Below 55	98
Below 60	100

Solution From the given table we can find the frequency and cumulative frequencies as given below :

Age (in years)	No. of policy holders (f_i)	Cumulative frequency
15-20	2	2
20-25	4	6
25-30	18	24
30-35	21	45
35-40	33	78
40-45	11	89
45-50	3	92
50-55	6	98
55-60	2	100
	N = 100	

Here, N = 100

$$\therefore \quad \frac{N}{2} = 50$$

Scanned with CamScanner

4.1 Merits & Demerits of mode

1. It can be easily understood and is easy to calculate.
2. It is not affected by extreme values.
3. It can be found by inspection in some cases.
4. It can be determined in distributions with open classes.
5. It can be represented graphically.

Scanned with CamScanner

Demerits of mode

1. It is ill-defined. It is not always possible to find a clearly defined mode.
2. It is not based upon all the observation.
3. It is affected to a greater extent by fluctuations of sampling.

Illustrations

Illustration 4. The following data gives the information on the observed lifetimes (in hours) of 225 electrical components :

Lifetimes (in hours)	0-20	20-40	40-60	60-80	80-100	100-120
Frequency	10	35	52	61	38	29

Determine the modal lifetimes of the components.

Solution Here the class 60-80 has maximum frequency, so it is the modal class.

$\therefore \quad \ell = 60,\ h = 20,\ f_1 = 61,\ f_0 = 52$ and $f_2 = 38$

Therefore, mode $= \ell + \left\{\dfrac{f_1 - f_0}{2f_1 - f_0 - f_2}\right\} \times h$

$$= 60 + \left(\frac{61 - 52}{2 \times 61 - 52 - 38}\right) \times 20$$

$$= 60 + \frac{9}{32} \times 20$$

$$= 60 + 5.625$$

$$= 65.625$$

Hence, the modal lifetimes of the components is 65.625 hours.

Illustration 5. Given below is the frequency distribution of the heights of players in a school.

Heights (in cm)	160-162	163-165	166-168	169-171	172-174
No. of students	15	118	142	127	18

Find the average height of maximum number of students.

Solution The given series is in inclusive form. We prepare the table in exclusive form, as given below :

Heights (in cm)	159.5-162.5	162.5-165.5	165.5-168.5	168.5-171.5	171.5-174.5
No. of students	15	118	142	127	18

We have to find the mode of the data.

Here, the class 165.5-168.5 has maximum frequency, so it is the modal class.

Scanned with CamScanner

Illustration 4. The following data gives the information on the observed lifetimes (in hours) of 225 electrical components :

Lifetimes (in hours)	0-20	20-40	40-60	60-80	80-100	100-120
Frequency	10	35	52	61	38	29

Determine the modal lifetimes of the components.

Solution Here the class 60-80 has maximum frequency, so it is the modal class.

$\therefore \quad \ell = 60,\ h = 20,\ f_1 = 61,\ f_0 = 52$ and $f_2 = 38$

$$\text{Therefore, mode} = \ell + \left\{\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right\} \times h$$

$$= 60 + \left(\frac{61 - 52}{2 \times 61 - 52 - 38}\right) \times 20$$

$$= 60 + \frac{9}{32} \times 20$$

$$= 60 + 5.625$$

$$= 65.625$$

Hence, the modal lifetimes of the components is 65.625 hours.

Illustration 5. Given below is the frequency distribution of the heights of players in a school.

Heights (in cm)	160-162	163-165	166-168	169-171	172-174
No. of students	15	118	142	127	18

Find the average height of maximum number of students.

Solution The given series is in inclusive form. We prepare the table in exclusive form, as given below :

Heights (in cm)	159.5-162.5	162.5-165.5	165.5-168.5	168.5-171.5	171.5-174.5
No. of students	15	118	142	127	18

We have to find the mode of the data.

Here, the class 165.5-168.5 has maximum frequency, so it is the modal class.

Scanned with CamScanner

Measures of Dispersion, Skewness, Moments and Kurtosis

§ 3·1. **Dispersion or Variation.** The averages give an idea of central tendency of the given distribution but it is necessary to know how the variates are clustered around or scattered away from the average. To explain it more clearly consider the works of two typists who typed the following number of pages in 6 working days of a weak :

	Mon.	Tues.	Wed.	Thu.	Fri.	Sat.	Total Page
I typist :	15	20	25	25	30	35	150
II typist :	10	20	25	25	30	40	150

We see that each of the typist I and II typed 150 pages in 6 working days and so the average in both the cases is 25. Thus there is no difference in the average, but we know that in the first case the number of pages varies from 15 to 35 while in the second case the number of pages varies from 10 to 40. This denotes that the greatest deviation from the mean in the first case is 10 and in the second case it is 15 *i.e.* there is a difference between the two series. The variation of this type is termed **scatter** or **dispersion** or **spread.**

Definition. *The degree to which numerical data tend to spread about an average value is said variation or dispersion or spread of the data.*

§ 3·2. **Desiderata for a Satisfactory Dispersion.**

To obtain a satisfactory dispersion, we require the following essential requisities :

(i) It should be easily calculated, rigidly defined and based on all observations.

(ii) It should be readily comprehensive and amenable to algebraic treatment.

(iii) It should be least affected by fluctuations.

§ 3·3. **Measures of Dispersion.** **[Ravishankar (old) 1989]**

Various measures of dispersion or variation are available, the most common are the following :

1. Range.
2. Quartile deviation or Semi-interquartile range.
3. Average deviation or mean deviation.
4. Standard deviation.

Scanned with CamScanner

§ 3·4. **The Range.** It is the simplest possible measure of dispersion. *The range of a set of numbers (data) is the difference between the largest and the least numbers in the set (i.e. values of the variable).* If this difference is small then the series of numbers is supposed regular and if this difference is large then the series is supposed to be irregular.

For Example. In the example of § 3·1 above the range for typist I is $35 - 15 = 20$ and that for typist II is $40 - 10 = 30$.

Demerits.

(i) It depends only on the extreme values of the variable.

(ii) It is subject to fluctuations of considerable magnitude from sample to sample.

(iii) If a giant or dwarf is included in the series representing the heights of persons, then the range will considerably change.

§ 3·5. Quartile Deviations. **[Agra 1983]**

Definition. The inter-quartile range of a set of data is defined by

$$\textbf{Inter-quartile range} = \mathbf{Q_3 - Q_1}$$

where Q_1 and Q_3 are respectively the first and third quartiles for the data.

Semi-interquartile range is denoted by Q and is defined by

$$Q = \frac{1}{2}(Q_3 - Q_1)$$

where Q_1 and Q_3 have the same meaning as given above.

The semi-interquartile deviation is a better measure of dispersion than the range and is easily computed. Its drawback is that it does not take into account all the items.

The interdecile range is defined by $D_9 - D_1$ where D_1 and D_9 are the first and the ninth deciles of a set of data. The 80% of total frequency lie in this range.

The 10 – 90 percentile range is defined by $P_{90} - P_{10}$, where P_{10} and P_{90} are the 10th and 90th percentiles of a set of data.

§ 3·6. Average Deviation or Mean Deviation.

Definition. **[Ravishankar 1985; Indore 89]**

The average of mean deviation of a set of N numbers $x_1, x_2, \ldots, x_N$ is defined by

$$\textbf{Mean Deviation (M. D.)} = \delta_m = \frac{1}{N}\sum_{i=1}^{N} |x_i - M|$$

$$= \frac{1}{N}\Sigma |x - M|$$

Where M is the mean or median or mode according as the mean deviation from the mean or median or mode is to be computed $|x_i - M|$ represents the absolute (or numerical) value. Thus $|-5| = 5$.

If $x_1, x_2, \ldots, x_k$ occur with frequencies $f_1, f_2, \ldots, f_k$ respectively, then the mean deviation (δ_m) is defined by

Scanned with CamScanner

$$\delta_m = \frac{1}{N}\sum_{j=1}^{k} f_j \mid x_i - M \mid = \frac{1}{N}\Sigma f \mid x - M \mid .$$

Mean deviation depends on all the values of the variables and therefore it is a better measure of dispersion than the range or the quartile deviation. Since signs of the deviations are ignored (because all deviations) are taken positive, some artificiality is created.

Illustrative Examples

Example 1. *Find the mean deviation from the arithmetic mean of the following distribution :*

Marks :	0—10	10—20	20—30	30—40	40—50
No. of Students :	5	8	15	16	6

[Jabalpur 1994; Vikram 78]

Sol. Let assumed mean $(A) = 25$ and $i = 10$

Class	Mid Value x	Frequency f	$u = \frac{x-A}{i}$	fu	$x_i - M$	f| x–M|
0—10	5	5	– 2	– 10	– 22	110
10—20	15	8	– 1	– 8	– 12	96
20—30	25	15	0	0	– 2	30
30—40	35	16	1	16	8	128
40—50	45	6	2	12	18	108
Total		$N = \Sigma F = 50$		$\Sigma fu = 10$		$\Sigma f \mid x - M \mid = 472$

$\therefore$ Arithmetic mean $M = A + \frac{\Sigma fu}{N} \times i = 25 + \frac{10}{50} \times 10 = 27$

$\therefore$ The required mean deviation from arithmetic mean

$$\delta_m = \frac{\Sigma f \mid x - M \mid}{N} = \frac{472}{50} = 9{\cdot}44.$$

Example 2. *Find the mean deviation from the mean from the following series :*

Age (Less then) :	10	20	30	40	50	60	70	80
No. of persons :	15	30	53	75	100	110	115	125

[Vikram 1978]

Scanned with CamScanner

Sol.

Age (less than)	Class	Cumulative frequency c. f.	Frequency f	Mid-Value x	$u=\frac{x-45}{10}$	fu	x–M	f|x–M|
10	0–10	15	15	5	– 4	– 60	– 30·16	452·40
20	10–20	30	15	15	– 3	– 45	– 20·16	302·40
30	20–30	53	23	25	– 2	– 46	– 10·16	233·68
40	30–40	75	22	35	– 1	– 22	– 0·16	3·52
50	40–50	100	25	45	0	0	9·84	246·00
60	50–60	110	10	55	1	10	19·84	190·84
70	60–70	115	5	65	2	10	29·84	149·20
80	70–80	125	10	75	3	30	39·84	398·40
	Total		$N = \Sigma F = 125$			$\Sigma fu = -123$		$\Sigma f\|x-M\| = 1976.44$

Let assumed mean $= A = 45$ and $i = 10$.

$$\therefore \qquad A.M.(M) = A + \frac{\Sigma fu}{N} \times i$$

$$= 45 + \frac{(-123) \times 10}{125} = 45 - 9{\cdot}84 = 35{\cdot}16.$$

$\therefore$ The required mean deviation (M. D.) is given by

$$\delta_m = \frac{\Sigma f|x-M|}{N} = \frac{1976{\cdot}44}{125} = 15{\cdot}8.$$

Example 3. *Find the mean deviation from the arithmetic mean for the following data, which represents the circumstances for the necks of a set of students :*

Mid-Value (in cm) x :	30	31·5	33	34·5	36	37·5	39	40·5
No. of Students f :	4	19	30	63	66	29	18	1

Sol.

x	f	$\xi = x - A$	$f\xi$	x – M	f | x – M |
30	4	– 6·0	– 24·0	– 5·0	20·0
31·5	19	– 4·5	– 85·5	– 3·5	66·5
33	30	– 3·0	– 90·0	– 2·9	60·0
34·5	63	– 1·5	– 94·5	– 0·5	31·5
36	66	0	0·0	1·0	66·0
37·5	29	1·5	43·5	2·5	72·5
39	18	3·0	54·0	4·0	72·0
40·5	1	4·5	4·5	5·4	5·5
	$N = \Sigma f = 230$		$\Sigma f\xi = -1192$		$\Sigma f\|x-M\| = 394$

Scanned with CamScanner