Measures of central tendency
or
Statistical averages

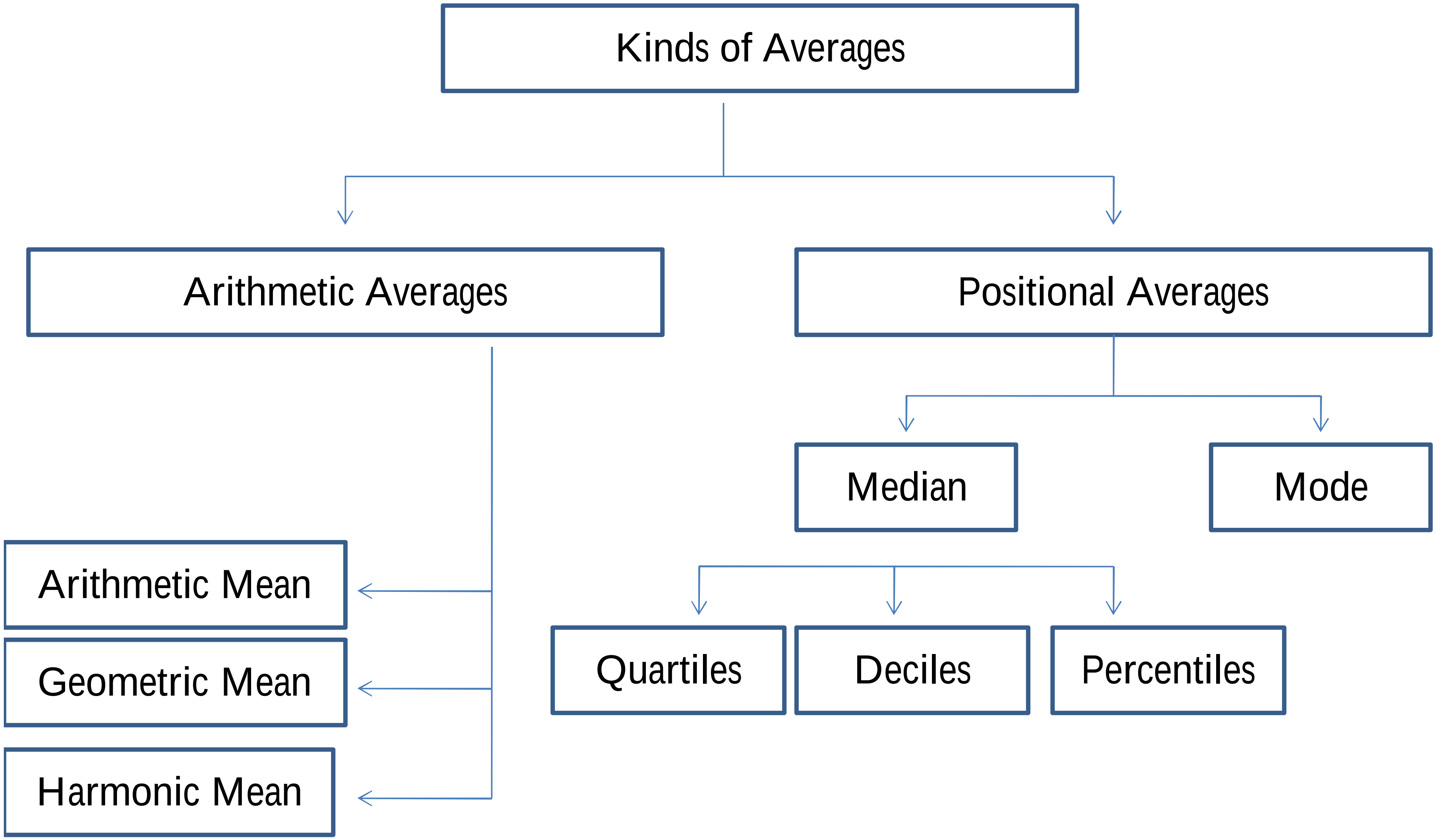
Statistical Averages :

“ A measure of central tendency is a typical value around which other figures congregate or which divides their number in half. Thus an average can be used to describe or represent a whole series of figures involves magnitudes of the same variable”.

“Averages are derived figures but not the original data”.

Definitions of Central Tendency :

Central Tendency is a statistical measure that identifies a single score as representative of the entire distribution. The goal of central tendency is to find the single score that is most typical or most representative of the entire group”.


Kinds of Averages
Arithmetic Averages
Positional Averages
Median
Mode
Arithmetic Mean
Geometric Mean
Harmonic Mean
Quartiles
Deciles
Percentiles

Characteristics (Essentials or Desirables) of a Good Average:

An average should be

(a) Rigorously defined.

(b) Easy to compute.

(c) Capable of simple interpretation.

(d) Dependent on all values of observation.

(e) Not unduly affected by one or two extremely large or small values.

(f) Should fluctuate relatively little from one random sample to another.

(g) Be capable of mathematical manipulation.

Arithmetic Mean of Ungrouped Data :

$$\text{Arithmetic Mean} = \frac{\text{(Sum of Values)}}{\text{(Number of Values)}}$$

$\Rightarrow$ Arithmetic Mean $\times$ (Number of Values) = (Sum of Values)

$\Rightarrow$ Arithmetic Mean If each value in a distribution of any series is replaced by the A.M. are added, the result = Sum of the values

If $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ are 'n' values of a variable, then their A.M. is (generally denoted by $\overline{x}$) and is given by,

$$\overline{x} = \frac{(x_1 + x_2 + x_3 + \ldots\ldots\ldots + x_n)}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

If f_1, f_2, f_3,………,f_n are respective frequencies of x_1, x_2, x_3,……..,x_n, then

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots\ldots\ldots + f_n x_n}{f_1 + f_2 + \ldots\ldots\ldots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Ex:

Find arithmetic mean of the values 5, 7, 1, 3, 4, 10.

Sol:

Here n=6, therefore

$$\bar{x} = \frac{5 + 7 + 1 + 3 + 4 + 10}{6} = \frac{30}{6} = 5$$

Ex:

Find arithmetic mean of the values 5, 3, 3, 5, 5, 7, 5, 1, 3, 4, 4, 10.

Sol:

The above data can be classified as,

Value	Frequency
1	1
3	3
4	2
5	4
7	1
10	1

$$\bar{x} = \frac{4\times 5 + 1\times 7 + 1\times 1 + 3\times 3 + 2\times 4 + 1\times 10}{4+1+1+3+2+1} = \frac{20+7+1+9+8+10}{12} = \frac{55}{12} = 4.58$$

Combined A.M. of Two Groups :

Combined A.M. of Two Groups G_1 and G_2 such that :

	Group I (G_1)	Group II (G_2)
A.M.	$\bar{x}_1$	$\bar{x}_2$
Number of values for observation	n_1	n_2

Combined A.M. of these two groups is given by :

$$\overline{x_{12}} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2}}{n_1 + n_2} \qquad \text{.........................(1)}$$

If there are p' groups their combined A.M. is given by :

$$\overline{x_p} = \frac{n_1 \overline{x_1} + n_2 \overline{x_2} + \dots\dots\dots n_p \overline{x_p}}{n_1 + n_2 + \dots\dots\dots + n_p}$$

If $n_1=n_2$ in ....(1) then,

$$\overline{x_{12}} = \frac{n_1\overline{x_1} + n_1\overline{x_2}}{n_1 + n_1} = \frac{n_1(\overline{x_1} + \overline{x_2})}{2n_1}$$

$$\overline{x_{12}} = \frac{\overline{x_1} + \overline{x_2}}{2}$$

Members; 4
$n_1 = 4$

Average Sales; ` 10 million
X_1 = ` 10 million

Members; 5
$n_2 = 5$

Average Sales; ` 13 million
$X_2 =$ ` 13 million

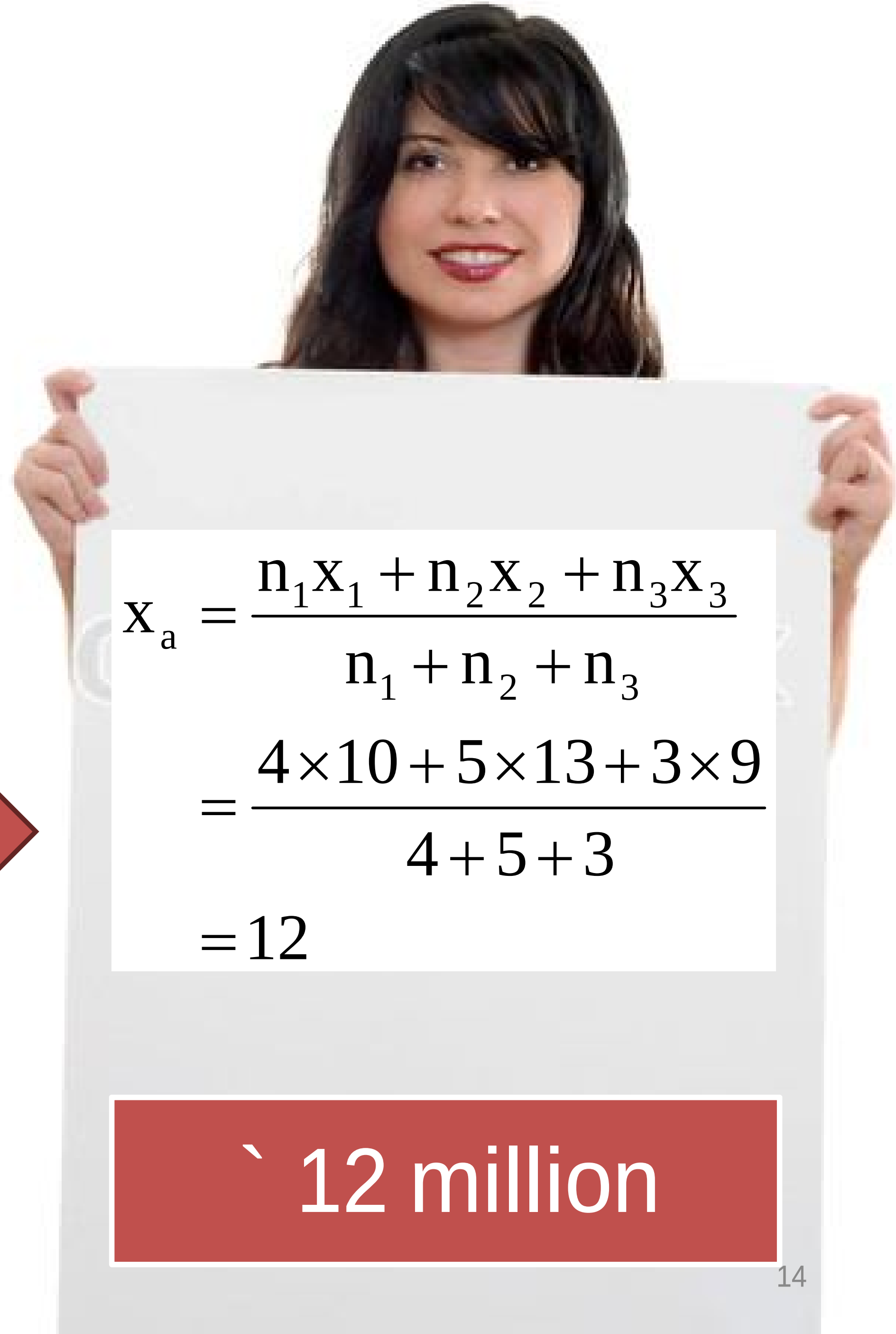
Members; 3

$n_3 = 3$

Average Sales; ` 9 million

X_3 = ` 9 million

Average sales per Team for company can be given by;



$$x_a = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3}{n_1 + n_2 + n_3}$$

$$= \frac{4 \times 10 + 5 \times 13 + 3 \times 9}{4 + 5 + 3}$$

$$= 12$$

` 12 million

Weighted Arithmetic Mean :

Suppose a man purchases mangoes @ rs 20 per kg, apples @ rs 30 per kg and oranges @ rs 16 per kg, then the average rate of purchase will be :

Rs (20 + 30 + 16)/3 = Rs 22 per kg

But what if man purchases 5 kgs of mangoes, 2 kgs of apples and 6 kgs of oranges.
For these cases the following formula is used,

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w} \Rightarrow \bar{x} = \frac{5\times 20 + 2\times 30 + 6\times 16}{5+2+6} = \frac{256}{13} = Rs19.69$$

Where *w'* is known as the weight given to the particular item.

Properties of A.M.:

(a) If a constant amount is added, subtracted, multiplied or divided from each value in the series, mean is also added, subtracted, multiplied or divided by the same constant amount.

For Instance:

Following are the salaries of employees in a firm of a particular department 30 k, 50 k, 90 k, 15 k, 16 k. Manager HR decides to give an incentive of Rs 2 k to each employee as a reward of the best performing department during the financial year.

The manager wants to calculate the increment in average salary which was Rs 40.2 k . What should he do?

Solution:

The manager HR should also add the average which was 40.2 k by 2 k. This way the new average results in 42.2 k.

(b) The algebraic sum of the deviations of all values from their mean = 0.

For Instance:

The mean of the values 4, 6, 9, 15, 16 is 10, the deviations of the values from mean can be calculated as:

4 – 10	- 6
6 – 10	- 4
9 – 10	-1
15 – 10	5
16 – 10	6
Total	0

(c) In case of frequency distribution, the algebraic sum of the deviations of all values from their mean

$$= \sum f(x-\bar{x}) = \sum fx - \bar{x}\sum f = \bar{x}\sum f - \bar{x}\sum f = 0$$

$$\text{because } \bar{x} = \frac{\sum fx}{\sum f} \Rightarrow \bar{x}\sum f - \sum fx$$

(d) If the mean and number of values are given but one value is missing in the series,

Missing value = (Number of values × Mean) – (Sum of known values)

(e) If one or two values have been taken wrongly while calculating the mean then,

$$\text{Correct Mean} = \frac{1}{n}\left[\begin{array}{l}\text{Sum of all values including wrong value – The value (or values taken} \\ \text{wrongly) + Correct value (or values)}\end{array}\right]$$

A.M of Grouped Data:

In Grouped data the middle value of each group is representative of the group i.e. in grouped data, each item in the group is considered to be equal to the middle value of that group.

Therefore the Sum of values in each group = Middle value of the group $\times$ Number of values.

Worksheet :

1. Find A.M. of 3, 6, 24, 48. (20.25)

2. Calculate A.M. of the following observations: (39.77)
   32. 35. 36. 37. 39. 41. 43. 47. 48.

3. Find the A.M. of following data:

x	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**
f	5	9	13	21	20	15	8	3

Ans : 22.127

4. Calculate A.M. of following frequency distribution :

Class Interval	**10-20**	**20-40**	**40-70**	**70-120**	**120-200**
Frequency	4	10	26	8	2

Ans : 57.4

5. The following figure relates to the monthly output of the cloth of a factory in given year;

Months	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
Output In ‘000 meters	80	88	92	84	96	92	96	100	92	94	98	86

Calculate the average monthly output.

Ans; 91.5 (‘000 meters)

Ex : From the data below, find the missing frequency if the Arithmetic Mean is 33.

Loss Per Shop	**No. of Shops**
0 – 10	**10**
10 – 20	**15**
20 – 30	**30**
30 – 40	**-**
40 – 50	**25**
50 – 60	**20**

Solution : Let the missing frequency be f.

Loss per Shop	**Mid Value x**	**No. of Shops**	fx
0 – 10	5	**10**	50
10 – 20	15	**15**	225
20 – 30	25	**30**	750
30 – 40	35	f	$35f$
40 – 50	45	**25**	1125
50 – 60	55	**20**	1100
		$\Sigma f = 100+f$	$\Sigma fx = 3250+35f$

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{3250 + 35f}{100 + f} = 33$$

$$\Rightarrow \quad 3250 + 35f = 3300 + 33f$$

$$\Rightarrow \quad f = 25$$

Short Cut Method of Arithmetic Mean :

This method is used when we deal with large figures.

Steps :

a. Take any value (though not necessary but the value among the given is taken). Call this value as the assumed mean, called as ‘A’.

b. Calculate the deviation of all the values from this assumed mean ‘A’ and take the mean of these deviations, i.e. calculate $\Sigma(x - A)/n$ or $\Sigma f(x - A)/\Sigma f$.

c. Add the mean of the deviation to the assumed mean, i.e.

Required A.M. of the given series = $A + \dfrac{\sum f(x - A)}{\sum f}$

$$= A + \frac{\sum fd}{\sum f}$$

Where $d = (x - A)$ are the deviations.

d. In case any common factor in the deviation, then divide each deviation by that common factor (in continuous series the common factor in most of the cases = the class interval), i.e. find *d/i = d'*, where *'i'* is the common factor.

e. Required mean = $A + \frac{\sum fd'}{\sum f} \times i$, This is also called the **Step Deviation Method**.

Ex : If the sum of the deviation of a set of 10 values taken from 35 is 60, find the value of the A.M.

Solution :

A.M. = $A + \frac{\sum d}{n} = 35 + \frac{60}{10} = 41$

Ex : Calculate the A.M. from the following data.

Class	**Frequency**
20 – 25	10
25 – 30	12
30 – 35	8
35 – 40	20
40 – 45	11
45 – 50	4
50 - 55	5

Sol :

Class	Frequency *f*	Mid-Value *x*	*d*=*x* – 37.5	*fd*	*d*/5 = *d'*	*fd'*
20 – 25 25 – 30 30 – 35 35 – 40 40 – 45 45 – 50 50 – 55	10 12 8 20 11 4 5	22.5 27.5 32.5 37.5=A 42.5 47.5 52.5	-15 - 10 - 5 0 5 10 15	-150 - 120 - 40 0 55 40 75	-3 - 2 - 1 0 1 2 3	-30 - 24 - 8 0 11 8 15
	N = 70			Σ *fd* = - 140		Σ *fd'* = *-28*

A.M. = $\overline{x} = A + \frac{\sum fd}{N} = 37.5 + \frac{(-140)}{70} = 35.5$

Ex : Company X pays wages to the workers, Department of finance is interested to know, what is the average wage paid to the workers? Following data provides the necessary information:

Wages	**No. of workers**
240	5
250	15
260	32
270	42
280	15
290	12
300	4

Let us take 250 as assumed mean.

Wages (X)	No. of Workers *(f)*	X - A = X - 250 = *d*	*fd*
240	5	-10	-50
250	15	0	0
260	32	10	320
270	42	20	840
280	15	30	450
290	12	40	480
300	4	50	200
Total	N = 125		$\sum fd = 2240$

$$Mean = A + \frac{\sum fd}{\sum f}$$

Mean of the given series = 267.92.

Wages (X)	No. of Workers *(f)*	X - A = X - 250 = *d*	d'= d/10	*fd'*
240	5	-10	-1	-5
250	15	0	0	0
260	32	10	1	32
270	42	20	2	84
280	15	30	3	45
290	12	40	4	48
300	4	50	5	20
Total	N = 125			224

$$Mean = A + \frac{\sum fd'}{\sum f} \times 10$$

Mean of the given series = 267.92.

Charlier's Accuracy Check:

Prof. Charlier has provided a formula for checking the accuracy of calculations involved to computing mean of a discrete or a continous frequency distribution. The formula is,

$$\sum f(d+1) = \sum fd + \sum f \quad or \quad \sum f(d'+1) = \sum fd' + \sum f$$

Case:

Following table gives the data collected by a soft drink manufacturer. The question of purpose was to know the age group, that likes a particular flavour most. The data was collected on a total sample size 120. The sample was containing the people from various age group of particular township.

Class	Frequency
0 - 10	12
10-20	15
20-30	25
30-40	36
40-50	18
50-60	14

Solution:

Class	*Mid - Value (x)*	*Frequency (f)*	*d = (x-35)*	*d' = d/10*	*fd'*	*f(d'+1)*	*f(d+1)*	*fd*
0 - 10	5	12	-30	-3	-36	-24	-348	-360
10-20	15	15	-20	-2	-30	-15	-285	-300
20-30	25	25	-10	-1	-25	0	-225	-250
30-40	35	36	0	0	0	36	36	0
40-50	45	18	10	1	18	36	198	180
50-60	55	14	20	2	28	42	294	280
Total		N = 120			-45	75	-330	-450

Since: $\sum f(d+1) = \sum fd + \sum f \ or \sum f(d'+1) = \sum fd' + \sum f$

Therefore, - 330 = - 450 + 120 = - 330

and, 75 = - 45 + 120 = 75

Case-let:

The following are hourly salaries in Rs of 20 employees of a firm:

130	62	145	118	125	76	151	142	110	98
65	116	100	103	71	85	80	122	132	95

The firm gives bonuses of Rs 10, 15, 20, 25 and 30 for individuals in the respective salary groups exceeding Rs 60 but not exceeding Rs 80, Exceeding Rs 80 but not exceeding Rs 100, and so on up-to exceeding Rs 140 but not exceeding Rs 160. Find the average hourly bonus point per employee.

Geometric Mean :

Sometimes when we are dealing with quantities that change over a period of time, we need to know an average rate of change, such as average growth rate over a period of several years. In such cases, the simple arithmetic mean is inappropriate, because it gives the wrong answers. We need to know 'geometric mean' simply called G.M.

G.M. is generally used in situations where small items are assigned large weights and vice versa. The following are the some uses of geometric mean:

1. G.M. is useful to calculate average percentage increase or decease.
2. In the construction of Index numbers, geometric mean is considered to be best average tool.

Geometric Mean :

Geometric Mean of n positive numbers is the nth root of their product. If $x_1, x_2, \ldots\ldots\ldots\ldots, x_n$ are the n values of a variable then their

$$G.M. = [x_1 \times x_2 \times \ldots\ldots x_n]^{\frac{1}{n}} \Rightarrow \log(G.M.) = \frac{1}{n}[\log x_1 + \log x_2 + \ldots\ldots + \log x_n]$$

If $f_1, f_2, \ldots\ldots\ldots\ldots, f_n$ are the frequencies of the values $x_1, x_2, \ldots\ldots\ldots\ldots, x_n$, then

$$G.M. = [x_1^{f_1} \times x_2^{f_2} \times \ldots\ldots\ldots \times x_n^{f_n}]^{\frac{1}{f_1 + f_2 + \ldots\ldots\ldots + f_n}}$$

$$\Rightarrow \log G.M. = \frac{1}{f_1 + f_2 + \ldots\ldots\ldots + f_n}[f_1 \log x_1 + f_2 \log x_2 + \ldots\ldots\ldots\ldots + f_n \log x_n]$$

i.e. the logarithm of G.M. of a set of values is equal to the arithmetic mean of their logarithms.

$$\Rightarrow G.M. = anti\ \log[\frac{1}{f_1 + f_2 + ... + f_n}\{f_1 \log x_1 + f_2 \log x_2 + ... + f_n \log x_n\}]$$

Generally, Geometric Mean of a number of values

$$= [\text{Product of the Values}]^{\frac{1}{[\text{Number of Values}]}}$$

If one of the values is zero G.M. can not be defined.

Combined Geometric Mean of Two Groups :

We know G.M. of n values $x_1, x_2, \dots\dots\dots\dots, x_n = [x_1 \times x_2 \times \dots\dots\dots\dots \times x_n]^{1/n}$

$\Rightarrow \quad [G.M.]^n = [x_1 \times x_2 \times \dots\dots\dots\dots \times x_n]$

$\Rightarrow \quad [G.M.]^{\text{Number of Values}} = [\text{Product of Values}]$

If G_1 and G_2 be geometric mean of two groups having n_1 and n_2 are the number of items respectively, then the combined G.M. 'G' of these two groups is defined by

$$G = [\{G_1\}^{n_1} \times \{G_2\}^{n_2}]^{\frac{1}{n_1 + n_2}}$$

$$\Rightarrow \text{Log } G = \frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2}$$

If $n_1 = n_2$, then

$$\Rightarrow \text{Log } G = \frac{\log G_1 + \log G_2}{2}$$

$$= \frac{1}{2} [\log G_1 G_2]$$

$$= \log [G_1 G_2]^{1/2}$$

$$G = [G_1 \times G_2]^{1/2} = \sqrt{[G_1 \times G_2]}$$

Ex:
Calculate the G.M. of the following data.
70, 10, 500, 75, 8, 250.

Given, log 70=1.845, log 500=2.69897, log 75=1.875, log 8=0.9030, log 250=2.3979
Antilog 1.619985 = 41.6855

Solution :

$$G.M. = [70 \times 10 \times 500 \times 75 \times 8 \times 250]^{\frac{1}{6}}$$

$$= \frac{1}{6}\log[70 \times 10 \times 500 \times 75 \times 8 \times 250]$$

$$= \frac{1}{6}[\log 70 + \log 10 + \log 500 + \log 75 + \log 8 + \log 250]$$

$$= \frac{1}{6}[1.845 + 1 + 2.69897 + 1.875 + 0.9030 + 2.3979]$$

$$= \frac{1}{6}[9.71991]$$

$$= 1.619985$$

$$G.M. = anti\ \log 1.619985 = 41.6855$$

Ex:

Find the G.M. of 3, 6, 24, 48.

Solution:

$$\text{G. M.} = (3 \times 6 \times 24 \times 48)^{\frac{1}{4}}$$

$$\text{Log GM} = \frac{1}{4}\log(3 \times 6 \times 24 \times 48)$$

$$\text{G. M.} = \text{antilog}\ (1.07918) = 12$$

Case:

The price of a commodity increased by 8 % from 1993 to 1994, 12 % from 1994 to 1995, and 76 % from 1995 to 1996. The average price increase from 1994 to 1996 Is quoted as 28.64 % and not 32 %. Explain and verify the result.

Solution:

Here the average of the percentage over a period of three years is to be computed. Thus, G.M. is the most appropriate average tool.

$\bar{x} = 32\%$

Computing the G.M.

Year	Percentage Increase	Price level at the end of the year with the preceding year’s level as base
1993-1994	8	108
1994-1995	12	112
1995-1996	76	176

GM 128.6426 Average Increase = 128.6426 – 100 = 28.6426

Case 1: When the commodity prices increases by 8 % from 1993 – 1994, 12 % from 1994 – 1995 and 76 % from 1995 – 1996.

Year	Rate of Increase	Total Increase	Price at the end of each year
1993-1994	8 % on 100	8	108
1994-1995	12 % on 108	12.96	120.96
1995-1996	76 % on 120.96	91.92	212.88

Case 2: When the average increase is 32 % (arithmetic mean) per year

Year	Rate of Increase	Total Increase	Price at the end of each year
1993-1994	32 % on 100	32	132
1994-1995	32 % on 132	42.24	174.24
1995-1996	32 % on 174.24	55.75	229.99

Case 3: When the average increase is 28.64266 % (geometric mean) per year

Year	Rate of Increase	Total Increase	Price at the end of each year
1993-1994	28.64266 % on 100	28.64266	128.64266
1994-1995	28.64266 % on 128.64266	36.84667972	165.4893397
1995-1996	28.64266 % on 165.4893397	47.40054891	212.88

Results are same in case 1 and 3.

Harmonic Mean :

The harmonic mean of a number of values of a variable is the reciprocal of the arithmetic mean of the reciprocals of the given values of the variable. Symbol H.M. is used to denote harmonic mean.

If $x_1, x_2, \ldots\ldots\ldots\ldots, x_n$ are n values of the variables then,

$$H.M. = \frac{1}{\dfrac{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots\ldots\ldots + \dfrac{1}{x_n}}{n}} = \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots\ldots\ldots + \dfrac{1}{x_n}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

If f_1, f_2, ..., f_n are the frequencies of the values x_1, x_2, ...,x_n then ,

$$H.M. = \frac{\sum_{i=1}^{n} f_i}{\dfrac{f_1}{x_1} + \dfrac{f_2}{x_2} + \ldots\ldots\ldots + \dfrac{f_n}{x_n}}$$

If each observation is divided by a constant. The H.M. is also divided by the same constant. In (1) if each observation is divided by ‘K’ then,

$$H.M. = \frac{n}{\frac{1}{x_1/k} + \frac{1}{x_2/k} + \dots\dots\dots\dots + \frac{1}{x_n/k}} = \frac{1}{k}\left[\frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots\dots\dots + \frac{1}{x_n}}\right]$$

= 1/k times the original H.M.

Similarly if each observation is multiplied by a constant H.M. is also multiplied by the same constant.
The use of harmonic Mean is very limited it is generally used in averaging speed price of articles, etc.

G.M. = √(A.M.×H.M.)

Following table shows salary ranges(in thousand rupees) and the number of employees for a manufacturing firm

Salary Range	Number of Employees
50 – 60	12
60 – 70	15
70 – 80	20
80 – 90	44
90 – 100	42
100 – 110	32
110 – 120	32
120 – 130	12
Total	209

Calculate HM and GM.

Salary	***Employees (f)***	***x***	***f/x***	***Log x***	***f log x***
50 - 60	12	55	0.218182	1.740363	20.88435
60 - 70	15	65	0.230769	1.812913	27.1937
70 - 80	20	75	0.266667	1.875061	37.50123
80 - 90	44	85	0.517647	1.929419	84.89443
90 - 100	42	95	0.442105	1.977724	83.06439
100 - 110	32	105	0.304762	2.021189	64.67806
110 - 120	32	115	0.278261	2.060698	65.94233
120 - 130	12	125	0.096	2.09691	25.16292
Total	**N = 209**		**2.354393**		**409.3214**

$$HM = \frac{N}{\sum \frac{f}{x}} = \frac{209}{2.354} = 88.78$$

$$\text{GM} = antilog\left[\frac{\sum f\ logx}{N}\right] = antilog\left[\frac{409.3211}{209}\right] = antilog\,(1.9584) = 90.88$$

MEDIAN

Median :

According to Connar, “The median is that value of the variable which divides the group in to two equal parts, one part comprising all values greater, and other all values less than median”.

To find the median of ungrouped data the data is arranged either in ascending order of magnitude or descending order of magnitude. After arranging the data in any of the two ways the middlemost value gives the median.

“If the number of items are odd, then **$\{(N+1)/2\}^{th}$** value gives the median and if the number of items are even, then the **arithmetic mean of $(N/2)^{th}$ and $\{(N/2)+1\}^{th}$ value** gives the median value”.

Ex:

Find the median of the following series

2,10, 5, 7, 15, 17, 21, 13, 4

and 2,10, 5, 7, 15, 17, 21, 13, 4, 22.

Sol:

After arranging the values in ascending order we have:

2, 4, 5, 7, 10, 13, 15, 17, 21......................................(1)

2, 4, 5, 7, 10, 13, 15, 17, 21, 22...............................(2)

Series (1) contains 9 items, therefore {(9+1)/2}=5th value = 10 is the median.

Series (2) contains 10 items, therefore

$$Median = \frac{[(\frac{N}{2})th + \{(\frac{N}{2}) + 1\}th]}{2} = \frac{[(\frac{10}{2})th + \{(\frac{10}{2}) + 1\}th]}{2} = \frac{5th + 6th}{2} = \frac{10 + 13}{2} = 11.5$$

Median is influenced by the position of the items in the array but not by the size of the items.

Median Of Grouped Data :

In grouped data, the individual items lose their identity, and the middle item can not be found by counting. Therefore, it is necessary to get inside of a class to find to find the value that divides the number of all items in half . If we divide the number of frequencies (N) in two halves.

We cumulate the frequencies until the we reach the lowest class whose cumulative frequency is greater than (N/2).

This class is a **median class.** Median class is the lowest class whose cumulative frequencies contain the value (N/2). When the class intervals are arranged in ascending order. Clearly the median value is more than the lower limit of the median class and less or equal to the upper limit of the median class.

Assuming that all items are evenly distributed over this median class, We proceed toward the upper limit of this class, Stopping when we have reached at the frequency equal to (N/2), this operation brings us to a value with in the median class which is presumed to have (N/2) items on each side of it.
As already stated, median value is at-least as high as the lower limit of the median class but can not be higher than the upper limit of the median class.

$$M_e = L + \frac{\frac{N}{2} - C.f._{-1}}{f_{med}} \times i_{med}$$

Where,

M_e = Median

L = Lower Limit of the class

N = Total frequency

$C.\ f._{-1}$ = Cumulative frequency of the preceding class

f_{med} = frequency of median class

i_{med} = class interval of median class

The Steps for Finding the Median for Grouped Data are as Follows:

1. Divide the number of items in the distribution by 2, i.e. compute the value of (N/2).

2. Accumulate i.e. cumulate the frequencies.

3. Find the class whose cumulative frequency is first to exceed N/2. This is the median class.

4. Find the actual lower limit of the median.

5. Then perform the following operations: Subtract from N/2 the frequencies we have accumulated before entering the median class. Divide this difference by the frequency of the median class. Multiply the quotient thus obtained by the size of the class interval of the median class.

6. Add the result of the operation in step 5 to the lower limit of the median class. This sum gives the median.

The formula for this procedure is

$$median = l_1 + \left(\frac{\frac{N}{2} - Cf_{-1}}{f_{med}} \right) \times i$$

Where,

l_1 = The lower limit of the Median Class.

N = Total Frequency.

Cf_{-1} = Cumulative Frequency of the Class Preceding the Median Class.

f_{med} = The frequencies of the Median Class.

i = The Size of the Class interval of the Median Class.

For finding the median for grouped data also, the class intervals should be arranged in ascending order of the magnitude.

The median can also be found for grouped data by entering the median class at its upper limit. In this case the formula is

$$median = l_2 - \left(\frac{\frac{N}{2} - \sum f_2}{f_{med}} \right) \times i$$

Where,

l_2 = The Upper limit of the Median Class.

N = Total Frequency.

Σf_2 = Sum of frequencies accumulated from the highest class to the class immediately above the median class in value, i.e. sum of the frequencies of all the classes succeeding the median classes.

f_{med} = The frequencies of the Median Class.

i = The Size of the Class interval of the Median Class.

Related Positional Measures i.e. Quartiles, Deciles and Percentiles :

There are some other measures that divides the series into equal parts like quartiles divide the data into four equal parts, Deciles divide the data into 10 equal parts and Percentile divides the data into 100 equal parts.

Following are the formulae for finding them :

$$Q_r = l_1 + \frac{r\left(\frac{N}{4}\right) - c}{f}(l_2 - l_1)$$

$$D_r = l_1 + \frac{r\left(\frac{N}{10}\right) - c}{f}(l_2 - l_1)$$

$$P_r = l_1 + \frac{r\left(\frac{N}{100}\right) - c}{f}(l_2 - l_1)$$

Where,

l_1 = The lower limit of the quartile class,

l_2= The upper limit of the quartile class,

f = The frequency of the quartile class,

c = Cumulative frequency of the class preceding the quartile class.

Solved Examples:

1. Find the median of the following Numbers
25, 24, 23, 32, 40, 27, 30, 25, 20, 10, 15, 45.

Solution:

Arranging the data in ascending order, we get
10, 15, 20, 23, 24, 25, 25, 27, 30, 32, 40, 45.

There are 12 items in the numbers in the series which is an even number

Therefore,

The A.M. of 6^{th} and 7^{th} item is the median

6^{th} item = 25

7^{th} item = 25

Hence median = (25+25)/2 = 25

2. From the following data, calculate the values of the upper and lower quartiles, D_2, P_{30}.

Marks	Below 10	10 – 20	20 – 40	40 – 60	60 – 80	Above 80
No. of Students	8	10	22	25	10	5

Solution :
We make the following cumulative frequency table :

Marks	No of Students (f)	Cumulative Frequency (*c.f.*)
0 – 10	8	8
10 – 20	10	18
20 – 40	22	40
40 – 60	25	65
60 – 80	10	75
80 – 100	5	80
Total	$N = \Sigma f = 80$	

For Q_1 :

N = 80 $\Rightarrow$ N/4 = 20

l_1 = 20

f = 22

$c.f_{-1}$ = 18

Therefore,

$$Q_1 = l_1 + \frac{r\left(\frac{N}{4}\right) - cf_{-1}}{f} \times i$$

$$\Rightarrow Q_1 = 20 + \frac{20 - 18}{22} \times 20 = 20 + \frac{2}{22} \times 20 = 21.82$$

For Q_3 :

N = 80 $\Rightarrow$ 3N/4 = 60
l_1 = 40
f = 25
$c.f_{-1}$ = 40

Therefore,

$$Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - cf_{-1}}{f} \times i$$

$$\Rightarrow Q_3 = 40 + \frac{60 - 40}{25} \times 20 = 40 + \frac{20}{25} \times 20 = 56$$

For D_2:

N = 80 $\Rightarrow$ 2N/10 = 16
l_1 = 10
f = 10
$c.f_{-1}$ = 8

Therefore,

$$D_2 = l_1 + \frac{2\left(\frac{N}{10}\right) - cf_{-1}}{f} \times i$$

$$\Rightarrow D_2 = 10 + \frac{16 - 8}{10} \times 10 = 10 + \frac{8}{10} \times 10 = 18$$

For P_{30} :

N = 80 $\Rightarrow$ 30N/100 = 24

l_1 = 20

f = 22

$c.f_{-1}$ = 18

Therefore,

$$P_{30} = l_1 + \frac{30\left(\frac{N}{100}\right) - cf_{-1}}{f} \times i$$

$$\Rightarrow P_{30} = 20 + \frac{24 - 18}{22} \times 20 = 20 + \frac{6}{22} \times 20 = 25.46$$

$$M_e = l_1 + \frac{\left(\frac{N}{2}\right) - cf_{-1}}{f_{med}} \times i$$

$$\Rightarrow 40 + \frac{40 - 40}{22} \times 20 = 40$$

GRAPHICAL METHOD OF FINDING MEDIAN

?

Mode :

Mode is the most common value, i.e., the value having maximum frequency in the series. For example

1. The retail price paid for a commodity by most of the customers in the model price.
2. In a factory the wage being paid to most of the workers in modal wage.
3. In grouped frequency distribution the class having the highest frequency is the model class.

Definitions :

According to Zizek, Mode is "the value occurring most frequently in a series of items and around which the other items are distributed most densely".

Mode of Ungrouped Data :

Find the mode of the following data :

41, 42, 45, 44, 45, 48, 50, 45, 47, 51, 56.

Arranging the distribution as frequency distribution, we get

Value	Frequency
41	1
42	1
45	3
44	1
48	1
50	1
47	1
51	1
56	1

Mode ← 45

Mode by Grouping Method:

For this method a grouping table has to be maintained:

Column 1: Contains the original frequencies arranged in ascending order and the maximum frequency is marked by putting a mark or a circle.

Column 2: The frequencies are grouped twos, the total remaining frequencies are grouped in tows and the highest total is marked.

Column 3: Leaving the first frequency, the remaining frequencies are grouped in twos and the highest total is marked.

Column 4: The frequencies are grouped in threes. The highest total is marked.

Column 5: Leaving the first frequency, the remaining frequencies grouped in threes and the highest total is marked.

Column 6: Leaving the first two frequencies, the remaining frequencies are grouped in threes and the highest total is marked.

"After completing the group table analysis table is formed for finding the value or the observation which is repeated the highest number of times".

Ex:

The following table gives the measurement of collar size of 230 students in a university. Determine the model size of the collar.

Collar Size (cms)	32	33	34	35	36	37	38	39	40	41
No. of Students	7	14	30	28	35	34	16	14	36	16

Solution : Determining the Mode by Grouping Method

Collar Size	Frequency					
	Col (1)	Col (2)	Col (3)	Col (4)	Col (5)	Col (6)
32	7	21				
33	14		44	51		
34	30	58			72	
35	28		**63**			**93**
36	35	**69**		**97**		
37	34		50		**85**	
38	16	30				64
39	14		50	66		
40	**36**	52			66	
41	16					

Analysis Table

Column No.	Collar Size values contributing to the highest Frequency									
	32	33	34	35	36	37	38	39	40	41
1									√	
2					√	√				
3				√	√					
4				√	√	√				
5					√	√	√			
6			√	√	√					
No. of Times	-	-	1	3	5	3	1	-	1	-

From analysis table we observe that size 36 is repeated the highest number of times. Therefore, Model size is 36 cms.

Determination of Mode in a Continuous Series :

In this method the class having the maximum frequency is known as the modal class. In case nearly equal concentration of frequencies is observed in two or more classes, the grouping method can be used to determine the model class. Afterwards following formula is applied to define mode:

$$Mode = l_1 + \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}} \times i$$

Where,

l_1 = The lower limit of the modal class,

f_m = The frequency of the modal class,

f_{m-1} = The frequency of the class preceding modal class,

f_{m+1} = The frequency of the class succeding modal class,

i = Size of the modal class, i.e., class interval of the modal class.

Ex:
Compute the mode from following data

Class	0 – 3	3 – 6	6 – 10	10 – 12	12 – 15	15 – 18
Frequency	4	8	10	14	16	20
Class	18 – 20	20 – 24	24 – 25	25 – 28	28 – 30	30 – 36
Frequency	24	14	16	11	10	6

Solution :
In this example the class intervals are unequal. For finding the mode class intervals has to be equal. So converting the class intervals in to equal lengths i.e.,
0 – 6, 6 – 12, 12 – 18, ...

We have the following arrangement

Grouping Table

Class	Frequency					
	Col (1)	Col (2)	Col (3)	Col (4)	Col (5)	Col (6)
0 – 6	4+8= 12	36		72		
6 – 12	10+14= 24		60		98	
12 – 18	16+20= 36	74				111
18 – 24	24+14= 38		75	81		
24 – 30	16+11+10= 37	43				
30 – 36	= 6					

Analysis Table

Column No.	Collar Size values contributing to the highest Frequency				
	6 - 12	12 - 18	18 - 24	24 – 30	30 - 36
1			√		
2		√	√		
3			√	√	
4			√	√	√
5	√	√	√		
6		√	√	√	
No. of Times	1	3	6	3	1

Now the Mode lies in the class 18 – 24. Whose frequency is 38.

Therefore,

l_1 = 18

f_m = 38

f_{m-1} = 36

f_{m+1} = 37

i = 6

Hence,

$$Mode = 18 + \frac{38 - 36}{2 \times 38 - 36 - 37} \times 6 = 22$$

Empirical Relation Between Mean, Median and Mode:

❖ Mode = 3 Median – 2 Mean

Selection of an Appropriate Average :

No single average is sufficient to serve each and every purpose. To find a suitable average, the situation has to be analyzed thoroughly and carefully. **Mean** takes in to account all values of distribution. If we wish to make total value estimations, **Mean** is the most appropriate average e.g., per capita income of the country.

In case of open ended class intervals mean can not be found, hence in such a case we have to depend on **Median** and **Mode**.

If the data is qualitative, e.g., beauty, honesty, intelligence, etc, then **Median** will be the most appropriate average. Regarding agricultural land holdings, psyche and social problems **Median** is used.

In case of business situations, where selection of ‘most common’, e.g., most common size of shoes for certain age group, most common size of garments, etc., is to be made, the **Mode** is most common average in use.

1. Graphical method of finding mode.
2. Limitations, Merits and Uses of Mean, Median and Mode.

?

Exercise :

Compute mean and mode of the following distribution of marks obtained by 125 students in statistics in a certain examination :

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Students	3	4	8	10	15
Marks	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
Students	35	20	16	8	6

Answer

Mode = 55.7

Mean = 55.8

Commercial Averages:

- Quadratic Average
- Moving Average
- Progressive Average
- Composite Average

www.v3solutions.in