Measures of Dispersion

BS – Unit II

What is Measures of Dispersion?

- The measure of dispersion indicates the scattering of data. It explains the disparity of data from one another delivering a precise view of the distribution of data. The measure of dispersion displays and gives us an idea about the variation and central value of an individual item.
- In other words, Dispersion is the extent to which values in a distribution differ from the average of the distribution. It gives us an idea about the extent to which individual items vary from one another and from the central value.

- The variation can be measured in different numerical measures namely:
- (i) **Range** – It is the simplest method of measurement of dispersion and defines the difference between the largest and the smallest item in a given distribution. Suppose, If Y max and Y min are the two ultimate items then
- Range = Y max – Y min
- (ii) **Quartile Deviation** – It is known as Semi-Inter-Quartile Range i.e. half of the difference between the upper quartile and lower quartile. The first quartile is derived as (Q), the middle digit (Q1) connects the least number with the median of the data. The median of a data set is the (Q2) second quartile. Lastly, the number connecting the largest number and the median is the third quartile (Q3). Quartile deviation can be calculated by
- $Q = \frac{1}{2} \times (Q3 - Q1)$

- **What Are the Merits and Demerits of Range?**
- **Merits**
- It is very easy to calculate and simple to understand.
- No special knowledge is needed while calculating range.
- It takes least time for computation.
- It provides the broad picture of the data at a glance.
- **Demerits**
- It is a crude measure because it is only based on two extreme values (highest and lowest).
- It cannot be calculated in case of open-ended series.
- Range is significantly affected by fluctuations of sampling i.e. it varies widely from sample to sample.

- **Merits and Demerits of Quartile Deviation**
- **Merits**
- It is also quite easy to calculate and simple to understand.
- It can be used even in case of open-end distribution.
- It is less affected by extreme values so, it a superior to ‘Range’.
- It is more useful when dispersion of middle 50% is to be computed.
- **Demerits**
- It is not based on all the observations.
- It is not capable of further algebraic treatment or statistical analysis.
- It is affected considerably by fluctuations of sampling.
- It is not regarded as very reliable measure of dispersion because it ignores 50% observations.

- (iii) Mean Deviation-Mean deviation is the arithmetic mean (average) of deviations $|D|$ of observations from a central value {Mean or Median or Mode denoted by A}.
- Mean deviation can be evaluated by using the formula:
- $M.D = 1/n \left[\sum |x_i - A|\right]$

- (iv) Standard Deviation- Standard deviation is the Square Root of the Arithmetic Average of the squared of the deviations measured from the mean. The standard deviation is given as
- $\sigma = [(\Sigma (y_i - \bar{y})^2 / n]^{1/2} = [(\Sigma y_i^2 / n) - \bar{y}^2]^{1/2}$

- **What Are the Merits and Demerits of Mean Deviation?**
- **Merits**
- It is based on all the observations of the series and not only on the limits like Range and QD.
- It is simple to calculate and easy to understand.
- It is not much affected by extreme values.
- For calculating mean deviation, deviations can be taken from any average.
- **Demerits**
- Ignoring + and – signs is bad from the mathematical viewpoint.
- It is not capable of further mathematical treatment.
- It is difficult to compute when mean or median are in fraction.
- It may not be possible to use this method in case of Open ended series.

- **Types of Measures of Dispersion**
- (1) Absolute Measures
- Absolute measures of dispersion are expressed in the unit of Variable itself. Like, Kilograms, Rupees, Centimeters, Marks etc.

- (2) Relative Measures
- Relative measures of dispersion are obtained as ratios or percentages of the average.
- These are also known as **'Coefficient of dispersion'**
- These are pure numbers or percentages totally independent of the units of measurements.

- **Characteristics of a Good Measure of Dispersion**
- It should be easy to calculate & simple to understand.
- It should be based on all the observations of the series.
- It should be rigidly defined.
- It should not be affected by extreme values
- It should not be unduly affected by sampling fluctuations.
- It should be capable of further mathematical treatment and statistical analysis.

- **What Are the Various 'Absolute Measures' of Dispersion?**
- Following Are the Different 'absolute Measures' of Dispersion:

- (1) Range
- It is the simplest method of measurement of dispersion.
- It is defined as the difference between the largest and the smallest item in a given distribution.
- **Range = Largest item (L) – Smallest item (S)**

- (2) Inter-quartile Range
- It is defined as the difference between the Upper Quartile and Lower Quartile of a given distribution.
- **Inter-quartile Range = Upper Quartile (Q_3)–Lower Quartile(Q_1)2**

- (3) Quartile Deviation
- It is known as Semi-Inter-Quartile Range i.e. half of the difference between the upper quartile and lower quartile.
- **Quartile Deviation = [upper quartile (Q3) – lower quartile (Q1)]/2**

- (4) Mean Deviation
- Mean deviation is the arithmetic mean (average) of deviations $|D|$ of observations from a central value {Mean or Median or Mode}.

- (5) Standard Deviation
- Standard deviation is the Square Root of the Arithmetic Average of the squared of the deviations measured from the mean.

Quartile deviation

- ***Quantiles*** are the extensions of the median concept because they are values which divide a set of data into equal parts.
  a. **Median** – divides the distribution into two equal parts.
  b. **Quartile** – divides the distribution into four equal parts.
  c. **Decile** – divides the distribution into ten equal parts.
  d. **Percentile** – divides the distribution into one hundred equal parts.
- Quartiles are values in a given set of distribution that divide the data into four equal parts. Each set of scores has three quartiles. These values can be denoted by Q_1, Q_2 and Q_3.
  - **First quartile -** Q_1 (lower quartile) – The middle number between the smallest number and the median of the data set (25^{th} Percentile).
  - **Second quartile -** Q_2 – The median of the data that separates the lower and upper quartile (50^{th} Percentile).
  - Third quartile - Q_3 – (upper quartile) - The middle value between the median and the highest value of the data set (75^{th} Percentile).
- The difference between the upper and lower quartiles is called the Interquartile range. ($IQR = Q_3 - Q_1$)
- Quartile deviation or Semi-interquartile range is one-half the difference between the first and the third quartiles. ($QD = Q_3 - Q_1/2$)

Getting the Quartile Deviation from Ungrouped Data

In getting the quartile deviation from ungrouped data, the following steps are used in getting the quartiles:

1. Arrange the test scores from highest to lowest.
2. Assign serial numbers to each score. The first serial number is assigned to the lowest test score, while the last serial number is assigned to the highest test score.
3. Determine the first quartile (Q_1). To be able to locate Q_1, divide N by 4. Use the obtained value in locating the serial number of the score that falls under Q_1.
4. Determine the third quartile (Q_3), by dividing $3N$ by 4. Locate the serial number corresponding to the obtained answer. Opposite this number is the test score corresponding to Q_3.
5. Subtract Q_1 from Q_3 and divide the difference by 2.

Consider the test scores in the table below:

Score	*Serial Number*
17	1
17	2
26	3
27	4
30	5
30	6
31	7
37	8
$N = 8$	

$$\frac{N}{4} = \frac{8}{4} = 2$$

$$Q_1 = \left(\frac{17 + 26}{2}\right)$$

$$= 21.5$$

$$\frac{3N}{4} = \frac{3(8)}{4} = 6$$

$$Q_3 = \left(\frac{30 + 31}{2}\right)$$

$$= 30.5$$

$$QD = \left(\frac{Q3 - Q1}{2}\right)$$

$$= \left(\frac{30.5 - 21.5}{2}\right)$$

$$= 4.5$$

Getting the Quartile Deviation from Grouped Data

In getting the quartile deviation from grouped data, the following steps are used in getting the quartiles:

1. Cumulate the frequencies from the bottom to the top of the grouped frequency distribution.

2. For the first quartile, use the formula $Q_1 = L + \frac{\frac{N}{4} - CF}{f}(i)$ where:

   L = exact lower limit if the Q_1 class
   N/4 = locator of the Q_1 class
   N = total number of scores
   CF = cumulative frequency below the Q_1 class
   i = class size/interval

3. For the third quartile, use the formula $Q_3 = L + \frac{\frac{3N}{4} - CF}{f}(i)$ where:

   L = exact lower limit if the Q_3 class
   3N/4 = locator of the Q_3 class
   N = total number of scores
   CF = cumulative frequency below the Q_3 class
   i = class size/interval

Example:

Computation of the Quartile Deviation for Grouped Test Scores

Classes	*Frequency (f)*	*Cumulative Frequency (CF)*
46-50	5	53
41-45	7	48
36-40	9	41
31-35	10	32
26-30	8	33
21-25	6	14
16-20	4	8
11-15	4	4
	N = 53	

The computational procedures for determining the quartile deviation for grouped test scores are reflected in the above table.

The computational procedures for determining the quartile deviation for grouped test scores are reflected in the above table.

For the first quartile

$$\frac{N}{4} = \frac{53}{4} = 13.25$$

$$CF = 8 \quad f = 6 \quad L = 20.5$$

$$Q_1 = L + \frac{\frac{N}{4} - CF}{f}(i)$$

$$= 20.5 + \frac{13.25 - 8}{6}(5)$$

$$= 20.5 + \frac{5.25}{6}(5)$$

$$= 20.5 + \frac{31.5}{6}$$

$$= 25.75$$

For the third quartile

$$\frac{3N}{4} = \frac{3(53)}{4} = 40.5$$

$CF = 32$ $f = 9$ $L = 35.5$

$$Q_3 = L + \frac{\frac{3N}{4} - CF}{f}(i)$$

$$= 35.5 + \frac{40.5 - 32}{5}(5)$$

$$= 35.5 + \frac{8.5}{5}(5)$$

$$= 35.5 + \frac{42.5}{6}$$

$$= 40.22$$

After obtaining the first and third quartiles, we can now compute QD. Thus $QD = \left(\frac{Q3 - Q1}{2}\right)$.

$$QD = \left(\frac{40.22 - 25.75}{2}\right)$$

$$= \left(\frac{14.47}{2}\right)$$

$$= 7.235 \text{ or } 7.24$$

Mean deviation(ungrouped data)

Find the mean deviation about the mean for the following data:

6, 7, 10, 12, 13, 4, 8, 12

Mean of the given data $= \frac{Sum\ of\ all\ terms}{Total\ number\ of\ terms}$

$\bar{x} = \frac{6 + 7 + 10 + 12 + 13 + 4 + 8 + 12}{8}$

$= \frac{72}{8}$

$= 9$

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
6	$6 - 9 = -3$	$\lvert -3 \rvert = 3$
7	$7 - 9 = -2$	$\lvert -2 \rvert = 2$
10	$10 - 9 = 1$	$\lvert 1 \rvert = 1$
12	$12 - 9 = 3$	$\lvert 3 \rvert = 3$
13	$13 - 9 = 4$	$\lvert 4 \rvert = 4$
4	$4 - 9 = -5$	$\lvert -5 \rvert = 5$
8	$8 - 9 = -1$	$\lvert -1 \rvert = 1$
12	$12 - 9 = 3$	$\lvert 3 \rvert = 3$
		$\sum_{1}^{8} \lvert x_i - \bar{x} \rvert = 22$

Mean deviation about mean

$= \frac{\sum \lvert x_i - \bar{x} \rvert}{8}$

$= \frac{22}{8}$

= 2.75

Find the mean deviation about the median for the following data:
3, 9, 5, 3, 12, 10, 18, 4, 7, 19, 21.

Arranging data in ascending order,

3, 3, 4, 5, 7, 9, 10, 12, 18, 19, 21

Here, number of observations = n = 11 (odd).

Since **n is odd**,

Median = $\left(\frac{11+1}{2}\right)^{th}$ observation

M = $\left(\frac{12}{2}\right)^{th}$ observation

= 6^{th} observation

= 9

3, 3, 4, 5, 7, 9, 10, 12, 18, 19, 21

Median (M) = 9

Mean deviation about median $= \frac{\sum |x_i - M|}{11}$

$$\text{M.D.(M)} = \frac{\begin{pmatrix}|3-9| + |3-9| + |4-9| + |5-9| + |7-9| + |9-9| \\ + |10-9| + |12-9| + |18-9| + |19-9| + |21-9|\end{pmatrix}}{11}$$

$$= \frac{\begin{pmatrix}|-6| + |-6| + |-5| + |-4| + |-2| + |0| \\ + |1| + |3| + |9| + |10| + |12|\end{pmatrix}}{11}$$

$$= \frac{(6+6+5+4+2+0+1+3+9+10+12)}{11}$$

$$= \frac{(58)}{11}$$

= 5.27

Find mean deviation about the mean for the following data :

x_i	2	5	6	8	10	12
f_i	2	8	10	7	8	5

First we will calculate mean

x_i	f_i	$f_i x_i$	$\lvert x_i - \bar{x}\rvert$	$f_i \lvert x_i - \bar{x}\rvert$
2	2	$2 \times 2 = 4$	$\lvert 2 - 7.5\rvert = \lvert -5.5\rvert = 5.5$	$2 \times 5.5 = 11$
5	8	$5 \times 8 = 40$	$\lvert 5 - 7.5\rvert = \lvert -2.5\rvert = 2.5$	$8 \times 2.5 = 20$
6	10	$6 \times 10 = 60$	$\lvert 6 - 7.5\rvert = \lvert -1.5\rvert = 1.5$	$10 \times 1.5 = 15$
8	7	$8 \times 7 = 56$	$\lvert 8 - 7.5\rvert = \lvert 0.5\rvert = 0.5$	$7 \times 0.5 = 3.5$
10	8	$10 \times 8 = 80$	$\lvert 10 - 7.5\rvert = \lvert 2.5\rvert = 2.5$	$8 \times 2.5 = 20$
12	5	$12 \times 5 = 60$	$\lvert 12 - 7.5\rvert = \lvert 4.5\rvert = 4.5$	$5 \times 4.5 = 22.5$
	$\sum f_i = 40$	$\sum f_i x_i = 300$		$\sum f_i \lvert x_i - \bar{x}\rvert = 92$

$$\text{Mean}(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

$$\bar{x} = \frac{300}{40}$$

$$\bar{x} = 7.5$$

Mean deviation about mean $= \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i}$

Putting $\sum f_i |x_i - \bar{x}| = 92$, $\sum f_i = 40$

$$\text{M.D. }(\bar{x}) = \frac{1}{40} \times 92$$

$$= \mathbf{2.3}$$

Find the mean deviation about the mean for the following data.

Marks obtained	Number of students(f_i)	Mid-point (x_i)	$f_i x_i$
10 – 20	2	$\frac{10+20}{2} = 15$	$2 \times 15 = 30$
20 – 30	3	$15 + 10 = 25$	$3 \times 25 = 75$
30 – 40	8	35	$8 \times 35 = 280$
40 – 50	14	45	$14 \times 45 = 630$
50 – 60	8	55	$8 \times 55 = 440$
60 – 70	3	65	$3 \times 65 = 195$
70 – 80	2	75	$2 \times 75 = 150$
$\sum f_i$	40		$\sum f_i x_i = 1800$

$$\text{Mean}(\bar{x}) = \frac{\sum x_i f_i}{\sum f_i}$$

$$= \frac{1800}{40}$$

$$\bar{x} = 45$$

Marks obtained	Number of students(f_i)	Mid-point (x_i)	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
10 – 20	2	15	$\lvert 15 - 45 \rvert = \lvert -30 \rvert = 30$	$2 \times 15 = 30$
20 – 30	3	25	$\lvert 25 - 45 \rvert = \lvert -20 \rvert = 20$	$3 \times 25 = 75$
30 – 40	8	35	$\lvert 35 - 45 \rvert = \lvert -10 \rvert = 10$	$8 \times 35 = 280$
40 – 50	14	45	$\lvert 45 - 45 \rvert = \lvert 0 \rvert = 0$	$14 \times 45 = 630$
50 – 60	8	55	$\lvert 55 - 45 \rvert = \lvert 10 \rvert = 10$	$8 \times 55 = 440$
60 – 70	3	65	$\lvert 65 - 45 \rvert = \lvert 20 \rvert = 20$	$3 \times 65 = 195$
70 – 80	2	75	$\lvert 75 - 45 \rvert = \lvert 30 \rvert = 30$	$2 \times 75 = 150$
	$\sum f_i = 40$			$\sum f_i x_i = 1800$

$$\sum f_i \lvert x_i - \bar{x} \rvert = 400$$

$$\therefore \text{ Mean Deviation} = \frac{\Sigma f_i \lvert x_i - \bar{x} \rvert}{f_i}$$

$$\text{Mean Deviation } (\bar{x}) = \frac{400}{40}$$

$$= \mathbf{10}$$

Mean deviation

Calculate the mean deviation about median for the following data :

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	6	7	15	16	4	2

Class	Frequency	Cumulative frequency	Mid-point x_i
0 – 10	6	6	5
10 – 20	7	7 + 6 = 13	15
20 – 30	15	13 + 15 = 28	25
30 – 40	16	28 + 16 = 44	35
40 – 50	4	44 + 4 = 48	45
50 – 60	2	48 + 2 = 50	55
	50		

$N = \sum f_i = 50$

$\text{Median Class} = \left(\frac{N}{2}\right)^{th} \text{term}$

$= \left(\frac{50}{2}\right)^{th} \text{term}$

$= 25^{th} \text{ term}$

In above data, cumulative frequency of class 20 - 30 is 28 which is slightly greater than 25.

∴ Median class = 20 - 30

$$\text{Median} = l + \frac{\frac{N}{2} - C}{f} \times h$$

Where,

l = lower limits of median class

N = sum of frequencies

f = frequency of median class

C = Cumulative frequency of class before median class

Here, l = 20, N = 50, C = 13, h = 10, f = 15

$$\text{Median} = l + \frac{\frac{N}{2} - C}{f} \times h$$

$$= 20 + \frac{\frac{50}{2} - 13}{15} \times 10 = 20 + \frac{25 - 13}{15} \times 10 = 20 + \frac{12}{15} \times 10$$

$$= 20 + 8 = \mathbf{28}$$

Finding **mean deviation about Median** = $\frac{\sum f_i |x_i - M|}{\sum f_i}$

Class	Frequency	Cumulative frequency	Mid-point x_i	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
0 – 10	6	6	5	$\lvert 5 - 28 \rvert = 23$	$6 \times 23 = 138$
10 – 20	7	7 + 6 = 13	15	$\lvert 15 - 28 \rvert = 13$	$7 \times 13 = 91$
20 – 30	15	13 + 15 = 28	25	$\lvert 25 - 28 \rvert = 3$	$15 \times 3 = 45$
30 – 40	16	28 + 16 = 44	35	$\lvert 35 - 28 \rvert = 7$	$16 \times 7 = 112$
40 – 50	4	44 + 4 = 48	45	$\lvert 45 - 28 \rvert = 17$	$4 \times 17 = 68$
50 – 60	2	48 + 2 = 50	55	$\lvert 55 - 28 \rvert = 27$	$2 \times 27 = 54$
	$\sum f_i = 50$				$\sum f_i \lvert x_i - M \rvert = 508$

$\sum f_i = 50$ & $\sum f_i |x_i - M| = 508$

$\therefore$ Mean Deviation (M) = $\frac{\sum f_i |x_i - M|}{f_i}$

$= \frac{508}{50}$

= 10.16

Standard deviation

Find the variance and standard deviation for the following data:

First we will calculate mean

x_i	f_i	$x_i - \overline{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \overline{x})^2$
4	3	$4 - 14 = -10$	$(-10)^2 = 100$	$3 \times 100 = 300$
8	5	$8 - 14 = -6$	$(-6)^2 = 36$	$5 \times 36 = 180$
11	9	$11 - 14 = -3$	$(-3)^2 = 9$	$9 \times 9 = 81$
17	5	$17 - 14 = 3$	$(3)^2 = 9$	$5 \times 9 = 45$
20	4	$20 - 14 = 6$	$(6)^2 = 36$	$4 \times 36 = 144$
24	4	$24 - 14 = 10$	$(10)^2 = 100$	$4 \times 100 = 400$
32	1	$32 - 14 = 18$	$(18)^2 = 324$	$1 \times 324 = 324$
	$\sum f_i = 30$			$\sum f_i(x_i - \bar{x})^2 = 1374$

$$\text{Mean}(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

$$\bar{x} = \frac{420}{30}$$

$$\bar{x} = 14$$

Now, finding variance

$$\sum f_i(x_i - \bar{x})^2 = 1374$$

$$\sum f_i = 30$$

Variance $(\sigma^2) = \dfrac{\sum f_i(x_i - \bar{x})^2}{\sum f_i}$

$$= \frac{1374}{30}$$

= 45.8

Standard deviation $(\sigma) = \sqrt{45.8}$

= 6.76

Calculate the mean, variance and standard deviation for the following distribution :

Class	Frequency (f_i)	Mid – point (x_i)	$f_i x_i$
30 – 40	3	35	35 × 3 = 105
40 – 50	7	45	45 × 7 = 315
50 – 60	12	55	55 × 12 = 660
60 – 70	15	65	65 × 15 = 975
70 – 80	8	75	75 × 8 = 600
80 – 90	3	85	85 × 3 = 255
90 – 100	2	95	95 × 2 = 190
	$\sum f_i = 50$		$\sum f_i x_i = 3100$

$\sum f_i x_i = 3100$

$\sum f_i = 50$

$$\text{Mean } (\overline{x}) = \frac{\sum f_i x_i}{\sum f_i}$$

$$= \frac{3100}{50}$$

$$\mathbf{= 62}$$

Finding Variance and Standard Deviation

Class	Frequency (f_i)	Mid – point (x_i)	$(x_i - \overline{x})^2$	$f_i(x_i - \overline{x})^2$
30 – 40	3	35	$(35 - 62)^2 = 729$	$3 \times 729 = 2187$
40 – 50	7	45	$(45 - 62)^2 = 289$	$7 \times 289 = 2023$
50 – 60	12	55	$(55 - 62)^2 = 49$	$12 \times 49 = 588$
60 – 70	15	65	$(65 - 62)^2 = 9$	$15 \times 9 = 135$
70 – 80	8	75	$(75 - 62)^2 = 169$	$8 \times 169 = 1352$
80 – 90	3	85	$(85 - 62)^2 = 529$	$3 \times 529 = 1589$
90 – 100	2	95	$(95 - 62)^2 = 1089$	$2 \times 1089 = 2187$
	$\sum f_i = 50$			Sum = 10050

$$\sum f_i(x_i - \overline{x})^2 = 10050$$

$$\sum f_i = 50$$

$$\text{Variance } (\sigma^2) = \frac{1}{N}\sum f_i(x_i - \bar{x})^2$$

$$= \frac{1}{50} \times 10050$$

$$\mathbf{= 201}$$

$$\text{Standard deviation } (\sigma) = \sqrt{201}$$

$$\mathbf{(\sigma) = 14.17}$$

What Are the Various 'Relative Measures' of Dispersion?

- Following Are the Relative Measure of Dispersion:
- (1) Coefficient of Range
- It refers to the ratio of the difference between two extreme items of the distribution to their sum.
- **Coefficient of Range** =

  [***Largest Item*(*L*)−*Smallest Item*(*S*)]/[*Largest I tem*(*L*)+*Smallest Item*(*S*)]**
- (2) Coefficient of Quartile Deviation
- It refers to the ratio of the difference between Upper Quartile and Lower Quartile of a distribution to their sum.
- Coefficient of Quartile Deviation = (Q3-Q1)/(Q3+Q1)

- (3) Coefficient of Mean Deviation
- Mean deviation is an absolute measure of dispersion.
- In order to transform it into a relative measure, it is divided by the particular average, from which it has been calculated.
- It is then known as the Coefficient of Mean Deviation.
- (4) Coefficient of Standard Deviation
- It is calculated by dividing the standard deviation (σ) by the mean $(\bar{X})$ of the data.
- Coefficient of Standard Deviation $= \sigma / \bar{X}$

- (5) Coefficient of Variation
- It is used to compare two data with respect to stability (or uniformity or consistency or homogeneity).
- It indicates the relationship between the standard deviation and the arithmetic mean expressed in terms of percentage.
- Coefficient of Variation (C.V.) = $(\sigma / \bar{X}) \times 100$
- Where, **C.V.** = Coefficient of Variation; σ = Standard Deviation;
- $\bar{X}$ = Arithmetic Mean