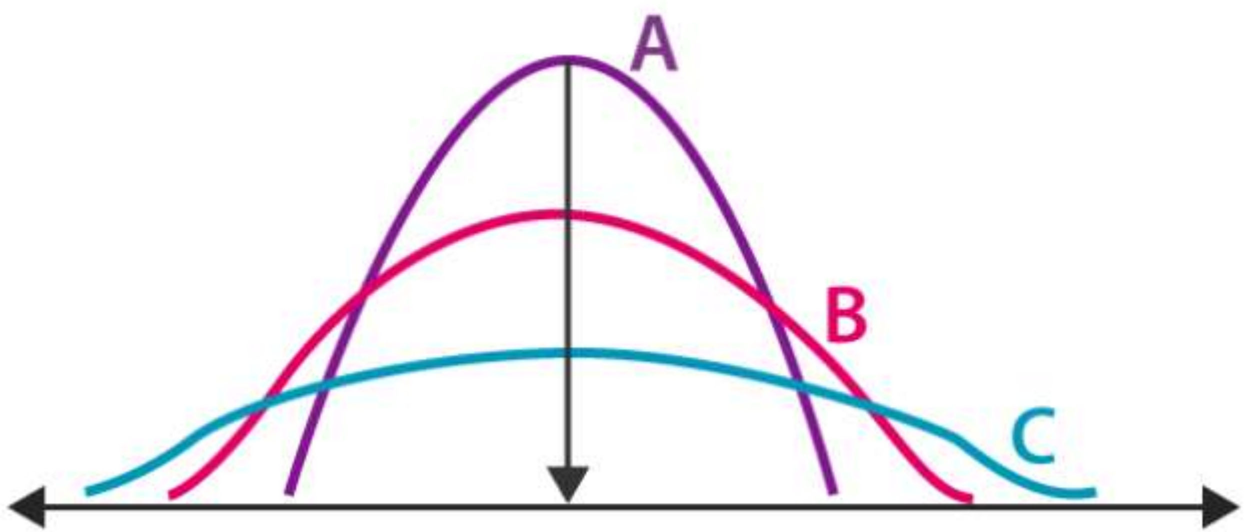
BASIC OF STATISTICS

UNIT-III

MEASURES OF DISPERSION

What is Dispersion in Statistics?

- Dispersion is the state of getting dispersed or spread. Statistical dispersion means the extent to which numerical data is likely to vary about an average value. In other words, dispersion helps to understand the distribution of the data.



EXAMPLE

Section A	Section B
Scores = 70, 70, 70, 70, 85, 85	Scores = 70, 72, 73, 75, 75, 85
Mean = 75	Mean = 75

Table 1. Exam scores for 2 sections of a class.

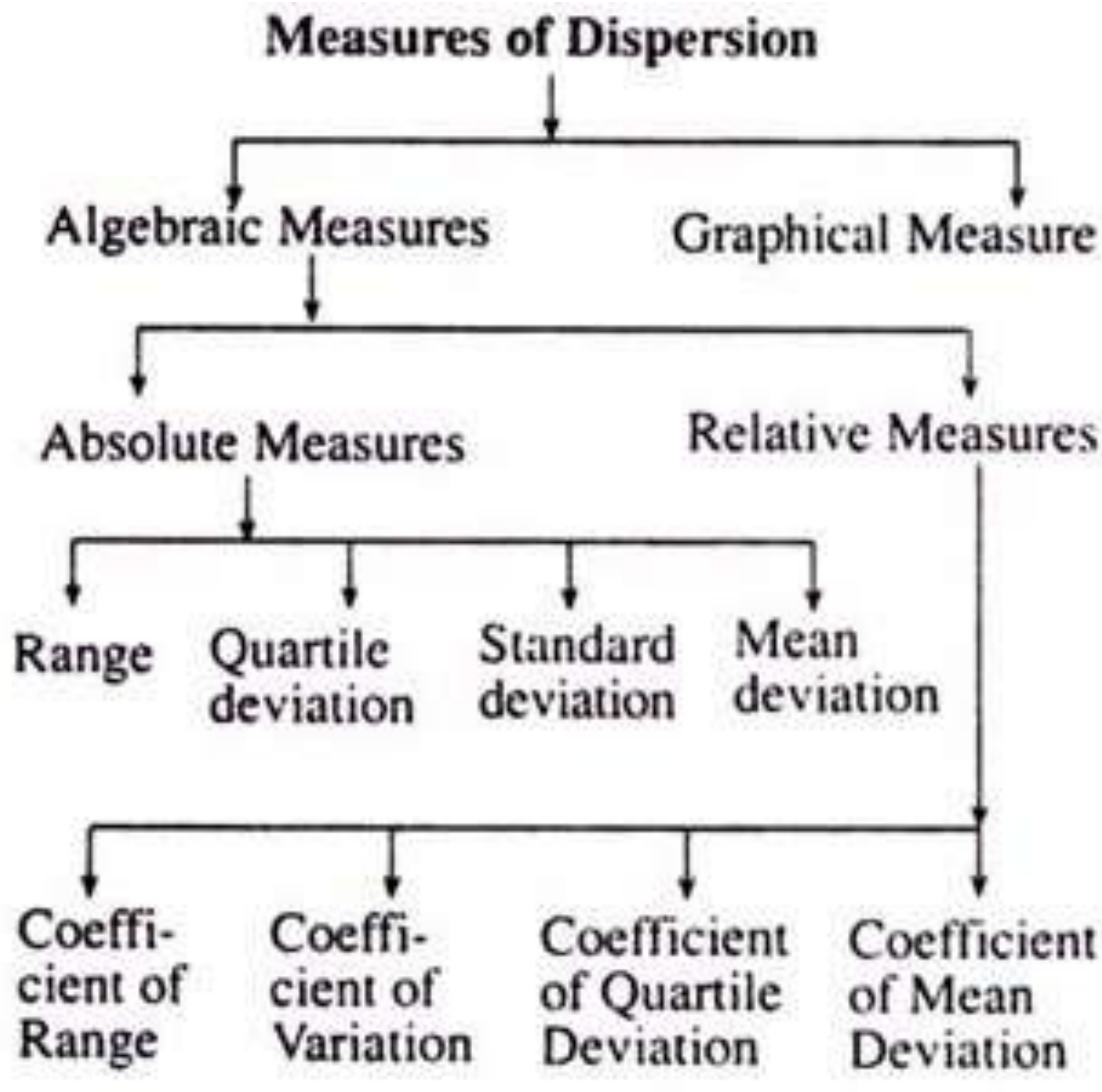
Measures of Dispersion

- In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogeneous the data is. In simple terms, it shows how squeezed or scattered the variable is.
- **Types of Measures of Dispersion**

  There are two main types of dispersion methods in statistics which are:

1. Absolute Measure of Dispersion
2. Relative Measure of Dispersion

The Usual Measures of Dispersion:



Absolute Measure of Dispersion

- An absolute measure of dispersion contains the same unit as the original data set. The absolute dispersion method expresses the variations in terms of the average of deviations of observations like standard or means deviations. It includes range, standard deviation, quartile deviation, etc.

The types of absolute measures of dispersion

- Range
- Variance
- Standard Deviation
- Mean and Mean Deviation
- Quartiles and Quartile Deviation

- **Range:** It is simply the difference between the maximum value and the minimum value given in a data set. Example: 1, 3,5, 6, 7 => Range = 7 -1= 6.

- **Variance:** Deduct the mean from each data in the set, square each of them and add each square and finally divide them by the total no of values in the data set to get the variance. Variance (σ^2) = $\sum(X-\mu)^2/N$.

- **Standard Deviation:** The square root of the variance is known as the standard deviation i.e. S.D. = $\sqrt{\sigma}$.

- **Quartiles and Quartile Deviation:** The quartiles are values that divide a list of numbers into quarters. The quartile deviation is half of the distance between the third and the first quartile.

- **Mean and Mean Deviation:** The average of numbers is known as the mean and the arithmetic mean of the absolute deviations of the observations from a measure of central tendency is known as the mean deviation (also called mean absolute deviation).

Range

- We calculate it by subtracting the smallest score from the largest score in the data set.
- **EXAMPLE:** What is the range of the following group of numbers: 10, 2, 5, 6, 7, 3, 4?

  Highest Number= 10

  Lowest Number= 2

  Therefore, Range = Highest Number- Lowest Number

  = 10 – 2 = 8

Coefficient of Range

- The coefficient of range is the ratio of difference between the highest and lowest value of frequency to the sum of highest and lowest value of frequency.

(I) Individual Series

- **Example 1:**The salaries of 8 factory employees are listed below. Calculate the range and the coefficient of the range. The following salaries are in Indian rupees: 1400, 1450, 1520, 1380, 1485, 1495, 1575, and 1440.
- **Solution:**
- In ascending order, the wages are 1380, 1400, 1440, 1450, 1485, 1495, 1520, 1575
- From the given values of salary, the Largest Item (L) = 1575, and the Smallest Item (S) = 1380

Range = L- S = 1575 – 1380=195

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

$$Coefficient\ of\ Range = \frac{1575-1380}{1575+1380}$$

$$Coefficient\ of\ Range = \frac{195}{2955}$$

Coefficient of Range = 0.065

(II) Discrete Series

- The values of the largest (L) and smallest (S) items in a discrete series should not be confused with the largest and smallest frequencies. They represent the largest and smallest values of the variable. Therefore, the range is determined without taking into account their frequencies by subtracting the smallest item from the largest item.

Example

- The number of homes and the number of people per home are shown in the distribution below. Find the range and coefficient of the range of the following distribution:

Number of Persons	1	2	3	4	5	6	7	8
Number of Houses	25	100	120	90	35	20	7	5

Solution:

Range (R) = Largest Item (L) – Smallest Item (S)

= 8 -1

Range = 7

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

$Coefficient\ of\ Range = \frac{8-1}{8+1}$

$Coefficient\ of\ Range = \frac{7}{9}$

Coefficient of Range = 0.77

(III) Continuous Series

- There are two ways to compute the range and coefficient of range for continuous frequency distributions:
- **1. First Method:** Calculate the difference between the lower limits of the lowest-class interval and the upper limit of the highest-class interval.
- **2. Second Method:** Calculate the difference between the mid-points of the lowest-class interval and the highest-class interval.
- **Note:** Both methods will provide different results. However, both answers will be accurate.

Example 1:

The following data represents the weight of students in kg. Find the range and coefficient of the range using both methods.

Weight (Kg)	50-52	52-54	54-56	56-58	58-60	60-62
Number of Students	10	15	20	40	25	5

Solution:

Weight (Kg)	Number of Students	Mid-value
50-52	10	51
52-54	15	53
54-56	20	55
56-58	40	57
58-60	25	59
60-62	5	61

Range and Coefficient of Range by the First Method:

Range (R) = Largest Item (L) – Smallest Item (S)

= 62 – 50

Range = 12

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

$$Coefficient\ of\ Range = \frac{62-50}{62+50}$$

$$Coefficient\ of\ Range = \frac{12}{112}$$

Coefficient of Range = 0.107

Range and Coefficient of Range by the Second Method:

Range (R) = Mid-point of the Highest Class – Mid-point of the Lowest Class

= 61 – 51

Range = 10

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

$$Coefficient\ of\ Range = \frac{61-51}{61+51}$$

$$Coefficient\ of\ Range = \frac{10}{112}$$

Coefficient of Range= 0.089

Quartile Deviation

- The Quartile Deviation can be defined mathematically as half of the difference between the upper and lower quartile. Here, quartile deviation can be represented as QD; Q_3 denotes the upper quartile and Q_1 indicates the lower quartile.
- **Quartile Deviation is also known as the Semi Interquartile range**.

Quartile Deviation Formula

- Suppose Q_1 is the lower quartile, Q_2 is the median, and Q_3 is the upper quartile for the given data set, then its quartile deviation can be calculated using the following formula.

  QD = $(Q_3 - Q_1)/_2$

Quartile Deviation for Ungrouped Data

- For an ungrouped data, quartiles can be obtained using the following formulas,

  Q_1 = [(n+1)/4]th item

  Q_2 = [(n+1)/2]th item

  Q_3 = [3(n+1)/4]th item
- Where n represents the total number of observations in the given data set.
- Also, Q_2 is the median of the given data set, Q_1 is the median of the lower half of the data set and Q_3 is the median of the upper half of the data set.
- Before, estimating the quartiles, we have to arrange the given data values in ascending order. If the value of n is even, we can follow the similar procedure of finding the median.

Quartile Deviation for Grouped Data

- For a grouped data, we can find the quartiles using the formula,

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

- Here,
- Q_r = the rth quartile
- l_1 = the lower limit of the quartile class
- l_2 = the upper limit of the quartile class
- f = the frequency of the quartile class
- c = the cumulative frequency of the class preceding the quartile class
- N = Number of observations in the given data set

Quartile Deviation Example

- Find the quartiles and quartile deviation of the following data:

  17, 2, 7, 27, 15, 5, 14, 8, 10, 24, 48, 10, 8, 7, 18, 28

  Solution:

  Given data:

  17, 2, 7, 27, 15, 5, 14, 8, 10, 24, 48, 10, 8, 7, 18, 28

  Ascending order of the given data is:

  2, 5, 7, 7, 8, 8, 10, 10, 14, 15, 17, 18, 24, 27, 28, 48

  Number of data values = n = 16

CALCULATION OF Q2

- Q_2 = Median of the given data set
- n is even, median = (1/2) [(n/2)th observation and (n/2 + 1)th observation]
- = (1/2)[8th observation + 9th observation]
- = (10 + 14)/2
- = 24/2
- = 12
- Q_2 = 12

CALCULATION OF Q1

- Now, lower half of the data is:
- 2, 5, 7, 7, 8, 8, 10, 10 (even number of observations)
- Q_1 = Median of lower half of the data
- = (1/2)[4th observation + 5th observation]
- = (7 + 8)/2
- = 15/2
- = 7.5

CALCULATION OF Q3

- Also, the upper half of the data is:
- 14, 15, 17, 18, 24, 27, 28, 48 (even number of observations)
- Q_3= Median of upper half of the data
- = (1/2)[4th observation + 5th observation]
- = (18 + 24)/2
- = 42/2
- = 21

CALCULATION OF QD

- Quartile deviation = $(Q_3 – Q_1)/2$
- = (21 – 7.5)/2
- = 13.5/2
- = 6.75
- Therefore, the quartile deviation for the given data set is 6.75.

Example 2

Calculate the quartile deviation for the following distribution.

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	5	3	4	3	3	4	7	9	7	8

Solution:

Let us calculate the cumulative frequency for the given distribution of data.

Class	Frequency	Cumulative Frequency
0 – 10	5	5
10 – 20	3	5 + 3 = 8
20 – 30	4	8 + 4 = 12
30 – 40	3	12 + 3 = 15
40 – 50	3	15 + 3 = 18
50 – 60	4	18 + 4 = 22
60 – 70	7	22 + 7 = 29
70 – 80	9	29 + 9 = 38
80 – 90	7	38 + 7 = 45
90 – 100	8	45 + 8 = 53

Here, N = 53

We know that,

$$Q_r = l_1 + \frac{r\left(\frac{N}{4}\right) - c}{f}(l_{2-}l_1)$$

Finding Q_1:

r = 1

N/4 = 53/4 = 13.25

Thus, Q1 lies in the interval 30 – 40.

In this case, quartile class = 30 – 40

l_1 = the lower limit of the quartile class = 30

l_2 = the upper limit of the quartile class = 40

f = the frequency of the quartile class = 3

c = the cumulative frequency of the class preceding the quartile class = 12

Now, by substituting these values in the formula we get:

Q_1 = 30 + [(13.25 – 12)/3] × (40 – 30)

= 30 + (1.25/3) × 10

= 30 + (12.5/3)

= 30 + 4.167

= 34.167

Finding Q_3:

$r = 3$

$3N/4 = 3 \times 13.25 = 39.75$

Thus, Q_3 lies in the interval 80 – 90.

In this case, quartile class = 80 – 90

l_1 = the lower limit of the quartile class = 80

l_2 = the upper limit of the quartile class = 90

f = the frequency of the quartile class = 7

c = the cumulative frequency of the class preceding the quartile class = 38

Now, by substituting these values in the formula we get:

$Q_3 = 80 + [(39.75 - 38)/7] \times (90 - 80)$

$= 80 + (1.75/7) \times 10$

$= 80 + (17.5/7)$

$= 80 + 2.5$

$= 82.5$

Finally, the quartile deviation = $(Q_3 - Q_1)/2$

$QD = (82.5 - 34.167)/2$

$= 48.333/2$

$= 24.1665$

Hence, the quartile deviation of the given distribution is 24.167 (approximately).

STANDARD DEVIATION

- A standard deviation (or σ) is a measure of how dispersed the data is in relation to the mean. Low, or small, standard deviation indicates data are clustered tightly around the mean, and high, or large, standard deviation indicates data are more spread out.

Standard Deviation

- Standard deviation (SD) is defined as the positive square root of variance. The formula is

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

and for a frequency distribution the formula is

$$SD = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=}^{k} f_i}}$$

where, all symbols have usual meanings. SD, MD and variance cannot be negative.

VARIANCE

- **Variance is the average of the square of deviations of the values taken from mean. Taking a square of the deviation is a better technique to get rid of negative deviations.**

Variance is defined as

$$Var(x) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and for a frequency distribution, the formula is

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2$$

where, all symbols have their usual meanings.

EXAMPLE: Calculate the variance for the data, 1, 2, 3, 4, 5, 6, 7

- SOLUTION

x	$(x-\bar{x})$	$(x-\bar{x})^2$
1	-3	9
2	-2	4
3	-1	1
4	0	0
5	1	1
6	2	4
7	3	9

We have $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 28$

Therefore, $\mathrm{Var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{7} \times 28 = 4$

EXAMPLE 2:

Calculate the variance of the data:

Class:	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**
Frequency:	**3**	**5**	**7**	**9**	**4**

Solution: We have the following data:

Class	Mid Value	Frequency (f)	$f(x - \bar{x})^2$
0-10	5	3	1470.924
10-20	15	5	737.250
20-30	25	7	32.1441
30-40	35	9	555.606
40-50	45	4	1275.504
		$\sum f = 28$	$\sum f(x_i - \bar{x})^2$ =4071.428

$$\text{Variance} = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} f_i}$$

$$= 4071.429 / 28 = 145.408$$

Merits of Variance

1. It is rigidly defined;
2. It utilizes all the observations;
3. Amenable to algebraic treatment;
4. Squaring is a better technique to get rid of negative deviations; and
5. It is the most popular measure of dispersion.

Demerits of Variance

1.In cases where mean is not a suitable average, standard deviation may not be the coveted measure of dispersion like when open end classes are present. In such cases quartile deviation may be used.

2. It is not unit free.

3. Although easy to understand, calculation may require a calculator or a computer.

4. Its unit is square of the unit of the variable due to which it is difficult to judge the magnitude of dispersion compared to standard deviation.