UNIT 6: MATHEMATICAL MEASURES OF CENTRAL TENDENCY

CONTENTS

6.0 Aims and Objectives

6.1 Introduction

6.2 Summation Notation and Its Properties

6.3 Arithmetic Mean (AM)

6.4 Geometric Mean (GM)

6.5 Harmonic Mean (HM)

6.6 Advantages and Application Areas of the Three Means

6.7 Summary

6.8 Model Examination Questions

6.9 Answers to Check Your Progress Questions

6.10 Glossary

6.11 References

6.0 AIMS AND OBJECTIVES

At the end of this unit, you will be able to:

- Apply properties of summation notation to solve problems which involve the summation operator.
- Understand the definition and properties of AM, GM and HM.
- Compute problems of the three means for ungrouped and grouped data.
- Identify the advantages and application areas of the three means.

6.1 INTRODUCTION

The definition should not leave anything to the description of the person who calculated averages. Averages should be computed with sufficient ease and rapidity or averages should not involve more of mathematical complexities. The most popular and widely used measure for representing the entire data by one value is arithmetic mean.

6.2 SUMMATION NOTATION AND ITS PROPERTIES

Summation operator, **Σ**, implies that the values that follow it are to be summed or added together.

$$\sum_{i=m\rightarrow lower\ limit}^{n\rightarrow upper\ limit} x_i \rightarrow the\ i^{th}\ variable\ of\ x$$

Example $\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$

Properties:

1. The summation of sums of differences

$$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \quad , \quad \sum_{i=1}^{n}(x_i - y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$$

Example: Suppose $x_1 = 1$, $x_2 = 3$, $x_3 = 4$, $y_1 = 2$, $y_2 = 5$, $y_3 = 3$

Then $\sum_{i=1}^{3}(x_i + y_i) = \sum_{i=1}^{3} x_i + \sum_{i=1}^{3} y_i$

$(x_1 + y_1) + (x_2 + y_2) + (x_3 + y_3) = (x_1 + x_2 + x_3) + (y_1 + y_2 + y_3)$

$(1 + 2) + (3 + 5) + (4 + 3) = (1 + 3 + 4) + (2 + 5 + 3)$

$3 + 8 + 7 = 8 + 10$

$18 = 18$

$\sum_{i=1}^{3}(x_i - y_i) = \sum_{i=1}^{3} x_i - \sum_{i=1}^{3} y_i$ ..... left for the student

2. <u>Multiplication by a constant</u>

$$\sum_{i=1}^{n} kx_i = k\sum_{i=1}^{n} x_i$$

<u>Example:</u> Suppose k = 7 and $x_1 = 2$, $x_2 = 5$, $x_3 = 4$

Then $\sum_{i=1}^{3} kx_i = k\sum_{i=1}^{n} x_i$

$kx_1 + kx_2 + kx_3 = k(x_1 + x_2 + x_3)$

$7(2) + 7(5) + 7(4) = 7(2 + 5 + 4)$

$14 + 35 + 28 = 7(11)$

$77 = 77$

7.0 <u>Summation of a constant</u>

Case 1: If lower limit equal to 1, $\sum_{i=1}^{n} k = nk$

<u>Example:</u> suppose k = 6 and upper limit = 4

Then $\sum_{i=1}^{4} k = nk$

$\sum_{i=1}^{4} 6 = 4(6)$

5 + 6 + 6 + 6 = 24

Case 2: if lower limit is different from 1 or lower limit is greater than 1,

$\sum_{i=m}^{n} k = [(n - m) + 1]k$ for m < n

Example: Suppose k = 8 and upper limit = 6 & lower limit = 4 then

$$\sum_{i=m}^{n} k = [(n - m) + 1]k$$

$$\sum_{i=4}^{6} 8 = [(6 - 4) + 1]8$$

$8 + 8 + 8 = 3(8)$

$24 = 24$

4. Sum of summations

$$\sum_{i=1}^{k} x_i + \sum_{i=k+1}^{n} x_i = \sum_{i=1}^{n} x_i \quad \text{for any } k < n$$

Example: Suppose $x_1 = 2$, $x_2 = 4$, $x_3 = 6$, $x_4 = 3$, $x_5 = 2$, $x_6 = 4$ let k = 3 then

$$\sum_{i=1}^{k} x_i + \sum_{i=k+1}^{n} x_i = \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{3} x_i + \sum_{i=4}^{6} x_i = \sum_{i=1}^{6} x_i$$

$(x_1 + x_2 + x_3) + (x_4 + x_5 + x_6) = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$

$(2 + 4 + 6) + (3 + 2 + 4) = (2 + 4 + 6 + 3 + 2 + 4)$

$21 = 21$

CYP 1 **Let** $\sum_{i=3}^{6} x_i = 10$, $\sum_{i=3}^{6} x_i^2 = 148$, $x_1 = 3$, $x_2 = 2$

Find i. $\sum_{i=1}^{6} x_i$ ii. $\sum_{i=1}^{6} x_i^2$ iii. $\sum_{i=1}^{6} x_i(x_i - 2)$

iv. $\sum_{i=1}^{6} (2x_i - 3)^2$ v. $\sum_{i=1}^{2} (ix_i + 4)$

6.3 ARITHMETIC MEAN (AM)

6.3.0 Definition

The arithmetic mean is the sum of the values in a group divided by the number of items in that group. Let $x_1, x_2, \ldots, x_n$ be n values of a variable x, then their arithmetic mean is defined by:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}$$

Where Σx – sum of all observations

n – total number of observations

6.3.1 Computation Of Arithmetic Mean For Ungrouped And Grouped Data

For ungrouped data:

Direct method: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ Short cut method:

$$\bar{x} = A + \frac{\sum d}{n}$$

Where n – number of items A = Assumed mean Σd = sum of deviations i.e. (x_i - A)

Example: Find the arithmetic mean for the following data by

i. direct method ii. short cut method

23.4 15.6 22.1 20.0 26.7 31.4 18.9 22.3

Solution:

i. $\sum_{i=1}^{8} x_i = 180.4$, n = 8 $\bar{x} = \frac{\sum_{i=1}^{8} x_i}{n} = \frac{180.4}{8} = 22.55$

ii. Let A = 22 then d_i : 1.4, -6.4, 0.1, -2, 4.7, 9.4, -3.1, 0.3

$\sum_{i=1}^{8} d_i = 4.4$, n = 8 $\quad \bar{x} = A + \frac{\sum_{i=1}^{8} d_i}{n} = 22 + \frac{4.4}{8} = 22 + 0.55 = \underline{22.55}$

For grouped data:

For Discrete Series:

Direct method: $\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{n} = \frac{\sum fx}{n}$ Short cut method:

$\bar{x} = A \pm \frac{\sum fd}{n}$

Where f - frequency d - deviation of items from assumed mean (x_i – A)

A - assumed mean n - number of observations

Example: Given data of 50 students of marks of a test in a class.

Calculate the arithmetic mean by i. direct method

ii. short cut method.

No. of Students	20	30	40	50	60	70
Marks	8	10	16	8	5	3

Solution:

Marks x_i	f_i	fx	**Di = (x – 40) where A = 40**	fd
20	8	160	-20	-160
30	10	300	-10	-100
40	16	640	0	0
50	8	400	10	80

60	5	300	20	100
70	3	210	30	90
	50	2010		10

i. $\bar{x} = \frac{\sum fx}{n} = \frac{2010}{50} = \underline{\underline{40.20}}$

ii. $\bar{x} = A \pm \frac{\sum fd}{n} = 40 + \frac{10}{50} = \underline{\underline{40.20}}$

For continuous series:

Direct method $\bar{x} = \frac{\sum fcm_i}{n}$ Step deviation method

$\bar{x} = A \pm \left(\frac{\sum fd^1}{n} \right) \times c$

Where , f – frequency n – number of observation

Cm_i – class mark A – assumed mean

d – derivation of class marks from assumed mean (cm_i – A)

d' – d/c c – class width

Example: In a survey, the number of persons at different ages is found as follows:

Age in Year	5 - 15	15 - 25	25 - 35	35 - 45	45 - 55	55 - 65
No. of Persons	8	10	14	20	16	12

Solution:

Classes	**f**	**cmi**	**fcm**	**d = cm – A where A = 30**	**df**	**d1 = d/c**	**fd1**
5 - 15	8	10	80	-20	-160	-2	-16
15 - 25	10	20	200	-10	-100	-1	-10

25 - 35	14	30	420	0	0	0	0
35 - 45	20	40	800	10	200	1	20
45 - 55	16	50	800	20	320	2	32
55 - 65	12	60	720	30	360	3	36
	80		3020		620		62

i. $\bar{x} = \frac{\sum fcm}{n} = \frac{3020}{80} = \underline{\underline{37.75}}$

ii. $\bar{x} = A + \frac{\sum fd}{n} = 30 + \frac{620}{80} = 30 + 7.75 = \underline{\underline{37.75}}$

iii. $\bar{x} = A + \left(\frac{\sum fd'}{n}\right) \times c = 30 + \left(\frac{62}{80}\right) 10 = 30 + (0.775)10 = \underline{\underline{37.75}}$

CYP 2 Find the arithmetic mean of the following data.

i. 53 54 52 32 30 60 47 46 35 29

ii.

Height (in inches)	64	65	66	67	68	69	70	71	72	73
No of Students	4	9	12	18	20	12	10	9	4	2

iii.

Classes	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35
Frequencies	5	6	7	7	5

6.3.2 Properties of the Arithmetic Mean

The mathematical properties of an arithmetic mean are discussed below:

1) The product of arithmetic mean $(\bar{x})$ and the number of observations (N) on which the mean based is equal to the sum of all given values i.e. $\bar{x}N = \sum_{i=1}^{n} x_i$

2) The algebraic sum of the deviations of the given values from the arithmetic mean is equal to zero. Mathematically,

$\sum (x_i - \bar{x}) = 0$... for ungrouped data

$\sum f_i (x_i - \bar{x}) = 0$... For grouped data

Because of this property, the arithmetic mean may be characterized as a point of 'Balance'

3) The sum of squares of deviations is minimum when deviations are taken from the arithmetic mean. i.e.

$\sum (x_i - \bar{x})^2 \leq \sum (x_i - A)^2$ ... For ungrouped data

$\sum f_i (x_i - \bar{x})^2 \leq \sum f_i (x_i - A)^2$ ...For grouped data. Where, A is different from mean.

4) Suppose the mean of the values $x_1, x_2, \ldots, x_n$ be $\bar{x}_0$. Then

i. if a constant k is added to each x_i, then the new mean $\bar{x}_n$ will be $\bar{x}_0 + k$.

Proof: Arithmetic mean of $x_1 + k, x_2 + k, \ldots, x_n + k$ is

$$A.M = \frac{\sum_{i=1}^{n} (x_i + k)}{n} = \frac{(x_1 + k) + (x_2 + k) + \ldots}{n}$$

$$A.M = \frac{(x_1 + x_2 + \ldots + x_n) + (k + k + \ldots + k)}{n}$$

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

ii. if each new mean will be $k\bar{x}_0$

Proof: A.M for kx_1 , kx_2, … kx_n, is

$$AM = \frac{\sum_{i=1}^{n} kx_i}{n} = \frac{kx_1 + kx_2 + \ldots + kx_n}{n}$$

$$AM + \frac{k(x_1 + x_2 + \ldots + x_n)}{n}$$

$$AM = \underline{\underline{k\bar{x}_0}}$$

Example: Given data 12, 10, 8, 6, 16, 7, 11. If each item is multiplied by $\frac{4}{5}$ and 8 is added, what will be the new mean?

$$\bar{x}_0 = \frac{\sum_{i=1}^{7} x_i}{7} = \frac{70}{7} = \underline{\underline{10}}$$ New mean

$$\bar{x}_n = \frac{4}{5}\bar{x}_0 + 8$$

$$\bar{x}_n = \frac{4}{5}(10) + 8 = \underline{\underline{16}}$$

CYP 3 Given data 3, 8, 9, 4, 7, 5, 10, 11, 6 if each item is multiplied by 2 and 6 is added, then

i. The new mean will be _____

ii. $\sum(x_i - \bar{x}) =$ _____

6.3.3 Combined (Pooled) Arithmetic Mean

Let $\bar{x}_1$ and $\bar{x}_2$ be the AM of two groups, having N_1 and N_2 observations respectively. If we combine the two groups in to a single group, then

the arithmetic mean of the combined group will be

$$\bar{x}_c = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2}$$

For n number of groups, $\bar{x}_c = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + \ldots + N_n x_n}{N_1 + N_2 + \ldots + N_n} = \frac{\sum_{i=1}^{n} N_i\bar{x}_i}{\sum_{i=1}^{n} N_i}$

Example: The mean height of 25 male and 20 female is 161.0cm and 155.6cm. What will be the combined mean height?

$\bar{x}_m$ = 161.0cm, $\bar{x}_F$ = 155.6cm, N_M = 25, N_F = 20

$$\bar{x}_c = \frac{\bar{x}_m N_m + \bar{x}_F N_F}{N_m + N_F}$$

$$\bar{x}_c = \frac{1610(25) + 1556(20)}{25 + 20} = \frac{7137}{45} = 158.60cm$$

CYP 4 In a factory, 120 workers get an average wage of birr 30 a day, 160 workers get Birr 50 a day, 80 workers get Birr 60 a day and 40 workers get birr 80 a day. Find

i. the average of averages.
ii. the general average.

6.3.4 Weighted Arithmetic Mean

An item or value may be relatively more important or less important than other items. This relative importance is technically known as weight. In case where the relative importance of the different items is not the same we compute weighted arithmetic mean.

If w_1, w_2, …, w_n are weights attached to the values x_1, x_2, … , x_n respectively, then the weighted AM is defined as

$$\bar{x}_w = \frac{x_1 w_1 + x_2 w_2 + \ldots + x_n w_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum wx}{\sum w}$$

Example: An auto ride costs Birr 5 for the first km, Birr 4 for the next 3kms and Birr 9 for each of the subsequent kms. Find the average cost per km for 10 kms.

Rate (Birr)	Distance (km) w	xw
5.00	1	5.00
4.00	3	12.00
9.00	6	54.00
	10	71.00

$$\bar{x}_w = \frac{\sum xw}{\sum w} = \frac{71.00}{10} = \underline{\underline{7.10 Birr}}$$

CYP 5 Given data

Designation	Class I Officer	Class II Officer	Subordinate Staff	Clerical Staff	Lower Staff
Monthly Salary (in birr) x	1500	800	500	250	100
Strength of Cadre. w	10	20	70	100	150

Calculate i. the simple arithmetic mean.

ii. the weighted arithmetic mean.

6.3.5 Correcting the Arithmetic Mean

Formula:

$$\text{Correct Mean} = \frac{\text{Wrong Sum} - \text{Wrong Entry} + \text{Correct Entry}}{\text{Total Number of Observations}}$$

Examples:

1. The average mark of 100 students was found to be 40 but latter it is discovered that a score of 33 was misread as 83. Find the correct average corresponding to the correct sum.

   $\left.\begin{matrix}\bar{x} = 40 \\ N = 100\end{matrix}\right\} \rightarrow \sum x_i = \bar{x}N = 40(100) = 4000$ wrong sum

   Wrong Entry = 83

   Correct Entry = 33

   $$Correct Mean = \frac{4000 - 83 + 33}{100} = \frac{3950}{100} = \underline{\underline{39.5}}$$

2. The average of a class having 35 pupils is 14 years. When the age of the class teacher is added to the sum of the ages of the pupils, the average rises by 0.5 year. What must be the age of the teacher?

   $\left.\begin{matrix}\bar{x} = 14 \\ N = 35\end{matrix}\right\} \rightarrow \sum x_i = 14(35) = 490$ ... Sum of ages of the pupils

   $\left.\begin{matrix}\bar{x} = 14.5 \\ N = 36\end{matrix}\right\} \rightarrow \sum x_i = 14.5(36) = 522$ ... Sum of ages of the pupils and the teacher

   $\therefore$ Age of the teacher is 522 - 490 = <u>32 years.</u>

3. Goals scored by a football team in successive matches are 5, 2, 4, 3, 6, 0, 4 and 6. What is the number of goals the team must score in the next match in order that the average comes to 4 goals per match?

   Total goal scored in 8 matches = 5 + 2 + 4 + 3 + 6 + 0 + 4 + 6 = 30

   Total goal scored in 9 matches = $\bar{x}$.N = $4 \times 9 = 36$

Hence the goals required in the 9th match to bring the average 4 = 36 – 30 = 6

CYP6 The mean of 200 items is 50. Later on it is discovered that two items were wrongly taken as 92 and 8 instead of 192 and 88. Find out the correct mean.

CYP7 The average rainfall for a week, excluding Sundays, was 10cm. Due to heavy rainfall on Sunday, the average rainfall for the week rose to 15cm. How much rain fall was there on Sunday?

6.4 GEOMETRIC MEAN (GM)

6.4.0 DEFINITION

Geometric mean is defined as the n^{th} root of the product of n items or values of a series. If there are two items, we take square root; if there are three items, the cube root and so on.

Symbolically, let $x_1, x_2, \ldots, x_n$ be the n values of a variable x, then their G.M is defined as

$$G.M = \sqrt[n]{x_1 . x_2 \ldots x_n}$$

If the number of observation is more than three or more, the computation of the n^{th} root is very tedious. To simplify computation, the logarithms are used. In terms of log.

$$Log G.M = Log \sqrt[n]{x_1 \,.\, x_2 \,.\, x_3 \ldots \, x_n}$$

$$= Log (x_1 \,.\, x_2 \ldots \, x_n)^{\frac{1}{n}}$$

$$= \frac{1}{n} \,.\, Log (x_1 \,.\, x_2 \ldots \, x_n)$$

$$= \frac{1}{n} \left[Log\, x_1 + Log\, x_2 + \ldots + Log\, x_n \right]$$

$$= \frac{1}{n} \,.\, \sum_{i=1}^{n} Log x_i$$

$$Antilog(LogGM) = Antilog \left[\frac{1}{n} \,.\, \sum_{i=1}^{n} Log x_i \right]$$

$$GM = Antilog \left[\frac{1}{n} \sum_{i=1}^{n} Log x_i \right]$$

6.4.1 Computation of Gm for Ungrouped and Grouped Data

For ungrouped data: $G.M = Antilog \left[\frac{1}{n} \sum_{i=1}^{n} Log x_i \right]$

For grouped data: $G.M = Antilog \left[\frac{1}{n} \sum_{i=1}^{n} f_i \,.\, Log x_i \right]$

Example: Compute GM of the following data.

i. **x** : 23 27 54 35 50

ii. **x**: 10 16 22 28 34
f: 5 4 3 6 2

iii. **Classes**: 30 – 40 40 – 50 50 – 60 60 – 70
fi 5 8 4 3

Solution:

i.

x:	23	27	54	35	50
Log x	1.3617	1.4314	1.7324	1.5441	1.6990

$\sum_{i=1}^{5} Log x_i = 9.402$ n = 5

$G.M = Anti\log\left[\frac{1}{5}(9.4021)\right] = Anti\log(1.5670) = \underline{\underline{369}}$

ii.

x_i	10	16	22	28	34
f_i	5	4	3	6	2

Log x_i	1	1.2041	1.3424	1.4472	1.5315
fi log x_i	5	4.8164	4.0272	8.6532	3.0630

$\sum_{i=1}^{20} f_i Log x_i = 25559$

$G.M = Anti\log\left[\frac{1}{20}(25598)\right] = Anti\log(1.2780) = \underline{\underline{186}}$

iii. **Classes**: 30 – 40 40 – 50 50 – 60 60 – 70

fi : 5 8 4 3

CMi : 35 45 55 65

Log CMi : 1.5441 1.6532 1.7401 1.8129

fi Log CMi : 7.7200 13.2256 6.9612 5.4387

$$G.M = Antilog\left[\frac{1}{20}(33345)\right] = Antilog(1.6673) = \underline{\underline{4581}}$$

CYP8 Calculate GM for the following data.

i. **x**: 8 40 175 1209 2000

ii. **x**: 2 3 4 5 6
f: 5 7 8 3 2

iii. **Classes**: 0 – 10 10 – 20 20 – 30 30 – 40 40 – 50 50 – 60
fi : 2 5 6 18 13 6

6.5 HARMONIC MEAN (H.M)

Definition:

Harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the given observations.

Symbolically, if $x_1, x_2, \ldots, x_n$ are n variables of x then their HM is given by:

$$H.M = \frac{1}{\frac{1}{n}\left[\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}\right]} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

6.5.1 Computations of Harmonic Mean for Ungrouped and Grouped Data

For ungrouped data : $H.M = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$ **For grouped data**:

$$H.M = \frac{n}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

Example: Find the harmonic mean of the following data

i. **x** : 20 30

ii.

x	2	3	4	5	6
f	5	7	8	3	2

iii.

Classes	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54
fi	11	18	32	37	21	47	13

Solution:

i. 20 30 $H.M = \dfrac{2}{\frac{1}{20} + \frac{1}{30}} = \dfrac{2}{\frac{5}{60}} = \dfrac{120}{5} = \underline{\underline{24}}$

ii.

x	2	3	4	5	6
f	5	7	8	3	2
f/x	2.5	2.33	2	0.6	0.33

$\sum_{i=1}^{25} \frac{f_i}{x_i} = 7.76$ $H.M = \dfrac{25}{7.76} = \underline{\underline{3.22}}$

iii.

Classes	20 - 24	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54
fi	11	18	32	37	21	47	13
Cm_i	22	27	32	37	42	47	52
f_i/Cm_i	0.5	0.67	1	1	0.5	1	0.25

$$\sum_{i=1}^{179} \frac{f_i}{CM_i} = 4.92 \qquad H.M = \frac{179}{4.92} = \underline{\underline{36.38}}$$

CYP9 Find the H.M of the following data

i. **x**: 10 25 12 8 5

ii.

Marks	40	50	60	70
No. of Students	20	30	50	10

iii.

Classes	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
fi	4	6	10	12	5	3

6.6 ADVANTAGES AND APPLICATION AREAS OF THE THREE MEANS

6.6.1 Advantages:

All are i. rigidly defined.

ii. based on all the observations.

iii. suitable for further mathematical tea.

AM iv. easy to calculate and understand.

v. is least affected by fluctuations of sampling compared to other averages.

GM iv. it gives highest weightage to smaller values and smaller weightage to large values.

v. it is a proper average to measure the relative change (like percentage increase in Population, sales over a period of time, etc.

HM iv. is not affected very much by fluctuation of sampling.

v. is particularly useful in averaging speed, special types of rates and ratios where time factor is involved.

vi. since the reciprocals of the variables are involved, it gives greater weightage to smaller values.

6.6.2 Application Problems

1. Prove that
   i. AM = GM = HM if all the values are equal in a series.
   ii. HM < GM < AM if the values are different in a series.

Solution:

i. Suppose there are two items x and y in the series

If x = y = 7, then

$$AM = \frac{x + y}{2}\ ,\ GM = \sqrt{xy}\ ,\ HM = \frac{2}{\frac{1}{x} + \frac{1}{y}}$$

$$AM = \frac{7 + 7}{2} = 7\ ,\ GM = \sqrt{7 \times 7} = 7\ ,\ HM = \frac{2}{\frac{1}{7} + \frac{1}{7}} = \frac{2 \times 7}{2} = 7$$

Therefore, AM = GM = HM

ii. Suppose there are two items x and y in the series

Then $AM = \frac{x + y}{2}$, $GM = \sqrt{xy}$

If $x \neq y$, then $x - y > 0$

$\sqrt{x} - \sqrt{y} > 0$

$(\sqrt{x} - \sqrt{y})^2 > 0$

$x + y - 2\sqrt{xy} > 0$

$x + y > 2\sqrt{xy}$

$\frac{x + y}{2} > \sqrt{xy}$

$\underline{\underline{AM > GM}}$

Consider $\frac{x + y}{2} > \sqrt{xy}$ This is proved above.

$\frac{\sqrt{xy}}{2} > \frac{xy}{x + y}$ by multiplying in both sides by $\frac{\sqrt{xy}}{x + y}$ we get

$$\sqrt{xy} > \frac{2xy}{x + y} = 2 \cdot \frac{xy}{x + y} = 2 \cdot \frac{1}{\frac{x + y}{xy}} = \frac{2}{\frac{x}{xy} + \frac{y}{xy}} = \frac{2}{\frac{1}{y} + \frac{1}{x}}$$

$$\therefore \sqrt{xy} > \frac{2}{\frac{1}{x} + \frac{1}{y}} \Rightarrow \underline{\underline{GM > HM}}$$

Therefore, HM < GM < AM

Note- We can have the following relationship between the three means.

$$AM \cdot HM = \frac{x + y}{2} \left(\frac{2xy}{x + y} \right) = xy$$

To equalize AM . HM to GM, we put AM . HM under square root

$$GM = \sqrt{xy} = \sqrt{\frac{x + y}{2} \cdot \left(\frac{2xy}{x + y} \right)} = \sqrt{AM \cdot HM}$$

$\therefore GM = \sqrt{AM \cdot HM}$ If there are only two positive observations in the series

2. The price of a commodity increased by 5% from 1979 to 1980, by 9% from 1980 to 1981 and by 73% from 1981 to 1982. The average

increase from 1979 to 1982 is quoted as 25.6% and not 29%. Verify.

Solution:

Year	Price at the end of the year taking preceding as 100%. (X)	Log X
1980	100 + 5 = 105	2.0212
1981	100 + 9 = 109	2.0374
1982	100 + 73 = 173	2.2380
		6.2966

AM = $\frac{5 + 9 + 73}{3} = \frac{87}{3} = 29$

GM = Antilog [1/3(6.2966)] = Antilog(2.0989) = 125.6

Therefore, Rise in price is 125.6 - 100 = 25.6%

Verification:

Year	**Rise**	**Price would be**	**Growth 25.6%**	**Growth 29%**
1979		100	100	100
1980	5%	105	125.6	129
1981	9%	114.45	157.75	166.41
1982	73%	198	198	214.67

Thus GM is the best average to give us the true rise in price.

3. World Population has increased from 5 billion to 6 billion within 12 years. Calculate the average increment per year.

Solution:

The average annual increase is computed by applying the formula

$$P_n = P_o(1 + r)^n \qquad \text{or} \qquad r = \sqrt[n]{P_n / P_0} - 1$$

Where P_n - the amount at the end of the n^{th} period

P_o - the amount at the beginning of the period

n - time (years)

r - rate of change

$P_n = 6, \quad P_o = 5, \quad n = 12 \qquad r = ?$

$$r = \sqrt[12]{6/5} - 1 = 1.01 - 1 = 0.01$$

$\therefore$The average increment per annum = 1%

Therefore, GM is used in determination of average percentage of change in amount.

4. A machine depreciates by 40% in the first year, by 25% in the second year and by 10% per annum for the next three years. Each percentage being calculated on the diminishing value, what is the average percentage of depreciation for the entire period?

Solution:

Depreciation (%)	**After depreciation (%) = X**	**Log X**
40%	60%	1.7782
25%	75%	1.8751
10%	90%	1.9542
10%	90%	1.9542

10%	90%	1.9542
		9.5159

GM = Antilog [1/5(9.5159)]

= Antilog (1.9032)

≈ 81

∴Rate of depreciation per annum is 100 - 81 = 19%

5. The weighted GM of 5 numbers 10, 15, 25, 12 and 20 is 17.15. If the weights of the first four numbers are 2, 3, 5 and 2 respectively, find out the weight of the fifth number.

Solution:

X	**W**	**Log X**	**(LogX).W**
10	2	1.0000	2.0000
15	3	1.1761	3.5283
25	5	1.3979	6.9895
12	2	1.0792	2.1584
20	x	1.3010	1.3010(x)

$$\text{Log}17.15 = \frac{14.6762 + 1.3010.x}{12 + x}$$

$$1.2343 = \frac{14.6762 + 1.3010.x}{12 + x}$$

$$-0.0667x = -0.1354$$

$$x = 2.03$$

∴The missing weight is 2.

6. A cyclist pedals from his house to his college at a speed of 8 kmph and back from the college to home at 12 kmph. Find the average speed.

Solution:

Let the distance between the house and the college be x kms. Then the distance from house to college is covered in x/8 hrs and from college to house in x/12 hrs.

And the total distance = 2 x (house to college and back) is covered in (x/8 + x/12)hrs.

$\therefore$Average Speed = $\frac{\text{Total distance traveled}}{\text{Time taken}}$

$$= \frac{2x}{x/8 + x/12} = \frac{2x}{5x/24} = \frac{48x}{5x} = 9.60\text{kmph}$$

7. Mr. Raga traveled a distance of 900 kms by train at an average speed of 60 kmph, 200 km by boat at speed of 20 kmph, 1000 km by plane at 800 kmph speed and finally 4 km by taxi at 25 kmph speed. What is the average speed for the entire distance?

Solution:

X	**W**	**X/W**
60	900	15.00
20	200	10.00
800	1000	1.25
25	4	0.16
	2104	26.41

Weighted HM = $\frac{\Sigma W}{\Sigma W/X}$

$$= \frac{2104}{} = 79.67 \text{ kmph.}$$

26.41

CYP 10 If the arithmetic mean and the geometric mean of two items is 12.5 and 10 respectively, then

i. find the HM of the two items.
ii. find the value of the two items.

CYP 11 A motorist travels at a uniform speed of 20 kmph, 60 kmph and 30 kmph from A to B, B to C and C to D respectively. Find the average speed.

CYP 12 In a factory, a unit of work is completed by A in 5 minutes, by B in 7 minutes, by C in 4 minutes, by D in 8 minutes and by E in 6 minutes.

i. What is their average rate of work?
ii. What is the average number of units of work completed per minute?
iii. At this rate, how many units of work will they complete in six hours a day?

CYP 13 Find the average rate of increase in Population which in the first decade had increased by 20%, in the next by 30% and in the third by 40%.

6.7. SUMMARY

Arithmetic mean is mostly used in practice of all areas because its characteristics value being represented to all items in the variable.

Geometric mean is widely used in averaging ratios and percentages and in computing average rates of increase or decrease.

Harmonic mean is useful in comparing the values of a variable with constant quantity of another variable, i.e. time, rate, distance covered, quantities purchased or sold per unit etc.

6.8. MODEL EXAMINATION QUESTIONS

1. Given data:

Income ('000)	Below 10	10-30	30-60	60-100	Above 100
# of Students	5	10	15	8	2

   Calculate the mean income.

2. Calculate the number of shops corresponding to class interval 30 - 40 of the following distribution if the mean profit is 28.

Profit per Shop	0-10	10-20	20-30	30-40	40-50	50-60
# of Shops	12	18	27	f_4	17	6

3. Find the class intervals if the AM of the following distribution is 30.1 and assumed mean is 31.5.

Step deviations (d')	-3	-2	-1	0	1	2
Frequency (f_i)	5	10	25	30	20	10

4. The mean weight of 150 students of a class is 60kgs. The mean weight of boys is 70kgs and that of girls is 55kgs. Find the number of boys and girls in the class.

5. The price of a commodity increased by 20% in 1989, decreased by 12% in 1990 and increased by 15% in 1991. Calculate the average annual change in price.

6. If the price of a commodity triples in a period of 6 years, what is the average percentage increase per year?

7. A train runs the first 40 kms at a speed of 60 kmph, the next 60 kms at a speed of 80 kmph and the last 80 kms at a speed of 100 kmph. What is the average speed of the train for the whole journey?

8. If the GM of two positive observations is 2/3 of their AM and the sum of the two observations is 18, then
   i. their HM is _____
   ii. the two observations are _____ and _____

6.9. ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

CYP 1 i. 15 ii. 161 iii. 128 iv. 518 v. 15

CYP 2 i. 43.8 ii. 68 iii. 22.67

CYP 3 i. 20 ii. 0

CYP 4 i. 55 ii. 49

CYP 5 i. Br 630 iii. Br 302.86

CYP 6 50.9

CYP 7 45cm.

CYP 8 i. 168.4 ii. 3.347 iii. 32.20

CYP 9 i. 9.12 ii. 52.98 iii. 33.44

CYP 10 i. 8 ii. 5 and 20

CYP 11 30 kmph.

CYP 12 i. 5.65 minutes. ii. 0.177 units of work / minute. iii. 63.72 units of work.

CYP 13 i. 29.7%

6.10. GLOSSARY

Assumed mean - Refers to an estimated or approximate value for the arithmetic mean or average that is used to simplify its calculation. The nearer it is to the mean, the smaller are the numbers involved.

Class Interval - The range of interval between the highest and lowest values allowed in a particular class.

Depreciate - Make or become less in value (being diminished in value).

Deviation - Refers to the difference between a value of a variable and the mean of its distribution.

Rate - Refers to standard of reckoning, obtained by bringing two numbers or amounts into relationship like a period of time and a number of people, Currencies, Tax, etc.

Time Series - Refers to a set of values of a variable recorded over a period of time.

6.11. REFERENCES

- Business Statistics, C.R. REEDY. M Com Ph. D., 1994
- Practical Business Statistics. T .K. Nag pal, P.S. Narayana. , 1988
- Business Statistics, Dr. J. S. CHANDAN, Prof. Jadjit Singh, Ph D. (USA), 1996

UNIT 7: POSITIONAL MEASURES OF CENTRAL TENDENCY

CONTENTS

7.0 Aims and Objectives
7.1 Introduction
7.2 Mode
7.3 Median
7.4 Quartiles, Deciles and Percentiles and Grouped Data
7.5 Summary
7.6 Model Examination Questions
7.7 Answers to Check Your Progress Questions
7.8 Glossary
7.9 References

7.0 AIMS AND OBJECTIVES

At the end of this unit, the students will be able to

- ☞ Define and state the importance of mode, median, quartiles, deciles and percentiles.
- ☞ Calculate the modal value, median, quartiles, deciles and percentiles of a certain distribution and interpret.

7.1 INTRODUCTION

The mode and median are called positional measures of central tendency. The term position refers to the place of a value in the series. The values being divided by a number of equal parts are called partition values. Besides median, which divides a series in to equal parts, the quartiles, deciles and percentiles are important measures.

7.2 MODE

A value, which occurs most frequently in a series of observations, is called Mode. So by looking the observations mode can be identified.

It is the value, which has the greatest frequency density in its immediate neighborhood.

Importance:

1. Mode can be used as a central location for qualitative as well as quantitative data, like the median. Example, if a beauty measurement turns in to three impressions or scores, which we rate ‘very beautiful’, ‘beautiful’ and ‘not beautiful’, then the modal value is beautiful.
2. Like the mean, the mode is not affected by extreme values.

3. Mode can be used when one or more of the classes are open-ended.

Computation of Mode for Ungrouped and Grouped Data

For ungrouped data: Mode $(\hat{x})$ = that value in the data set, which occurs most often.

For grouped data: Discrete Series: Mode $(\hat{x})$ = the value of the variable corresponding to the maximum frequency.

Continuous Series: The class corresponding to the maximum frequency is called the modal class. The value of mode is obtained by the following interpolation formula.

$$Mode(\hat{x}) = l + \left[\frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)}\right] \times c$$

or

$$Mode\hat{x} = l + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) c$$

Where l – LCB of the modal class
f_1 – maximum frequency
f_0 – frequency preceding f1
f_2 – frequency succeeding f_1
f_2 – frequency succeeding f_1
C – magnitude of the class
$\Delta_1 = f_1 - f_0$
$\Delta_2 = f_1 - f_2$

Example: Find the value of mode for the following data

i. 25, 15, 23, 40, 27, 25, 23, 25, 20, 19, 22, 24, 25

ii.

x	10	20	30	40	50	60
f	4	9	16	25	22	15

iii.

Classes	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 - 79
fi	328	350	720	664	598	524	378	244

Solution:

i. Mode = value which occurs most often

Mode = 25

ii. Mode = Value of the variable with maximum frequency

Mode = 40

iii. Modal Class = 19.5 - 29.5

$l = 19.5$ $f_0 = 350$ $f_1 = 720$ $f_2 = 664$ c = 10

$$Mode(\hat{x}) = l + \left[\frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)}\right] \times c$$

$$= 195 + \left[\frac{720 - 350}{(720 - 350) + (720 - 664)}\right] \times 10 = 195 + \frac{3700}{426} = \underline{\underline{28185}}$$

CYP 14 Find the modal value of the following data.

i. 27, 33, 42, 25, 23, 27, 25, 33, 27, 28, 16, 18

ii.

Height (in inches)	58	60	61	62	63	65	68	70
No. of Persons	4	6	10	8	20	24	9	5

iii.

Classes	0 - 400	400 - 800	800 - 1200	1200 - 1600	1600 - 2000
fi	4	12	40	41	27

7.3. MEDIAN

The median is that value of the variable, which divides the group in to two equal parts, one part comprising all the values greater and the other all the values less than median. Or median can be defined as the middle value of a set of data values when they are arranged in ascending or descending order.

Importance:

- In dealing with qualitative data, median is more suitable average
- Median is recommended if the distribution has unequal classes, since it is simple to compute than the mean.
- Median is especially useful incase of open-ended classes since it is only positional and not calculated average.
- The magnitudes of extreme deviations do not influence the median.

Commutation of Median for Ungrouped and Grouped Data

For ungrouped data:

First, rearrange the values in the order of magnitude.
Then apply the following formula.

$$\text{Median}(\tilde{x}) = \text{value of the}\left(\frac{N+1}{2}\right)^{th}\text{item (where } n \text{ is odd)}$$

$$= \frac{x_{n+1}}{2}$$

$$\text{Median}(\tilde{x}) = \frac{1}{2}\left[\text{Value of}\left(\frac{N}{2}\right)^{th}\text{item} + \text{Value of}\left(\frac{N}{2}+1\right)^{th}\text{item}\right] \text{ (Where } n \text{ is even)}$$

$$= \frac{1}{2}\left[x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right]$$

For grouped data:

Discrete Series:

1. Compute the < cf_i column.
2. Search for the value of $\frac{N+1}{2}$, if not available, consider the value just greater than it, in the column of < cf_i.

Continuous Series:

1. Compute the < cf_i column.
2. Search for the value of $\frac{N}{2}$ if not available, consider the value just greater than it, in the column of < cf_i.
3. Then the following interpolation formula is used to calculate the median.

$$\text{Median}(\tilde{x}) = l + \frac{c}{f}\left(\frac{n}{2} - c.f\right)$$

where l - LCB of the median class

c - class interval of the median class

f - frequency of the median class

c.f - cumulative frequency jut less than $\frac{N}{2}$

Example: Find the median of the following data.

i. a) 27, 33, 42, 25, 23, 25, 33, 28, 27, 16, 18, 12

b) 8, 5, 2, 6, 15, 10, 25

ii.

x	4	6	8	10	12	14	16
f	2	4	5	3	2	4	1

iii.

x	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100	100 - 110
fi	20	21	50	40	53	16

Solution:

i. a. Rearranging:

12 16 18 23 25 25 27 27 28 33 33 42

n = 12 ... even

$$\tilde{x} = \frac{1}{2}\left[x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right] = \frac{1}{2}\left[x_6 + x_7\right] = \frac{1}{2}\left[25 + 27\right] = \underline{\underline{26}}$$

b. Rearranging: 2 5 6 8 10 15 25

n = 7 ... odd

$$\tilde{x} = \frac{x_{n+1}}{2} = x_4 = \underline{\underline{8}}$$

ii.

x	4	6	8	10	12	14	16
f	2	4	5	3	2	4	1
<*cfi*	2	6	11	14	16	20	21

n = 21 Median = The value of $\frac{N+1}{2}^{th}$ item

$= \left(\frac{21+1}{2}\right)^{th}$ item

= The value of the 11^{th} item

= 8

iii.

x	50 - 60	60 - 70	70 - 80	80 - 90	90 - 100	100 - 110
fi	20	21	50	40	53	16
<cfi	20	41	91	131	184	200

Median class = Value of

$\left(\frac{n}{2}\right)^{th}$ item $= 100^{th}$ item $= 80 - 90$

l = 80, c = 10, f = 40, c.f = 91

$$\text{Median}(\tilde{x}) = l + \frac{c}{f}\left(\frac{n}{2} - c.f\right) = 80 + \frac{10}{40}(100 - 91) = 80 + \frac{9}{4} = 82.25$$

CYP 15 Find the median of the following data.

i. a. 20 15 21 13 22 24 22 25 26 27 25

b. 120 125 112 137 129 127

ii.

x	28	30	32	34	36	38	40	42
f	14	15	16	24	16	10	6	4

iii.

x	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59
f_i	5	10	15	20	6	4

Determining the value of median Graphically:

Draw both the more than and less than ogives on the same graph. From the point of intersection of these two curves, draw a perpendicular line to the x – axis. The foot of the perpendicular line is the value of the median.

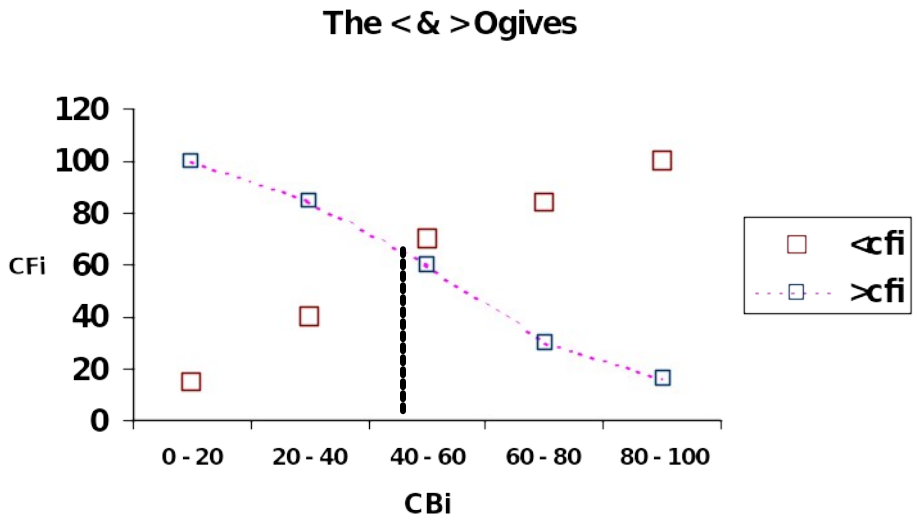
Examples: Find the median of the following frequency distribution graphically.

Classes	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
fi	15	25	30	14	16

Solution:

Classes	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100
fi	15	25	30	14	16
<cfi	15	40	70	84	100

>cfi	100	85	60	30	16



The perpendicular line drawn from the intersection point meets the x-axis approximately at 46. Therefore, the Median of the distribution is 46.

7.4 QUARTILES, DECILES AND PERCENTILES

Definitions:

Quartiles: are the three values, which divide the given data in to four equal parts. They are denoted by Q_1, Q_2 and Q_3.

Q_1 - The lower or first quartile. It covers 25% of the distribution.

Q_2 - The middle or second quartile. It covers 50% of the distribution.

Q_3 - The upper or third quartile. It covers 75% of the distribution.

Deciles: are the nine values, which divide the series in to ten equal parts. They are denoted by $D_1, D_2, \ldots, D_9$.

D_1 = Covers 10% of the distribution

D_2 = Covers 20% of the distribution

.

.

.

D_9 = Covers 90% of the distribution

Percentiles: are the 99 values, which divide the series in to 100 equal parts. They are denoted by $P_1, P_2, \ldots, P_{99}$.

Note that: i. $Q_1 = P_{25}$ $Q_2 = D_5 = P_{50}$ = median $Q_3 = P_{75}$

ii. $D_1 = P_{10}$, $D_2 = P_{20}$, $D_3 = P_{30}$, $\ldots, D_9 = P_{90}$.

Importance:

The quartiles are more widely used in Economics and Business while the deciles and percentiles are important in Psychology and Educational Statistics concerning grades, rates, ranks, etc. The working principle for computing the partition value is basically the same as that of computing the median.

Computation of Quartiles, Deciles and Percentiles for Ungrouped and Grouped Data

For ungrouped data and discrete series:

First, for ungrouped data, rearrange the values in the order of magnitude and for discrete series, compute the $<Cf_i$ column. Then apply the following formula.

$$Q_i = \text{value of } \frac{i(N+1)^{th}}{4} \text{ item}$$

$$D_i = \text{value of } \frac{i(N+1)^{th}}{10} \text{ item}$$

$$P_i = \text{value of } \frac{i(N+1)^{th}}{100} \text{ item}$$

For continuous series:

1. Compute the $<cf_i$ column.
2. Determine the quartile, decile or percentile class.
3. Apply the following interpolation formula.

$$Q_i = l + \frac{c}{f}\left(\frac{iN}{4} - c.f\right)$$

$$D_i = l + \frac{c}{f}\left(\frac{iN}{10} - c.f\right)$$

$$P_i = l + \frac{c}{f}\left(\frac{iN}{100} - c.f\right)$$

Example: For the data given below, compute the value of Quartiles, D_3, D_7, P_{15} and P_{88} and interpret.

Marks	Below 10	10 - 20	20 - 40	40 - 60	60 - 80	Above 80
No. of Students	10	15	25	30	14	6
$<cf_i$	10	25	50	80	94	100

Solution:

Q_1 – size of $\frac{N^{th}}{4}$ item = 25th item 10 – 20 quartile class

l = 10, c = 10, f = 15, c.f = 10

$$Q_1 = l + \frac{c}{f}\left(\frac{n}{4} - c.f\right) = 10 + \frac{10}{15}(25 - 10) = \underline{\underline{20}}$$

Mark of 25% of students is less than 20.

Q_2 – size of $\frac{2N^{th}}{4}$ item = 50th item 20 – 40 quartile class

l = 20, c = 20, f = 25, c.f = 25

$$Q_2 = l + \frac{c}{f}\left(\frac{n}{2} - c.f\right) = 20 + \frac{20}{25}(50 - 25) = \underline{\underline{40}}$$

Mark of half of students is below 40.

Q_3 – size of $\frac{3N}{4}^{th}$ item = 75th item 40 – 60 quartile class

l = 40, c = 20, f = 30, c.f = 50

$$Q_3 = l + \frac{c}{f}\left(\frac{3n}{4} - c.f\right) = 40 + \frac{20}{30}(75 - 25) = \underline{\underline{73.33}}$$

Mark of $\frac{3}{4}$th of students is below 73.33.

D_3 – size of $\frac{3N}{10}^{th}$ item = 30th item 20 – 40 decile class

L = 20, c = 20, f = 25, c.f = 25

$$D_3 = l + \frac{c}{f}\left(\frac{3n}{10} - c.f\right) = 20 + \frac{20}{25}(30 - 25) = \underline{\underline{24}}$$

Mark of 30% of students is below 24.

D_7 – size of $\frac{7N}{10}^{th}$ item = 70th item 40 – 60 decile class

L = 40, c = 20, f = 30, c.f = 50

$$D_7 = l + \frac{c}{f}\left(\frac{7n}{10} - c.f\right) = 40 + \frac{20}{30}(70 - 50) = \underline{\underline{53.33}}$$

Mark of 70% of students is below 53.33.

P_{15} – size of $\frac{15N}{100}^{th}$ item = 15th item10 – 20 percentile class

L = 10, c = 10, f = 15, c.f = 10

$$P_{15} = l + \frac{c}{f}\left(\frac{15n}{100} - c.f\right) = 10 + \frac{10}{15}(15 - 10) = \underline{\underline{13.3}}$$

Mark of 15% of students is below 13.3.

P_{88} – size of $\frac{88N^{th}}{10}$ item = 88th item 60 – 80 percentile class

L = 60, c = 20, f = 14, c.f = 80

$$P_{88} = l + \frac{c}{f}\left(\frac{88n}{100} - c.f\right) = 60 + \frac{20}{14}(88 - 80) = \underline{\underline{71.43}}$$

Mark of 88% of students is below 71.43.

CYP 16 Compute the value of Quartiles, D_4, P_{69} and interpret for the data given below.

i. 46 35 28 52 54 43 35 49 46 50 41

ii.

Daily Wages	40	45	50	55	60	65	70
No. of Workers	9	22	26	18	13	8	5

iii.

Rent in Birr	150-250	250-350	350-450	450-550	550-650	650-750	750-850	850-950
No. of Houses	8	10	15	25	40	20	15	7

7.5 SUMMARY

The arithmetic mean and median satisfy the conditions of definition and stability. Media has a distinct merit over mean insofar as easy calculations. Mode can be located just by inspection. In case, every value occurs the same number of times mode is useless measure. It is

observed that the median, quartiles, deciles and percentiles have good relation.

7.6 MODEL EXAMINATION QUESTIONS

1. In a class of 15 students, three failed in a test. The marks of 12 students who passed were 9, 6, 7, 8, 8, 9, 6, 8, 7, 5, 4 and 7. What was the median of all the 15 students?

2. Calculate the mode and median for the distribution of the weights in kgs of 150 people from the data given below.

Weight in Kgs	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
No. of People	18	37	45	27	15	8

3. For the data given below, find the missing frequencies if median is 37 and mode is 43 million birr.

Fund raised in millions of birr	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
No. of NGO's	3	F_2	16	20	F2	16

4. The following distribution shows the marks of 60 students in Economics. Calculate Q3, D5, P57 and the median.

Marks	31 - 39	41 - 49	51 - 59	61 - 69	71 - 79	81 - 89	91 - 99

No. of Students	12	10	12	9	6	7	4

5. For the following data Q1 is found to be 41. Find the missing frequency.

Classes	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59
f_i	8	10	f_3	20	12	25

7.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

CYP14 i. 27 ii. 65 iii. 1226.6667

CYP15 i. a. 22 b. 126 ii. 34 iii. 44.5

CYP16

i.	$Q_1 = 35$	$Q_2 - 46$	$Q_3 = 50$	$D_4 = 43$	$P_{69} = 50$
ii.	$Q_1 = 45$	$Q_2 = 50$	$Q_3 = 60$	$D_4 = 50$	$P_{69} = 55$
iii.	$Q_1 = 458$	$Q_2 = 580$	$Q_3 = 685$	$D_4 = 542$	$P_{69} = 646.5$

7.8 GLOSSARY

Interpolation - Refers to the process of calculating an unknown value of a variable that is between two or more known values in the series.

Open-ended Classes - A class that allows either the upper or lower end of a quantitative classification scheme to be limitless or indeterminate.

Percentage - Refers to the proportion or rate per hundred parts. Symbol is %. To write a number as a

percentage, all that is needed is to multiply it by 100.

Stability - Refers to quality of being firm or not likely to move or change.

7.9 REFERENCES

- Practical Business Statistics. T .K. Nag pal, P.S. Narayana. , 1988
- Statistics for Business and Economics. J. S. CHAN DAN., 1998
- Business Statistics [A Text Book for B. Com. Students of Indian Universities] R.H DHARESHWAR, M.Sc. M. Phil., 1999

UNIT 8: MEANING AND TYPES OF DISPERSION

CONTENTS

8.0 Aims and Objectives

8.1 Introduction
8.2 Meaning of Variation
8.3 Need for Measures of Dispersion
8.4 Objectives of Measure of Dispersion
8.5 Characteristics of Good Measures of Dispersion
8.6 Types of Measures of Dispersion
8.7 Absolute and Relative Measures of Dispersion
8.8 Summary
8.9 Answers to Check Your Progress Questions (CYP)
8.10 Model Examination Questions
8.11 Glossary
8.12 References

8.0 AIMS AND OBJECTIVES

The aim of this unit is to present the meaning, need, objectives, characteristics and types of dispersion.

At the end of this unit, you will be able to:

- ☞ explain the meaning of depression
- ☞ understand why we need measures of dispersion
- ☞ describe the objectives of measures of dispersion
- ☞ recognize the types of dispersion
- ☞ distinguish between absolute and relative measures of dispersion.

8.1 INTRODUCTION

Measures of central tendency are not sufficient to tell the difference between two or more distributions, as there might be distributions with the same average. This unit focuses on the discussion of measures of variation or dispersion and their types.

8.2 MEANING OF DISPERSION

`Dispersion' or `variation' in statistics means the degree of spread of each individual item or value from the central value in the given distribution. According to Bowley it is "the measure of the variation of the items." In the words of Spiegel" the degree to which numerical data tend to spread about an average is called the variation or dispersion of the data". According to Berenson and Leuine" dispersion is the amount of variation, scatter, or spread in the data." The term dispersion indicates the extent the items in the data differ from one another. In other words, they measure the lack of uniformity in the distribution.

According to John I. Griffin " A measure of variation or dispersion describes the degree of scatter shown by observations and is usually measured as an average deviation about some central value". The object of measuring dispersion is to arrive at summary figure, which adequately demonstrates the extent of scattered ness of the variables in the distribution. The measures of dispersion are also called the average of second order, since they measure the average of the deviations taken from the central tendency of the distribution.

8.3 NEED FOR MEASUREMENT OF DISPERSION

The need for measurement of dispersion arises from the facts that,

a) The distribution may have similar averages, but the degree of scatterdness in the individual values is different, and

b) The distribution may have different averages, but the variation among the values in the distribution may be quite less. Hence, the difference between the averages is not important. It also arises from the fact that changes in the behavior of a group of values are often reflected in a change in variation rather than the change in central values. The need for measurement dispersion, when the averages are some but the variations in the individual items are different can better be illustrated with the following example.

Production of Teff (In Metric Tonnes Per hectar) in three fields.

Year	Field'A'(X1)	Field `B'(x2)	Field`C'(X3)
1984	8	6	2
1985	8	13	8
1986	8	8	30
1987	8	13	28
1988	8	15	10
1989	8	11	2
1990	8	7	0
1991	8	6	0
1992	8	0	0
1993	8	1	0
	$\Sigma X_1 = 80$	$80\Sigma X_2 = 80$	$\Sigma X_3 = 80$

Arithmetic mean of:

Field `A' = $\frac{\sum x_1}{n} = \frac{80}{10} = 8$

Field `B' = $\frac{\sum x_2}{n} = \frac{80}{10} = 8$

Field `C' = $\frac{\sum x_3}{n} = \frac{80}{10} = 8$

In this example, the average production of Teff in all the fields is the same. But yearly production figures differ widely from the average. In field `A' the average production and yearly values of production for all the years is the same. Hence, the average perfectly represents the values. In field `B' except in 1986, the production figures for all other years show variation from the average. In field `C' wide variations are observed in the production for all the years from the average. Hence, comparison among these three fields on the basis of their average would be misleading, because the degree of deviations of figures in field `B' and field `C' from their average are altogether different in each case. Therefore, measurement of these variations is necessary to make rational decisions. Sometimes, the average may be different, but the nature of variation may be the same. This can be illustrated with the help of the example below.

Production of three computer Assembly firms (in thousands)

Year / Firms	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
Yassem (X1)	6	10	14	18	22	26	30	34	38	42
Tianes (X2)	16	20	24	28	32	36	40	44	48	52
Winner Cap (X3)	150	154	158	162	166	170	174	178	182	186

Solution:

Year	**Firms**					
	Yassem	Variation	Tianes	Variation	WinnerCap	Variation
	X_1=24		X_2=34		X_3=168	
1984	6	-18	16	-18	150	-18
1885	10	-14	20	-14	154	-14

1986	14	-10	24	-10	158	-10
1987	18	-6	28	-6	162	-6
1988	22	-2	32	-2	166	-2
1989	26	+2	36	+2	170	+2
1990	30	+6	40	+6	174	+6
1991	34	+10	44	+10	178	+10
1992	38	+14	48	+41	182	+14
1993	42	+18	52	+18	186	+18
	$\Sigma x_1 = 240$	= 0	$\Sigma x_2 = 340$	=0	$\Sigma x_3 = 1680$	= 0

$$\overline{x_1} = \frac{\sum x_1}{n} = \frac{240}{10} = 24$$

$$\overline{x_2} = \frac{\sum x_2}{n} = \frac{340}{10} = 34$$

$$\overline{x_3} = \frac{\sum x_3}{n} = \frac{1680}{10} = 168$$

In the above example, the average production of the firm's yassem, Tianes, and winnerCap is 24,000 computers, 34,000 computers, and 168000 computers respectively. Thought the average differs with one another, variations are similar in nature. If the variation in time series is ignored, one may conclude that WinnerCap is efficient as it shows a higher average value. But it could be a hasty conclusion, as the variations in the production figure are occurring in the same manner for all the three assembly firms. Hence, the measurement of the variations is significant to arrive at appropriate conclusions.

8.4. OBJECTIVES OF MEASURES OF DISPERSION

Measures of dispersion are studied with the following objectives.

i) To determine the reliability of average an average found from a homogeneous set of observations is considered to be representative and reliable. Measure of dispersion help to study the representative ness and reliability of the average. When the dispersion is small, it can be considered that greater uniformity is ensured in the distribution and average is considered to be fairly representative and reliable. But greater value of dispersion indicates that the average is unreliable and not representative one. According to W.A. Spurr and C.P. bonini, "When dispersion is small, the average is a typical value in that it closely represents the individual values and it is reliable in that it is a good estimate of the corresponding average in the population. On the other hand, when the dispersion is great the average is not so typical and unless the sample is very large, the average may be unreliable".

ii) To analyze the nature and causes of dispersion: measurement of variation is an effective technique for analyzing the nature of variation and detecting the cause for such variation in the given set of observations. It also helps to determine whether the variation is

due to random causes or due to assignable causes in the original values. Thus it serves as a mechanism to control variations that arise due to assignable reasons.

iii) To compare the variability among the distributions: Two or more sets of observations are comparable when they are consistent. Measures of dispersion serve as an effective tool for comparison with regard to variability among the distributions. A high degree of dispersion shows lack of uniformity and low degree of dispersion accounts for more uniformity and consistency. If, for example, the prices of a commodity over a period of time are to be compared, fewer variations in the prices denote more uniformity and representative ness and Vice-Versa.

iv) To facilitate the use of other statistical measures: Variation or dispersion is very useful in the application of advanced statistical techniques like correlation, regression, analysis of time series, tests of significance etc. It is also useful in determining the effective production control technique.

8.5 CHARACTERISTICS OF A GOOD MEASURE OF DISPERSION

Since measurement of dispersion occupies an important place in various statistical investigations, it is necessary to use an appropriate and satisfactory measure of dispersion. For this it should possess the following characteristics.

i) it shows be simple to understand and easy to calculate: A good measure of dispersion should be simple to understand and should not require lengthy computations.

ii) It must be rigidly defined: It should be rigidly defined so as to ensure the same interpretations at all times and places. This avoids confusion and ambiguity relating to its meaning. If it is not so, it may be influenced by the personal bias of the statistician. As far

possible it must be algebraically defined to arrive at uniform computation.

iii) It should be based on all the values in the distribution: Any measure of dispersion is considered to be fully representative, when its computations are based on all the values. Further, it should not be affected by the extreme values in the distribution. If small or large items influence the value of dispersion, its usefulness is lessened.

iv) It must be suitable for further algebraic treatment: Since the measures of dispersion are used in various fields to facilitate the application of advanced statistical cost and production control techniques, it must be amenable to further computations

v) It must possess sampling stability: Any value of dispersion which is not affected by the variation in sampling is considered to be a satisfactory measure. In other words, the values would be uniform for all the samples drawn from same population.

CYP 1

List out the characteristics of good measures of dispersion.

_____

_____

_____

_____

8.6 TYPES OF MEASURES OF DISPERSION

There are two types of dispersion. They are (i) distance measures or positional measures and (ii) measures of deviations. The distance measures study the dispersion in terms of distance between the value of selected observations. The measures of deviations are employed to study the variations from the respective measures of

central tendency. The two types of dispersions-distance and deviation can be studied by

a) Numerical methods and
b) Graphical methods. Numerical methods explain the dispersion in numbers.

They include (i) Range (ii) Quartile deviation,(iii) Mean deviation (iv) standard deviation. Graphic method presents the dispersion by graphic means. The important graphic method used to show the dispersion is Lorenz curve.

CYP 2

Write the different types of dispersion.

8.7 ABSOLUTE AND RELATIVE MEASURES OF DISPERSION

Absolute measures of dispersion are expressed in the same statistical unit in which the original values in the distribution are expressed. This useful to assess the absolute magnitude of variability. It is used to compare two or more distributions, which are expressed in homogenous units.

Relative measures of dispersion also known as `Coefficients' are considered to be pure numbers, independent of the units of measurement. These measurements are expressed as ratio of absolute dispersions to the appropriate average used for the measurement of variation. They can also be expressed in terms of percentages. Relative measures are used in comparing the distributions expressed in different statistical units of measurement and the relative accuracy of the data.

Detailed discussion of the methods of dispersion is dealt in the subsequent sections.

8.8 SUMMARY

Dispersion is the degree of variation of individual items in the distribution from the central value. The techniques used to measure such variations are called measures of dispersion or variation. They help to determine the reliability of an average and compare the variability among distributions. Dispersions studied with the help of measures like Range. Interquartile range, Quartile deviation, Mean deviation, Standard deviation and Lorenz curve. Dispersion can be measured both in absolute and relative terms (Coefficients) While an absolute value is used to compare the distributions expressed in homogeneous statistical units, the relative measures are used to compare distributions expressed in different statistical units. A low value of a measure of dispersion indicates uniformity and consistency of data and Vice-versa.

8.9 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

CYP 1

The characteristics of good measures of dispersion are:

(i) It should be simple to understand and easy to compute

(ii) It must be rigidly defined

(iii) It should be based on all the values of distribution

(iv) It should be amenable for further algebraic treatment

(v) It must possess the sampling stability.

CYP 2

There are two types of dispersion. They are (i) distance measures or positional measures and (ii) measures of deviations. The distance measures study the dispersion in terms of distance between the value of selected observations. The measures of deviations are

employed to study the variations from the respective measures of central tendency. The two types of dispersions-distance and deviation can be studied by

c) Numerical methods and
d) Graphical methods. Numerical methods explain the dispersion in numbers.

They include (i) Range (ii) Quartile deviation, (iii) Mean deviation, (iv) Standard Deviation. Graphic method presents the dispersion by graphic means. The important graphic method used to show the dispersion is Lorenz curve.

8.10 MODEL EXAMINATION QUESTIONS

Short Answer questions

1. Define dispersion
2. What is relative measure of dispersion
3. What is absolute measure of dispersion
4. Explain the need for measure of dispersion
5. What are the objectives of measure of dispersion
6. Explain the characteristics of dispersion
7. What are the types of measures of dispersion?

Essay Questions

1. What are absolute and relative measures of dispersion? Explain their importance.
2. What are the characteristics of good measure of dispersion? Take any two measures of dispersion an evaluate them in accordance with these characteristics

8.11 GLOSSARY

Absolute measures of dispersion: They express the dispersion in terms of original unit of data.

Dispersion or variation: It refers to the degree of spread of each individual item or value from the central value in a given distribution**.**

Measure of dispersion: A measure that used to express the scattered ness or homogeneity of data is called measure of dispersion or variation.

Relative measures of dispersion: They are pure numbers, independent of unit of measurement. These are called coefficients.

8.12 REFERENCES

- Bersenso, Mark L, Basic Business statistics concepts and leuine, David M. and applications. New Jersey: prentice-hall, 1983.
- Mason,Rober D, and Lind, Douglas A. Statistical techniques in Business and Economics. U.S.A: IRWIN, 1996.

UNIT 9: RANGE AND QUARTILE DEVIATION

CONTENTS

9.0 Aims and Objectives
9.1 Introduction
9.2 Meaning of Range
9.3 Calculation of Range
9.4 Merits and Limitations of Range
9.5 Utility of Range
9.6 Inter-Quartile Deviation
9.7 Meaning of Quartile Deviation
9.8 Calculation of Quartile Deviation
9.9 Merit and Limitations of Quartile Deviation
9.10 Summary
9.11 Model Questions & Answers
9.12 Model Examination Questions
9.13 Glossary
9.14 References

9.0 AIMS AND OBJECTIVE

This section aims at explaining the range, methods of calculating it, advantages, disadvantages, and uses of range. It also deals with the meaning, uses, and limitation of Quartile deviation.
After going though this section, you should be able to: -

- ☞ define range
- ☞ compute range
- ☞ list the merits and limitations of range
- ☞ explain the utility of range
- ☞ define inter-Quartile range
- ☞ calculate inter-quartile range
- ☞ define Quartile deviation
- ☞ calculate quartile deviation
- ☞ list merits and limitation of quartile deviation

9.1 INTRODUCTION

We have thoroughly discussed measure of dispersion in section 5.0. In this section we describe two of the measures of dispersion (i.e Range and quartile deviation).

Range, which measures the different between the maximum and the minimum values in the distribution, is the simplest measure of dispersion. Unlike range, quartile deviation takes into account the middle 50% of items, which are delimited by third quartile (Q3) form above and first quartile Q1 from below. Then the difference between Q3 and Q1 is called the Inter-quartile range and when Inter-quartile range is divided by two, it is called quartile deviation.

9.2 MEANING OF RANGE

Range (R) can be defined as the absolute difference between the maximum value (Xmax) and the minimum value (X min) in a given distribution.

Symbolically

$R = X_{max} - X_{min}$ or $R = L - S$

Where, R = Range

$L = X_{max}$ = maximum value or largest value

$S = X_{min}$ = minimum or smallest value.

Range is an about measure of dispersion. Such measures are compared if the unit of measurement is homogeneous.(i.e if all the sets of distribution are expressed in the same statistical unit such as birr, letter, meters, etc. But if different sets of distributions are expressed in different statistical units, the absolute measure of range can not be effectively compared .For ensuring comparability, relative measure of range, known as coefficient of range is used. This is obtained by dividing the absolute range by the sum of the largest and smallest values in the distribution. Its calculation follows the formula.

Coefficient of Range $\frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{L - S}{L + S}$

If the distributions have similar averages, smaller range represents uniformity and consistency in the distribution. It means that the average is representative in the distribution. On the other hand, large value of range shows lack of uniformity and consistency in the distribution. It explains that the average is inadequate and not representative.

9.3 CALCULATION OF RANGE

Indiscrete series, range is calculated by

i) Locating the largest and smallest values in the series

ii) Obtaining the range by calculating the difference (L - S)

R = L - S

iii) Calculating the coefficient of range by dividing the range by the sum of the smallest and largest value in the series. In discrete series, the frequencies are to be ignored.

Example 1 The following are times in which two Athletes finished a race.

Find out which athlete is consistent in the race.

Athlete A	6	8	12	25	15	16	18	3
Athlete B	6	22	28	30	17	10	14	

Solution: Athlete - A

$X_{max} = 25, \quad X_{min} = 3$

Hence,

Range = $X_{max} - X_{min}$

Range = 25 – 3 = 22

Coefficient of Range = $\frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{25 - 3}{25 + 3} = \frac{22}{28} = 0.786$

Athelete - B

$X_{max} = 30, \quad X_{min} = 6$

Coefficient of Range = $\frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{30 - 6}{30 + 6} = \frac{24}{36} = 0.667$

Since the variation in time of Athlete-B in terms of Coefficient of smaller than that of Athlete–A, we can conclude that groups B is more consistent.

Discrete Series

Example2: Calculate range and its coefficient from the following data relating to Salaries of 400 Lectures in Unity University.

Salary (In birr)	1800	2000	2400	2800	3000	3200	3400	4000	4500	5000	5400	7000
Number of Lectures	20	25	26	27	30	30	20	40	24	24	94	40

Solution:

Range $= X_{max} - X_{min}$

$= 7000 - 1800$

$= \underline{5200}$ Birr

Coefficient of Range =

$$\frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{7000 - 1800}{7000 + 1800} = \frac{5200}{8800}\ birr = \underline{0.591}$$

Coefficient of range is high in the distribution. Hence it is more variable and less consistent.

Continuous series

Example 3: Calculate range and its coefficient fro the mark obtained by distance students in statistics.

Mark	11-20	21-30	31-40	41-50	51-60	61-70
Number of Students	15	28	37	10	48	63

Solution: maximum mark is 70 and minimum mark is 11.

$X_{max} = 70$

$X_{min} = 11$

Range $= X_{max} - X_{min}$

$= 70 - 11$

$= \underline{\underline{59}}$

Coefficient of Rage $= \frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{70 - 11}{70 + 11} = \frac{59}{81} = \underline{\underline{0.728}}$

The coefficient of range of marks in statistics in 0.723, hence the marks are more vaiable. Sometimes, range may also be calculated from the mid-value of the classes. In such cases, first, mid values are obtained and then range is calculated by subtracting smallest the mid-value from the largest mid-value.

Example4 From the following distribution, compute range and its coefficient by taking themid-value

Class	5-12	13-20	21-28	29-36	37-44
Frequency	10	9	6	8	4

Solution: Calculation of range and coefficient of range.

Class	Frequency	Mid-value

5 - 12	10	$\frac{5+12}{2} = \frac{17}{2} = 8.5$
13 - 20	9	$\frac{13+20}{2} = \frac{33}{2} = 16.5$
21 - 28	6	$\frac{21+28}{2} = \frac{49}{2} = 24.5$
29 - 36	8	$\frac{29+36}{2} = \frac{65}{2} = 32.5$
37 - 44	4	$\frac{37+44}{2} = \frac{81}{2} = 40.5$

Range $X_{max} - X_{min}$

$X_{max} = 40.5$

$X_{min} = 8.5$

Range $= X_{max} - X_{min}$

$= 40.5 - 8.5$

$= \underline{\underline{32}}$

Coefficient of Range $\frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{40.5 - 8.5}{40.5 + 8.5} = \frac{32}{49} = \underline{\underline{0.653}}$

The coefficient of range is large, Hence, the distribution is more variable

9.4 MERITS AND LIMITATION OF RANGE

a) Merits

i) Range is the simplest method of studying dispersion
ii) Range can be easily calculated. It doesn't require ungthly calculations.

b) Limitations

i) Since range is computed by taking only two extreme value, it ignores all other values in the distribution
ii) Its value is affected by extreme items as it does not describe all other items.
iii) It is considered to be unreliable measure of dispersion, because changes in the value of range.
iv) Often, range is influenced by the fluctuations in sampling .Its computation is not practicable incase of grouped observations which include open-end classes.
v) Range is not amenable for algebraic treatment.

On account of these limitations, the utility of range is limited. In the words of W.O king, Range is too indefinite to be used as practical measure of dispersion.

9.5 UTILITY OF RANGE

Though range is a crude measure of dispersion, it has a specific significance due to its simplicity. It can be usefully applied in certain cases. For instance, it is most conveniently used in weather for casts, where identifying the extreme values are enough for making appropriate decisions. Usually, meteorological department makes weather forecasts on the bases of minimum and maximum limits of temperature, rainfall, etc. In industries, the quality of the products manufactured will be determined by studying a sample out of the total production. In such cases, the limit for variations will be fixed and test checks are conducted to see whether the quality is within the prescribed limits. If the variations or consistency in the pattern of

production exceeds the limits, proper care will be taken to inspect the equipment. Range is normally applied in constructing quality control Charts. Range also helps to understand the price behavior of stocks and shares in stock markets.

9.6 INTER-QUARTILE RANGE

The range, which takes into account only two extreme values, is considered a crude measure of dispersion. To overcome certain limitations of range another method known as `inter-quartile range' has been developed. Inter-quartile range is the absolute difference between the third quartile (Q_3) and the first quartile (Q_1) of a given frequency distribution. In other words, it includes only the middle 50% of items in the distributions and ignores one quarter of the observations on either end of the distribution. The inter-quartile range is calculated with the help of the formula.

Inter-quartile range (I.R) $= Q_3 - Q_1$

Where I.R = Inter-quartile range.

Q_3 = the third quartile

Q_1 = the first quartile

For example, if a frequency distribution Q_3 and Q_1 are 72 and 20 respectively, its inter-quartile range is

I.R = 72 – 20 = 52

Inter-quartile range is often used in the data pertaining to socio-economic problems and it studies the absolute difference between two quartiles.

9.7 MEANING OF QUARTILE DEVIATION

Another method of dispersion developed to overcome the limitations of range and inter-quartile deviation is quartile deviation or semi

inter-quartile-range. It can be defined as the average absolute difference between the third (upper) quartile and the first (lower) quartile of the frequency distribution. It studies the range of spread various items on either side of the median and it ignores nearly 50% of the items on either the extreme ends of the distribution. High degree of quartile deviation means low uniformity, and low degree of variation quartile deviation is calculated with the help of the following formula.

$$Q.D = \frac{Q_3 - Q_1}{2} \quad or \quad \frac{I.R}{2}$$

Where Q.D = Quartile deviation (Semi-Inter-quartile range)

Q_3 = third quartile (upper quartile)

Q_1 = first quartile (lower quartile)

Quartile deviation, which is used in comparing distributions in a homogenous statistical unit, is an absolute measure of dispersion. In comparing distributions expressed in different statistical units, however, we use a relative measure of dispersion called coefficient of quartile deviation.

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

9.8 CALCULATION OF QUARTILE DEVIATION

In the previous blocks, you are well acquainted with the procedures to calculate the third (upper) quartile and the first or lower quartile. Thus, we shall discuss few examples.

Example 1 The following is data corrected on the result of 11 student in a final examination of statistics.

Result: 72 65 75 69 35 43 52 37 61 58 41

Solution The data when arranged in ascending order will read as follows:

35 37 41 52 58 61 65 69 72 75

Calculating the quartiles,

Q_3 = value of the $3\left(\frac{N+1}{4}\right)^{th}$

Q_1 = Value of the $\left(\frac{N+1}{4}\right)^{th}$

Here N = 11

Q_3 = value of the $3\left(\frac{11+1}{4}\right)^{th}$

= Value of $3 \times \left(\frac{12}{4}\right)^{th}$ item

$\therefore$ Value of the 9^{th} item = 69

Q1 = the value of $\left(\frac{N+1}{4}\right)^{th}$ item

= Value of $\left(\frac{11+1}{4}\right)^{th}$ item

= Value of $\left(\frac{12}{4}\right)^{th}$ item

$\therefore$ Value of the 3^{rd} item = 41

Substituting 69 & 41 for the upper & lower quartiles respectively in the formula, we have:

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{69 - 41}{2} = \frac{28}{2} = 14$$

Coefficient of quartile deviation =

$$\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{69 - 41}{69 + 41} = \frac{28}{110} = 0.255$$

From the coefficient calculated above, it can be seen that there is a greater uniformity in the distribution because the coefficient of quartile deviation is small.

Example 2 Compute coefficient of quartile deviation form the following data:

Value	Frequency
5-8	6
9-12	10
13-16	18
17-20	30
21-24	15
25-28	12
29-32	10
33-36	6
37-46	4

Solution

Value	Frequency	Cumulative Frequency
5-8	6	6
9-12	10	16
13-16	18	34
17-20	30	64
21-24	15	79
25-28	12	91
29-32	10	101
33-36	6	107
37-40	4	111

Coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ value

$= \left(\frac{111+1}{4}\right)^{th}$ Value

$= \left(\frac{112}{4}\right)^{th}$ Value

$= (28)^{th}$ Value

Q_1 lies in class 13 – 16

Hence, $Q_1 = LQ_1 + \frac{\left(\frac{N}{4} - c.f\right)w}{f}$ $Q_3 = 3 \times$

$\frac{\left(\frac{N}{4} - cf\right)w}{f}$

$= 125 + \frac{(28 - 16)}{18} 4$ $= (3 \times 28)^{th}$

$= 125 + \frac{12 \times 4}{18}$ $= 84^{th}$ value

$= 125 + \frac{8}{3}$ Q_3 lies in class 25 - 28

$= 12.5 + 2.67$ hence $Q_3 = 245 + \frac{(84 - 79)}{12} 4$

$= \underline{15.17}$ $= 245 + \frac{5}{3}$

$= \underline{26.17}$

Coefficient of QD $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{2617 - 1517}{2617 + 1517}$

$= \frac{11}{41.34}$

$= \underline{0.266}$

The distribution above is considered to be consistent or less variable.

Check your progress-1

Calculate the quartile deviation from the following data

Value	41 - 60	61 - 80	81 - 100	101 - 120	121 - 140	141 - 160

Frequency	8	16	14	22	38	12

9.9 MERITS AND LIMITATIONS OF QUARTILE DEVIATION

A) Merits

i) Quartile deviation is easy to understand and simple to calculate. It is also easy to compute even in the case of open-ended classes.

ii) Since 50% of the observations in the distribution are taken into account it can be treated as a better measure of dispersion than the range.

B) Limitations

i) It doesn't take every value into account. So, it is not a good measure of dispersion.

ii) We cannot use quartile deviation for further statistical treatment.

iii) The value of quartile deviation is affected by sampling fluctuations.

iv) It does not reveal the extent of variability of the values.

Thus, due to these limitations, quartile deviation is not used as a measure of dispersion. To overcome some of the limitations of quartile deviation, certain other measures of dispersion like mean deviation, and standard deviation have been developed.

9.10 SUMMARY

Range is the absolute difference between the maximum and the minimum values in the distribution. It is generally used in weather forecast, quality control, stock markets, etc.

Inter-quartile range is the absolute difference between Q_3 and Q_1. The relative measure of dispersion called coefficient of quartile deviation is used in comparing distributions of different (heterogeneous) statistical units.

Even though it is simple to calculate and easy to understand, it is not the representative measure since it excludes 25% the items on each of the extreme ends of the distribution.

9.11 MODEL ANSWER

$$Q_1 = LQ_1 + \frac{\left(\frac{N}{4} - c.f\right)w}{f} = 85.5$$

$$Q_3 = LQ_3 + \frac{\left(3\frac{N}{4} - c.f\right)w}{f} = 132.34$$

$Q.D = Q_3 - Q_1 = 46.84$

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \underline{\underline{0.22}}$$

9.12 MODEL EXAMINATION QUESTIONS

Short Answer Questions

1. Define
    a) Range
    b) Relative range
    c) Quartile deviation
2. Explain the merits and limitations of quartile deviation.

Exercises

1. Calculate the range (R) and its coefficient (RR) from the following data on average price of a quintal of maize.

Jan	Feb	March	Apr.	May	June	July	Aug	Sep.	Oct	Nov.	Dec
30	42	69	70	83	130	144	150	150	80	30	36

Answer. R = 120

RR = 0.67

2. The following relate the annual income of college teachers. Find out the range and the coefficient of range.

Income (birr)	18,000	20,000	24,000	28,000	40,000	50,000
Number of teachers	250	412	570	45	56	80

3. Calculate the quartile deviation and coefficient of quartile deviation from the following data.

   Annual income 1360,1450,1470,960,1210,1490,2000,1800,1780,1660,1530(birr)

   QD = 210

   Answer: 13.4% or 0.134

4. Calculate quartile deviation and coefficient of quartile deviation. Weight (kg)

Weight (kg)	35	37	39	41	43	45	49	51	53	55
Number of students	12	11	16	25	32	54	43	46	30	21

5. Find out quartile deviation and its coefficient

Class	2-10	11-19	20-28	29-37	38-46
	18	23	9	4	29

Answer: Coefficient of quartile deviation = 0.543

9.13 GLOSSARY

Inter-quartile range: It is two absolute differences between the upper quartile (Q_3) and the lower quartile (Q_1).

Quartile deviation: It is the average absolute difference between the upper quartile (Q_3) and the lower quartile (Q_1).

Coefficient of quartile deviation: The relative measure of quartile deviation

Coefficient of Range: The relative measure of range.

Range: It is the absolute difference between the highest (maximum) value and the lowest (minimum) value of distribution.

9.14 REFERENCES

- Berenson, Mark L, and Levine, David M. Basic Business Statistics Concepts and Applications. New Jersey: Prentice Hall, 1983
- Mason, Robert D, and Lind, Douglas A. Statistical techniques in Business and Economics. U.S.A: IRWIN, 1996.