MEASURES OF CENTRAL TENDENCY

DESCRIPTIVE STATISTICAL MEASURES

Graphical representation summarizes information in the data. In addition to the diagrammatic and graphic representations there are numerical methods which summarize the information in data sets, called the measures of central tendency and dispersion. A measure of central tendency is a central value around which the measurements have a tendency to cluster. Measures of dispersion estimate the extent of the spread of the measurements of data sets. These measures put together are called descriptive measures.

Whenever data are summarized, information on individual observations is lost. A measure of central tendency can be computed for a sample or a finite population. The former is called statistic and the later parameter. Sample mean is statistic while population mean is parameter.

MEASURES OF CENTRAL TENDENCY

A measure of central tendency is a typical value that serves as a representative of all the measurements. The measurements obtained from a common source are not likely repetitions. It is undesirable to keep all the measurements in focus. Consequently, a representative of all the measurements is required. Such a representative is one of the three popular measures of central tendency. Viz., arithmetic mean, median and mode. A measure of central tendency is a numerical value around which the measurements have a tendency to cluster.

We come across five measures of central tendency (i) arithmetic mean, (ii) median, (iii) mode, (iv) geometric mean and (v) harmonic mean.

Arithmetic Mean (or Simply Mean):

It is defined as the sum of the given observations divided by the total number of observations.

$$\text{Arithmetic Mean (A.M.)} = \bar{X} = \frac{\text{Sum of all observations}}{\text{Total number of observations}}, \quad \bar{X} = \frac{\sum X}{n}$$

where $\sum X$ = Sum of all observations. (Read $\sum$ as capital Sigma)

n = Total number of observations.

Case A : Raw Data

Let $X_1, X_2, \ldots, X_n$ be 'n' measurements. The arithmetic mean of this data set can be computed by using formula:

$$\bar{X} = \frac{\sum X}{n}, \text{ where } \sum X = X_1 + X_2 + \ldots, + X_n.$$

n = No. of observations in the given data.

Case B: Discrete frequency distribution

Consider the following discrete frequency distribution of variable values and their corresponding frequencies

Variable Value (X)	X_1	X_2	…	X_k	*Total*
Frequency (f)	f_1	f_2	…	f_k	N

The Arithmetic mean is then defined as,

$$\bar{X} = \frac{\sum fX}{N}, \text{ where } \sum fX = f_1X_1 + f_2X_2 + \ldots + f_kX_k$$

N = Total Frequency ($\sum f$)

Case C: Continuous frequency distribution

In this case, A.M. is given by $\bar{X} = \frac{\sum fX}{N}$,

where $\sum fX$ = Sum of products of midvalues of class intervals and the corresponding frequencies.

N = Total frequency. When mid values of class intervals are large in magnitude, the step deviation method (or short cut method) can be employed to find A.M.

Step deviation method : Under this method, A.M. can be calculated, using the following formula :

$$\bar{X} = A + \left(\frac{\sum fX}{N}\right)C$$

where, $d = \frac{X - A}{C}$ = Scaled deviation of X.

X's are mid values of class intervals.

A is a suitable origin parameter, usually chosen as the midpoint of the middle most class interval.

C is the scale parameter whose value is the common width of the class intervals.

$\sum fd$ is the sum of products of deviation values and their corresponding frequencies.

Properties of Arithmetic mean:

(a) Sum of the deviations of observations taken from their A.M. is always zero. Symbolically, we can write this property as

$$\sum (X - \bar{X}) = 0.$$

(b) The sum of the squares of the deviations of observations is minimum, when taken from A.M. Symbolically, we write this as $\sum (X - \bar{X})^2$ which is always minimum.

COMBINED ARITHMETIC MEAN

If $\bar{X}_1$ and $\bar{X}_2$ be arithmetic means of two series of n_1 and n_2 observations respectively, then the arithmetic mean of combined data is as follows.

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$$

The combined arithmetic mean is based on $n_1 + n_2$ observations.This formula can be generalized for many number of groups say k

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 ... + n_k\bar{X}_k}{n_1 + n_2 + ... + n_k}$$

Uses of Arithmetic Mean

It is one of the most commonly and widely used average. In the general usage, we talk about average profits of a business concern, average production of an industry, average benefit of a group of persons etc., we imply the arithmetic mean. Comparison of several means is an important problem in statistical analysis.

Weighted Arithmetic Mean

In calculating arithmetic mean, we suppose that all the items in the distribution have equal importance. But in practice, this may not be so. Some items may be more important than others. For example, rice, wheat, oil, electricity, fuel, pulses etc are more important than coffee, tea, cigarettes, etc. Therefore, proper weight age has to be given to various items in proportion to their relative importance for calculating arithmetic mean. In such cases, weighted arithmetic mean is a suitable measure, for which we assign different weights to different items according to their relative importance.

Suppose the weights assigned to the variable values ($x_1, x_2, \ldots, x_n$) be ($w_1, w_2, \ldots, w_n$) respectively. Then the weighted arithmetic mean ($\bar{X}_w$)is given by

$$\bar{X}_w = \frac{\sum WX}{\sum W}$$

where $\sum WX$ = Sum of the products of variable values and their corresponding weights
$\sum W$ = Sum of weights.

SOLVED PROBLEMS ON ARITHAMETIC MEAN:

Problem: Calculate the mean height of the following 10 measurements
Height (in cms)**:** 120, 115, 140, 141, 125, 124, 127, 130, 130, 133

Solution:

$$\sum X = 1285$$

Number of measurements: n = 10

$$\bar{X} = \frac{\sum X}{n} = \frac{1285}{10} = 128.5$$

The mean height is 128.5 cms

Problem: Compute the arithmetic mean of daily wages of workers in a factory.

Worker :	A	B	C	D	E	F	G	H	I	J	K	L
Daily Wages:(in Rs.)	75	60	90	95	80	75	70	65	65	60	75	70

Solution:

We have $\sum X = 880$
n = 12

$$\bar{X} = \frac{\sum X}{n} = \frac{880}{12} = 73.33$$

The arithmetic mean of daily wages of workers is Rs. 73.33.

Problem: The following data gives the number of children born to 350 women.

No. of children :	0	1	2	3	4	5	6
No. of women :	171	82	50	25	13	7	2

Calculate the mean number of children born per woman.

Solution:

No. of children (x)	No. of women (f)	fx
0	171	0
1	82	82
2	50	100
3	25	75
4	13	52
5	7	35
6	2	12
Total	350	356

From the table, we have,

$\sum fx = 356 \quad N = 350$

$$\bar{X} = \frac{\sum fx}{N} = \frac{356}{350}$$

$\bar{X} = 1.017$

$\therefore$ The mean no. of children born to a woman =1.017

Problem: Four teachers of Engineering reported grades of 80, 90, 50 and 60 respectively for 30, 40, 50 and 60 students, what is the average grade ?

Grade (x)	No. of Students (f)	fx
80	30	2400
90	40	3600
50	50	2500
60	60	3600
Total	180	12100

From table we have $\sum fx = 12100$, N = 180

$$\bar{x} = \frac{\sum fx}{N} = \frac{12100}{180} = 67.22$$

Average Grade = 67.22

Problem: The following data relates to the marks of 100 students in statistics. Calculate the A.M. marks of students.

Marks	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	7	13	20	30	18	12

Solution: **(a) Direct Method :**

Marks	No. of students (f)	Mid value (x)	fx
20-30	7	25	175
30-40	13	35	455
40-50	20	45	900
50-60	30	55	1650
60-70	18	65	1170
70-80	12	75	900
Total	10	-	5250

From table we have, $\sum fx = 5250$, $N = 100$

$$x = \frac{\sum fx}{N} = \frac{1520}{100} = 52.5$$

Alternative Method:

(b) Step Deviation Method :

Let A = 55 and C = Length of class = 10

Marks	No: of students	Mid value x	$d = \frac{X-A}{C}$	fd
20-30	7	25	-3	-21
30-40	13	35	-2	-26
40-50	20	45	-1	-20
50-60	30	55 = A	0	0
60-70	18	65	1	18
70-80	12	75	2	24
Total	100	-	-	-25

From table we have $\sum fd = -25$, C = 10, A = 55, N = 100

$$\bar{X} = A + \left(\frac{\sum fd}{N}\right)C$$

$$= 55 + \left(\frac{-25}{100}\right)10 = 55 - 2.5 = 52.5$$

Average Marks of Students = 52.5

Problem: compute arithmetic mean for the 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

Solution: Given, C =Common length of class internals = 10

Class	Frequency (f)	Mid value X	$d=\dfrac{X-A}{C}$	fd
50-59	1	54.5	-3	-3
60-69	3	64.5	-2	-6
70-79	8	74.5	-1	-8
80-89	17	84.5 = A	0	0
90-99	35	94.5	1	35
100-109	4	104.5	2	8
110-119	2	114.5	3	6
Total	70	-	-	32

From the table, we have, $\sum fd = 32$, $C = 10$, $A = 84.5$, $N = 70$

$$\bar{X} = A + \left(\frac{\sum \text{fd}}{N}\right)C = 84.5 + \left(\frac{32}{70}\right)10$$

$$= 84.5 + 4.5714$$

$\therefore$ Arithmetic Mean = $\bar{X}$ = 89.0714

Problem : The mean wage of workers in a factory running two shifts of 60 and 40 workers are Rs.40 and Rs.35 respectively. Find the mean wages of all the 100 workers put together.

Solution: Given that $n_1 = 60$ and $n_2 = 40$

$\bar{X}_1 = 40$ and $\bar{X}_2 = 35$

Combined A. M. = $\bar{X} = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$

$$= \frac{(60)(40)+(40)(35)}{60+40}$$

$$= \frac{2400+1400}{100} = \frac{3800}{100} = 38$$

$\therefore$ Mean wage of all the 100 workers = Rs. 38.

Problem: Calculate the weighted A.M. of the index numbers for the following data:

Group	Index No.	Weight
Food	126	9
Clothing	130	5
Fuel and light	140	6
House Rent	175	2
Miscellaneous	182	3

Solution:

	Index No. (x)	Weight (w)	Wx
Food	126	9	1134
Clothing	130	5	650
Fuel and light	140	6	840
House Rent	175	2	350
Miscellaneous	182	3	546
Total	-	25	3520

Weighted A.M = $\bar{X}_w = \frac{\sum wx}{\sum w} = \frac{3520}{25} = 140.8$

Weighted A.M. of index numbers is 140.8

Problem: Calculate the average salary paid in the whole industry, using the data given below:

Income Group (Rs.)	60-80	80-100	100-200	200-300	300-600
No. of Firms	16	13	12	10	14
Average No. of workers	4	7	5.25	2.2	1.5

Solution : We are given frequency distribution with weights. We calculate weighted arithmetic mean.

Class	No. of firms (y)	Average No. of workers (z)	Mid value of class (x)	Total No. of workers (weight) $w = yz$	wx
60-80	16	4	70	64	4480
80-100	13	7	90	91	8190
100-200	12	5.25	150	63	9450
200-300	10	2.2	250	22	5500
300-600	14	1.5	450	21	9450
Total	-	-	-	261	37070

$$\text{Weighted Mean } = \frac{\sum wx}{\sum w} = \frac{37070}{261} = 142.0307$$

Average salary paid to the workers in the industry is Rs. 142.03

Problem: Given the following cumulative frequency distribution of less than upper bound marks (x), obtained by 140 students in an examination, find the mean marks of students.

X	10	20	30	40	50	60	70	80	90	100
Cumulative Frequency	140	133	118	100	75	45	25	9	2	0

Solution : Recover the frequency distribution from the given cumulative frequency distribution constructing the following table

Lower boundary	More than cum. *f*	Marks Class	No. of students Frequency (*f*)	Mid value *x*	*fx*
10	140	10-20	7	15	105
20	133	20-30	15	25	375
30	118	30-40	18	35	630
40	100	40-50	25	45	1125
50	75	50-60	30	55	1650
60	45	60-70	20	65	1300
70	25	70-80	16	75	1200
80	9	80-90	7	85	595
90	2	90-100	2	95	190
100	0				
Total	-	-	140	-	7170

From the table, we have

$\sum fx = 7170$ N = 140

$$\bar{X} = \frac{7170}{140} = 51.2143$$

Arithmetic mean marks of students = 51.2143

Problem: Find the missing frequency from the following data, given the average mark is 16.82

Marks	Frequency
0-5	10
5-10	12
10-15	16
15-20	f_4
20-25	14
25-30	10
30-35	8

Solution: Computation of Arithmetic mean :

Marks	Midvalue X	Frequency (f)	$d = \frac{x-17.5}{5}$	fd
0-5	2.5	10	-3	-30
5-10	7.5	12	-2	-24
10-15	12.5	16	-1	-16
15-20	17.5 = A	f_4	0	0
20-25	22.5	14	1	14
25-30	27.5	10	2	20
30-35	32.5	8	3	4
Total	-	N=70 + f_4	-	$\sum fd$=-12

From the table, we have, N = 70 + f_4, $\sum fd$ = -12 , C=5

We know that, $\bar{X}$ -A+$\left[\frac{\sum fd}{N}\right]C$

$$16.82 = 17.5 + \left[\frac{(-12)}{70+f_4}\right]5$$

$$16.82 - 17.5 = \frac{-60}{70+f_4} \Rightarrow -0.68 = \frac{-60}{70+f_4}$$

$\Rightarrow$ 0.68 (70+f_4) = 60

$\Rightarrow$47.6 + 0.68 f_4 = 60

$\Rightarrow$0.68 f_4 = 60 – 47.6 =12.4

$$\therefore f_4 = \frac{12.4}{0.68} = 18.2353 \simeq 18$$

Problem: The mean wage of 100 workers in a factory who work in two shifts of 60 and 40 workers respectively is Rs. 38. The mean wage of 60 laborers working in the morning shift is Rs, 40. Find the mean wage of laborers working in the evening shift.

Solution : Given $n_1 = 60$, $n_2 = 40$

$$\bar{X}_1 = 40 \text{ and } \bar{X} = 38$$

we have, $\bar{X} = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$

$$38 = \frac{(60)(40) + (40)(\bar{X}_2)}{60 + 40}$$

$3800 = 2400 + \bar{X}_2$

$40\,\bar{X}_2 = 3800 - 2400 = 1400$

$$\bar{X}_2 = \frac{1400}{40} = 35$$

Mean wage of 40 workers working in the evening shift is Rs. 35.

Problem: The mean monthly salary paid to all employees in a certain company was Rs. 600/-, the mean monthly salaries paid to male and female employees were Rs. 620 and 520 respectively. Obtain the percentage of male to female employees in the company.

Solution : Mean salary paid to male and female employees $\bar{X} = 600$

Mean salary paid to male employees $\bar{X}_1 = 620$

Mean salary paid to female employees = $\bar{X}_2 = 520$

We have, $\bar{X} = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} = 600 = \dfrac{620n_1 + 520n_2}{n_1 + n_2}$

$600\,(n_1 + n_2) = 620\,n_1 + 520\,n_2$

$600\,n_1 + 600\,n_2 = 620\,n_1 + 520\,n_2$

$600\,n_2 + 520\,n_2 = 620\,n_1 - 600\,n_1 \Rightarrow 80n_2 = 20n_1$

$$\frac{n_2}{n_1} = \frac{20}{80} = \frac{1}{4}$$

$\Leftrightarrow n_1 : n_2 = 4:1$

Percentage of male employees = $\left(\dfrac{4}{5}\right)100 = 80\%$

Percentage of Female Employees = $\left(\dfrac{1}{5}\right)100 = 20\%$

Problem: For a certain frequency table, which has been partly reproduced here, the mean was found to be 1.46.

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

Find the missing frequencies.

Solution : Let f_1 and f_2 be missing frequencies, then

No. of accidents	*Frequency*	*fx*
0	46	0
1	f_1	f_1
2	f_2	$2f_2$
3	25	75
4	10	40
5	5	25
Total	$86+f_1+f_2$	$140+f_1+2f_2$

From table we have, $\sum f = N = (86 + f_1 + f_2)$

$\sum fx = (140 + f_1 + 2f_2)$

But given that

$\sum f = N = 200$

Also $\bar{X} = \dfrac{\sum fx}{N} = 1.46$

$\therefore \sum fx = N\bar{X} = (200)(1.46) = 292$

$\therefore$ We have $86 + f_1 + f_2 = 200$ (1)

and $140 + f_1 + 2f_2 = 292$ (2)

From (1) : $f_1 + f_2 = 200 - 86 = 114$ (3)

From (2) : $f_1 + 2f_2 = 292 - 140 = 152$ (4)

Solving (3) and (4), we get f_1 and f_2.

Consider (3) : $f_1 + f_2 = 114$

(4) : $f_1 + 2f_2 = 152$

By subtraction $-f_2 = -38$ or $f_2 = 38$

By substituting $f_2 = 38$ in (3)

we get, $f_1 + 38 = 114$

$\therefore f_1 = 114 - 38 = 76$

Hence, the missing frequencies are $f_1 = 76$ and $f_2 = 38$

Problem: The mean salary paid to 1,000 employees of an establishment was found to be Rs. 180.40. Later on, after disbursement of salary, it was discovered that the salary of two employees were wrongly entered as Rs. 297 and Rs. 165. Their correct salaries were Rs. 197 and 185. Find the true arithmetic mean.

Solution: Given that $\bar{X}$ = 180.40 and n = 1000

Since $\bar{X} = \dfrac{\sum X}{n}$ $or \sum X = n\bar{X} = (1000)(180.40) = 180400$

Since, two values were wrongly entered, consider the correction as :

Corrected $\sum$X = Wrong $\sum$X – (Sum of wrong values) + (Sum of correct values)

= 180400 – (297 + 165) + (167 + 185)
= 180400 – 462 + 382
= 180320

Corrected Mean = $\dfrac{\text{Corrected } \sum X}{n} = \dfrac{180320}{1000} = 180.32$

True mean is Rs.180.32.

Problem: The mean weight of a student in a group of six students is 119 pounds. The individual weights of five of them are 115, 109, 129, 117 and 114 pounds. What is the weight of the sixth student.

Solution : Given that $\bar{X}$ = 119 and n = 6

Since $\bar{X} = \dfrac{\sum X}{n}$ $or \sum X = n\bar{X}$, we get, $\sum X$ = (6) (119) = 714

$\sum X$ = 714 = Sum of six observations

The sum of given five observations = 115 + 109 + 129 + 117 + 114 = 584

The value of sixth observation = Sum of 6 observations – Sum of five observations.

= 714 – 584 = 130

The weight of sixth student = 130 pounds

Median

Median of data is the value of the variable which divides arranged data into two equal parts. When the observations in the data are arranged either in ascending (increasing) order or in descending (decreasing) order, the Median is defined as the value of the variable, which divides the arranged data into two equal parts. We denote it by $\tilde{X}$

Case A : Raw data : When the number of observations in the data is an 'odd' number, then the Median is equal to the middle most term in the arranged data. If the number of observation is even, the simple average of the two middle most values of the arranged data is median.

Case B :
Discrete frequency distribution : In this case, median is defined as the value of the variable that corresponds the less than cumulative frequency just above half the total frequency.

Case C :
Continuous frequency distribution : In the case of continuous frequency distribution, median can be obtained by using the following formula :

$$\text{Median} = \tilde{X} = L + \left(\frac{\frac{N}{2} - m}{f} \right) C$$

where, L = Lower boundary of the median class
f = Frequency of the median class
m = Less than cumulative frequency of the class that precedes the median class
C = Length of the median class
N = Total Frequency

Here, Median class is the class that corresponds the less than cumulative frequency just greater than (N/2) value.

Remark : When, we are given the class intervals in descending order, the median can be found by using the following formula.

$$\text{Median} = \widetilde{X} = U - \left(\frac{\frac{N}{2} - m}{f} \right) C$$

where, 'U' is the upper bound of the median class. Other symbols such as *N*, *m*, *f* and *C* remain the same as in the previous formula.

USES OF MEDIAN

It is used frequently in practice. It is readily used in cases where quantitative measurement of all items is difficult but ordering of items is relatively easy. When the data contains extreme values, it is advantageous to use median. Its usefulness as a positional average is recognised in statistical analysis.

Remarks :

1. For a continuous frequency distribution Median can be found by plotting gives.

2. The sum of the absolute deviations of observations taken from the median is always minimum, i.e., $\sum |X - \widetilde{X}| \leq \sum |X - A|$, where A is any arbitrary value.

SOLVED PROBLEMS ON MEDIAN

Problem 2.10.1: Determine Median for the following data : 26, 20, 15, 45, 18, 8, 10, 38, 13.

Solution : Arrange the observations in ascending order as
8, 10, 13, 15, 18, 20, 26, 38, 45
Number of observations = 9 (odd number)
Value of the middle term = Median = 18.

Problem 2.10.2: Find the median for the following values :
16, 12, 5, 8, 9, 5, 10, 28.

Solution : Arrange the observations in ascending order as,
5, 5, 8, 9, 10, 12, 16, 28.
Number of observations = 8 (even number)
Median = Average of values of two middle terms

$$\text{Median} = \frac{9+10}{2} = 9.5$$

Problem 2.10.3: Calculate the median for the following frequency distribution.

Variable X	3	4	5	6	7	8
Frequency (f)	2	4	7	9	10	16

Solution :

X	f	Less than cum. f
3	2	2
4	4	6
5	7	13
6	9	22
7 Median	10	32 Median
8	16	48
Total	N = 48	-

N/2 = 24

From the above table, $\frac{N}{2} = \frac{48}{2} = 24$

Here, '32' is just greater than $\left(\frac{N}{2}\right)$ value.

$\therefore$**Median** = The value of the variable that corresponds less than cumulative frequency just greater than N/2 value = 7.

Problem: The weight of 30 students are given in the following frequency distribution. Determine the median weight of the students.

Weight in kgs	60-64	65-69	70-74	75-79	80-84
No. of students	2	8	12	5	3

Solution :

Class Boundaries	F	Less than cum. f
59.5-64.5	2	2
64.5-69.5	8	10 = m
69.5-74.5 Median class	12 =f	22
74.5-79.5	5	27
79.5-84.5	3	30
Total	N = 30	-

N/2 = 15

From table, $\frac{N}{2} = \frac{30}{2} = 15$ Here, 69.5 – 74.5 is the median class.

We have, = 69.5; m = 10; f = 12; and C = 5

$$Median = U + \left(\frac{\frac{N}{2} - m}{f}\right)C$$

$$= 69.5 + \left[\frac{15-10}{12}\right]5 \quad = 69.5 + \frac{25}{12} \quad = 69.5 + 2.08 \quad = 71.58$$

Median weight = 71.58 kg.

Problem: Find the median weight for the following table:

Weight in kgs	84-80	79-75	74-70	69-65	64-60
No. of students	3	5	12	8	2

Solution: The class intervals are given in the descending order:

Class Boundaries	f	Less than cum. f
84.5-79.5	3	3
79.5-74.5	5	8 = m
74.5-69.5 Median class	12 = f	20
69.5-64.5	8	28
64.5-59.5	2	30
Total	30	-

Here 74.5 – 69.5 is the median class.

From table , $\frac{N}{2} = \frac{30}{2} = 15$

U = 74.5; m = 8, f = 12, and C = 15

Median = $\bar{X}$ = U - $\left(\frac{\frac{N}{2} - m}{f}\right)C$

$$= 74.5 - \left[\frac{15-8}{12}\right]5 = 74.5 - \frac{35}{12}$$

$$= 74.5 - 2.92 = 71.58$$

Median weight = 71.58 kg.

Problem: In the frequency distribution of 170 families given below, the number of families corresponding to expenditure groups 20-30 and 40-50 are missing in the table. However, the median is known to be 35. Find the missing frequencies.

Expenditure	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of families	10	20	?	40	?	25	15

Solution : Let the missing frequencies of the expenditure groups 20-30 and 40-50 be f_1 and f_2 respectively.

Given , Median = 35 and total frequency = 170

Class	f	Less than cum. f
0-10	10	10
10-20	20	30
20-30	f_1	$30 + f_1 = m$
30-40 Median class	$40 = f$	$70 + f_1$
40-50	f_2	$70 + f_1 + f_2$
50-60	25	$95 + f_1 + f_2$
60-70	15	$110 + f_1 + f_2$
Total	170	-

Since, median = 35, the median class is given by 30-40.

From the table, we have

L = 30, f = 40, m = 30 + f_1

$$\frac{N}{2} = \frac{170}{2} = 85 \; and \; C = 10$$

$$\therefore \text{Median} = L + \left(\frac{\frac{N}{2} - m}{f} \right) C$$

$$\Rightarrow \; 35 = 30 + \left[\frac{85 - (30 + f_1)}{40} \right] 10$$

$$\Rightarrow 35 - 30 = \frac{55 - f_1}{4}$$

$$\Rightarrow 55 - f_1 = 20 \; or \; f_1 = 55 - 20 = 35$$

Since, the total frequency N = 170, or $110 + f_1 + f_2 = 170$, and $f_1 = 35$, we get

$110 + 35 + f_2 = 170$ or $f_2 = 170 - 145 = 25$

$\therefore$ The missing frequencies are respectively $f_1 = 35$ and $f_2 = 25$.

MODE

Mode is the value of the variable, which occurs most frequently in the measurements. The word 'Mode' is derived from the French word "La Mode' which signifies fashion. It is often used as a positional average in practice. Mode need not be unique. A data set may have two or more modes. If a data set contains two modes then the data is said to be bi-model data. Mode is usually denoted by X.

Case A :
Raw data : In this case, mode is defined as the value of the variable which occurs most frequently in the given data set.

Case B :
Discrete frequency distribution : In the case of discrete frequency distribution, mode is the value of the variable for which the frequency is maximum.

Case C :
Continuous frequency distribution : In the case of continuous frequency distribution, mode can be obtained by using the following formula.

$$\text{Mode} = \widehat{X} = \text{L} + \left(\frac{f - f_1}{2f - f_1 - f_2}\right)C$$

where L : Lower bound of the model class
f : Frequency of the model class
f_1 : Frequency of the class that precedes the model class
f_2 : Frequency of the class that succeeds model class
C : Length of the model class

Here, model class is the class that corresponds largest frequency.

Remarks :

In the following three cases, mode can not be obtained by using the above formula:

(a) When the highest frequency is observed at the beginning of the frequency table.

(b) When the highest frequency is observed at the ending of the frequency table.

(c) When two or more class intervals contain the same maximum frequencies.

However, in the above three cases, mode can be obtained by using either a method called 'Grouping method' or 'empirical relationship between arithmetic mean, median and mode.

Empirical Relationship between A.M. Median and Mode

If mean, median and mode are equal for a frequency distribution, then the distribution is called symmetrical distribution. In a moderately asymmetrical distribution, mean, median and mode approximately satisfy the following empirical relationship:

(A.M. - Mode) = 3 (A.M. - Median)

Symbolically $(\bar{X} - \widehat{X}) = 3\,(\bar{X} - \widetilde{X})$

This relation is observed from experience and it is not mathematically derived. When we are given the values of any two of these three measures, then the third measure can be found from this relation. This relation is occasionally used to find mode, when A.M. and Median are known.

Mode = 3 Median – 2 A.M.

Uses of Mode :

Mode finds an important place in marketing studies, where a manager of a business concern is interested in knowing about the size which has the highest concentration of items. For example, in placing an order for shoes or ready made garments, the model size helps because this size and other sizes around it are in common demand. It is also used in dealing with non-quantitative data.

SOLVED PROBLEMS ON MODE

Problem: Find the mode for the following set of values of a variable :

2, 7, 3, 2, 1, 3, 2, 2, 5

Solution : In the given data the value 2 occurs most frequently than the other values. Hence,

mode of the given data is 2.

Problem: Determine the mode for the following data

1, 0, 2, 3, 6, 7, 5, 4, 8.

Solution : In the given data, no single value repeats more frequently, when compared to other

values. Therefore, we conclude that there is no mode for the above data.

Problem: Find the modal age of married women at first child birth :

Age at the birth of first child (in years)	13	14	15	16	17	18	19	20	21	22	23
No. of married women	37	150	300	360	270	435	160	200	85	65	25

Solution : From the given discrete frequency distribution, it is observed that the highest frequency is 435. The age of married women corresponding to this highest frequency is 18 years.

Hence, 18 years is the model age of married women at first child birth.

Problem Calculate the mode for the following data :

Class	130-134	135-139	140-144	145-149	150-154	155-159	160-164
Frequency	5	15	28	24	17	10	1

Solution :

Class	Frequency
130-134	5
135-139	15 = f_1
140-144 Model class	28 = f
145-149	24 = f_2
150-154	17
155-159	10
160-164	1
Total	100

The highest frequency corresponds the class 140 –144. Therefore, it is the model class.

From table, we have,

L = 139.5, f = 28, f_1 = 15, f_2 = 24 and C = 5

$$\text{Mode} = X = L + \left[\frac{f - f_1}{2f - f_1 - f_2}\right]C$$

$$= 139.5 + \left[\frac{28 - 15}{2x28 - 15 - 24}\right]5$$

$$=139.5+\frac{65}{17}=139.5+3.8235$$
$$=143.3235$$

Problem: In a moderately skewed distribution (Asymmetrical distribution) A.M. = 15 and Mode = 12. Find the value of the Median of the given distribution.

Solution : Consider the empirical relationship between mean, median and mode.

(Mean - Mode) = 3 (Mean - Median)

Hence, 15 - 12 = 3 (15 - Median)
Therefore, 3 Median = 45 – 3 = 42

Median = $\frac{42}{3}=14$

Problem: Find the Mode for the following data :
2, 5, 3, 2, 1, 4, 6, 3, 7

Solution : Since 2 and 3 have maximum frequencies, (2 occurs 2 tm ice 3 occurs 2 tm ice) the given data is a bimodal data. Therefore, we use the empirical relationship between mean, median and mode.

$$A.M=\overline{x}=\frac{\sum x}{n}=\frac{33}{9}=3.6667$$

To find median arrange the data in ascending order:
1, 2, 2, 3, 3, 4, 5, 6, 7
Median = 3
Consider the empirical relationship (AM. - Mode) = 3 (Mean - Median)
Substituting the values of A.M. and Median. we get.

3.6667 – Mode = 3 (3.667-3)
3.6667 – Mode = 2.0001
$\therefore$ Mode = 3.667 – 2.0001 = 1.6666

Problem: Compute Mode for the following data :

Size	0-4	4-8	8-12	12-16	16-20
Frequency	10	20	30	35	35

Solution: The highest frequency is 35 and it corresponds the two bottom most class intervals of frequency table. Hence, the given distributions is a bimodal. In this case, we use the empirical relationship between A.M., Median and Mode.

Class	Frequency f	Mid values x	fx	Less than cumulative frequency
0-4	10	2	20	10
4-8	20	6	120	30
8-12	30	10	300	60 =m
12-16	35 = f	14	490	95
16-20	35	18	630	130
Total	130	-	1560	-

Form the table, $\sum fx$ = 1560; N = 130

$\therefore$ A.M. = $\bar{x} = \frac{\sum fx}{N} = \frac{1560}{130} = 12$

Since, N/2 = 130/2 = 65

12-16 is the median class

L=12, m = 60, f = 35, N = 130, C = 4

$$\text{Median} = L + \left(\frac{\frac{N}{2} - m}{f}\right)C$$

$$= 12 + \left[\frac{65-60}{35}\right]4 = 12 + \frac{20}{35} = 12 + 0.5714 = 12.5714$$

Median = 12.5714

Consider the empirical relationship

(A.M. - Mode) = 3 (Mean - Median)

12 – Mode = 3 (12 – 12.5714)

12 – Mode = 3 (- 0.5714)

12 – Mode = - 1.7142

$\therefore$ Mode = 12 + 1.7142

Mode = 13.7142

Problem: The median and mode of the following wage distribution are known to be Rs. 335 and Rs. 340 respectively. Find the three missing frequencies in the frequency distribution given below.

Wages (in Rs.)	Frequency

0-100	4
100-200	16
200-300	60
300-400	?
400-500	?
500-600	?
600-700	4
Total	230

Solution : Let f_1, f_2 and f_3 be the missing frequencies.

Wages (Rs.)	Frequency	Less than cumulative frequence
0-100	4	4
100-200	16	20
200-300	60	80
300-400 Median class	f_1	$80 + f_1$
400-500	f_2	$80 + f_1 + f_2$
500-600	f_3	$80 + f_1 + f_2 + f_3$
600-700	4	$84 + f_1 + f_2 + f_3$
Total	230	

We have, $230 = 84 + f_1 + f_2 + f_3$

$f_3 = 230 - 84 - f_1 - f_2$

$= 146 - (f_1 + f_2)$……….(1)

Since the median and mode are 355 and 340, both lie in the same class, viz., 300 – 400.

$$\text{Median} = 335 = 300 + \left[\frac{\frac{230}{2} - 80}{f_1}\right]100$$

$$335\text{-}300 = \frac{(115-80)100}{f_1} \Rightarrow 35 f_1 = 3500$$

$$\therefore f_1 = \frac{3500}{35} = 100$$

$$\text{Mode} = L + \left[\frac{100-60}{2\times100-60-f_2}\right]100$$

$$40 = \frac{4,000}{140-f_2} \Rightarrow 40(140-f_2) = 4000$$

$$\Rightarrow 5600 - 40 f_2 = 4000$$

$$40 f_2 = 5600 - 4000 = 1600$$

$$\therefore f_2 = \frac{1600}{40} = 40$$

From (1), we get $f_3 = 146 - (100 + 46) = 6$

Hence, the missing frequencies are 100, 40 and 6 respectively

Requisites for an Ideal Measure of Central Tendency :

According to Prof. Yule, the following are the chief characteristics to be satisfied by an ideal measure of central tendency.

i) It should be rigidly defined

ii) It should be easy to understand and easy to calculate

iii) It should be based on all the observations.

iv) It should be suitable for further mathematical manipulations

v) It should be affected as little as possible by sampling fluctuations.

vi) It should not be affected much by extreme values.

Now, we shall consider the merits and demerits (or advantages and disadvantages) of Arithmetic mean, median and mode.

Measure of Central Tendency	Merits	Demerits :
a) Arithmetic Mean	1) It is rigidly defined. 2) It is easy to understand and easy to calculate. 3) It is based on all the observations. 4) It is suitable for further mathematical manipulations. 5) Of all the averages, it is affected least by sampling fluctuations.	1) It is affected very much by extreme values 2) It cannot be found for distributions with open end classes. 3) It cannot be used if measurements are qualitative.
b) Median	1) It is rigidly defined. 2) It is easy understand and easy to calculate. 3) It is not affected by extreme values. 4) It can be found for distributions with open end classes. 5) Median is the only average to be used while dealing with the qualitative data.	1) It is not based on all the observations. 2) It is not suitable for further mathematical manipulations. 3) It is affected by sampling fluctuations. 4) It is not so generally familiar as the arithmetic mean.
c) Mode	1) It is easy to understand and easy to calculate. 2) It is not affected by extreme classes. 3) It can be found for distribution with open end classes.	1) Mode is ill-defined 2) It is not based on all the observations. 3) It is not suitable for further mathematical manipulations.

.