Measures of Dispersion and Shape of the Distribution (Part 1)

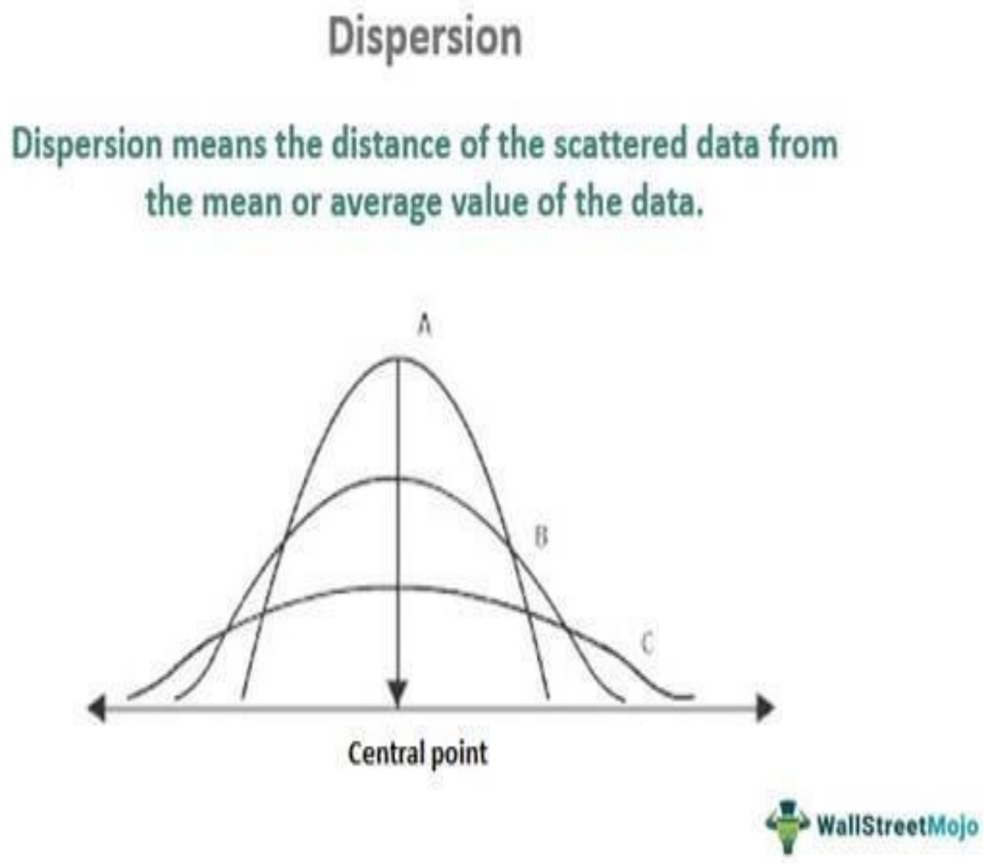
Munmun Akter

Lecturer,

NFE, DIU

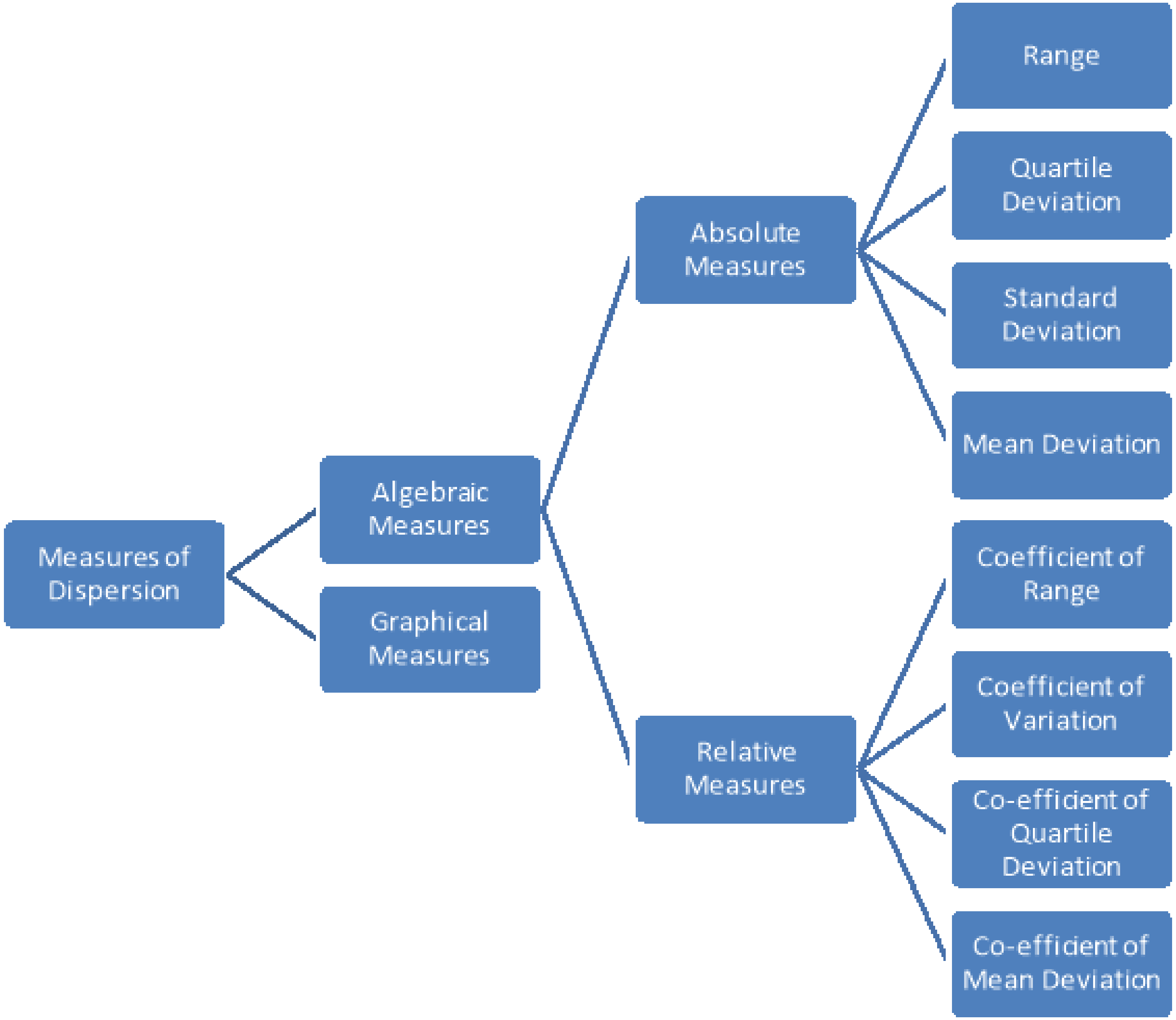
Meaning of Dispersion

- In statistics, dispersion (also called variability, scatter, or spread) refers to the extent to which data points in a dataset are spread out or vary from the central tendency (such as the mean, median, or mode).
- It is a measure of the variability or scatter of the data points around the center.



Measures of dispersion

Measures of dispersion are important because they provide information about the spread or variability of the data, which is essential for understanding the distribution and characteristics of a dataset.



Measures of dispersion

Range and Coefficient of range

The range is the simplest measure of dispersion and is calculated as the difference between the maximum and minimum values in the dataset. It gives an idea of the total spread of the data but does not take into account the distribution of values within that range.

The coefficient of range on the other hand is the ratio of difference between the highest and lowest value of frequency to the sum of highest and lowest value

Range = L – S; Large Value – Small Value

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

Mean Absolute Deviation (MAD):

MAD is the average of the absolute differences between each data point and the mean.

It provides a measure of dispersion that is also in the original units of the data, like the standard deviation.

$$MAD = \frac{\sum_{i=1}^{n} |xi - \bar{x}|}{n}$$

xi : Performance Value for Period i

$\bar{x}$: Average Value

n : Number of Data

Variance:

The variance is the average of the squared differences between each data point and the mean of the dataset.

It quantifies how much the data points deviate from the mean.

Population Variance	Sample Variance
$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
σ^2 = population variance x_i = value of i^{th} element μ = population mean N = population size	s^2 = sample variance x_i = value of i^{th} element $\bar{x}$ = sample mean n = sample size

Standard Deviation:
The standard deviation is the square root of the variance. It is one of the most widely used measures of dispersion because it is in the original units of the data and is easier to interpret than the variance. It represents the average deviation of data points from the mean.

Standard Deviation Formula	
Population	**Sample**
$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$	$s = \sqrt{\frac{\sum(X-\bar{x})^2}{n-1}}$
X – The Value in the data distribution *μ – The population Mean* *N – Total Number of Observations*	*X – The Value in the data distribution* *$\bar{x}$ – The Sample Mean* *n - Total Number of Observations*

Coefficient of Standard Deviation
is a relative measure of Standard Deviation and is determined by dividing Standard Deviation by the Mean of the given data set. It is also known as the Standard Coefficient of Dispersion

$$\text{Coefficient of S.D.} = \frac{\text{Standard Deviation}}{\text{Mean}}$$

$$= \frac{\sigma}{\bar{x}}$$

Coefficient of Variation (CV):

- a statistical measure used to quantify the relative variability of a dataset, particularly when the means of the datasets are on different scales.
- It expresses the standard deviation as a percentage of the mean, allowing for the comparison of the dispersion of datasets with different units or scales.
- The formula for calculating the coefficient of variation is:

Coefficient of Variation (CV)=(Standard Deviation/Mean)×100

Number of guavas, $n = 7$

x_i	x_i^2
3	9
5	25
6	36
4	16
3	9
5	25
4	16
$\Sigma x_i = 30$	$\Sigma x_i^2 = 136$

$$\text{Mean } \bar{x}_1 = \frac{30}{7} = 4.29$$

$$\text{Standard deviation } \sigma_1 = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

$$\sigma_1 = \sqrt{\frac{136}{7} - \left(\frac{30}{7}\right)^2} = \sqrt{19.43 - 18.40} \simeq 1.01$$

Coefficient of variation for guavas

$$\text{C.V}_1 = \frac{\sigma_1}{\bar{x}_1} \times 100\% = \frac{1.01}{4.29} \times 100\% = 23.54\%$$

Number of oranges $n = 7$

x_i	x_i^2
1	1
3	9
7	49
9	81
2	4
6	36
2	4
$\Sigma x_i = 30$	$\Sigma x_i^2 = 184$

$$\text{Mean } \bar{x}_2 = \frac{30}{7} = 4.29$$

$$\text{Standard deviation } \sigma_2 = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

$$\sigma_2 = \sqrt{\frac{184}{7} - \left(\frac{30}{7}\right)^2} = \sqrt{26.29 - 18.40} = 2.81$$

Coefficient of variation for oranges

$$C.V_2 = \frac{\sigma_2}{\bar{x}_2} \times 100\% = \frac{2.81}{4.29} \times 100\% = 65.50\%$$

follow: Calculate the mean, mean deviation, variance and standard deviation.

W(Kg)	f	x	fx	$\lvert x-\bar{x}\rvert$	$f(\lvert x-\bar{x}\rvert)$	$f(\lvert x-\bar{x}\rvert)^2$
54 – 57	5	55.5	277.5	11.44	57.20	654.368
58 – 61	7	59.5	416.5	7.44	52.08	387.4752
62 – 65	10	63.5	635	3.44	34.40	118.336
66 – 69	12	67.5	810	0.56	6.72	3.7632
70 – 73	6	71.5	429	4.56	27.36	124.7616
74 – 77	5	75.5	377.5	8.56	42.80	366.368
78 – 81	4	79.5	318	12.56	50.24	631.0144
82 – 85	1	83.5	83.5	16.56	16.56	274.2336
	50		3347		287.36	2560.32

$$\text{Variance} = \frac{\Sigma f(\lvert x-\bar{x}\rvert)^2}{\Sigma f} = \frac{2560.32}{50}$$

$$\text{Standard deviation} = \sqrt{\text{Variance}}$$

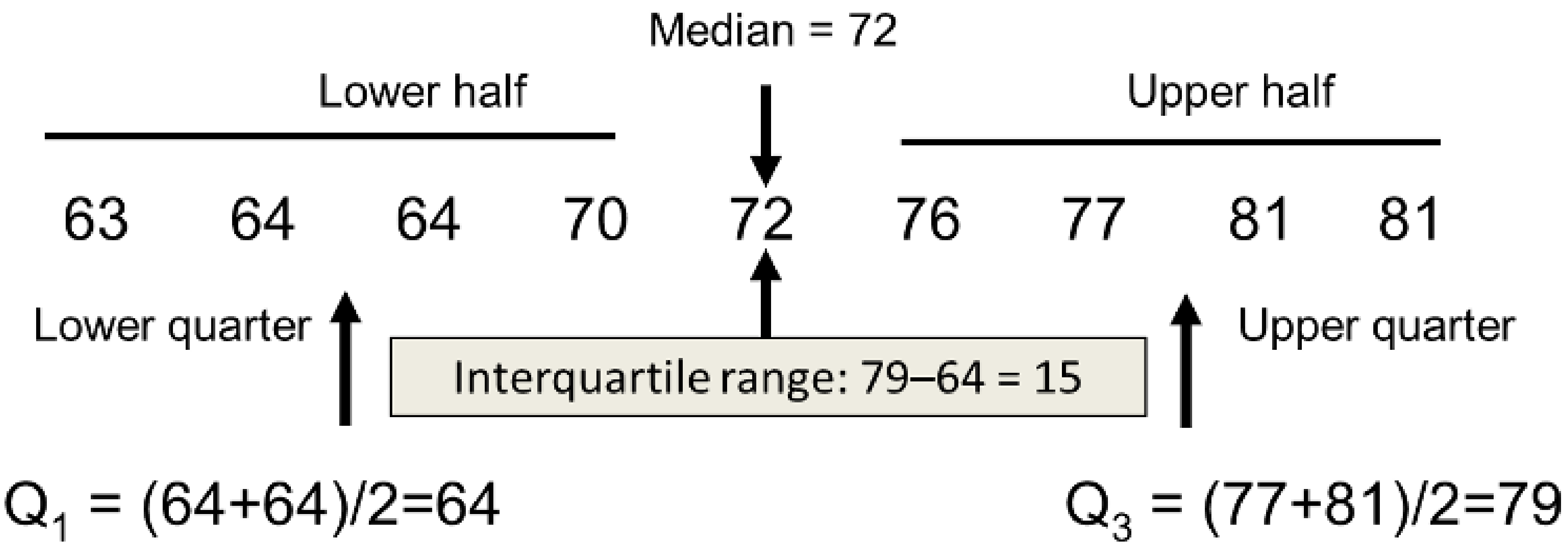
$$= \sqrt{51.2064}$$

Marks scored	Frequency	Mid-point	Frequency $\times$ Mid-point
0 - 9	3	$\frac{0+9}{2} = 4.5$	$3 \times 4.5 = 13.5$
10 - 19	5	$\frac{10+19}{2} = 14.5$	$5 \times 14.5 = 72.5$
20 - 29	8	$\frac{20+29}{2} = 24.5$	$8 \times 24.5 = 196$
30 - 39	4	$\frac{30+39}{2} = 34.5$	$4 \times 34.5 = 138$
	n = 20		Total = 420

Exam Score	Frequency
51-60	4
61-70	8
71-80	15
81-90	8
91-100	5

Interquartile Range (IQR):

The IQR is the range between the first quartile (25th percentile) and the third quartile (75th percentile) of the data. It is a measure of spread that is not affected by extreme values (outliers) as much as the range or standard deviation.



Quartile deviation and Coefficient of quartile deviation

The Quartile deviation is a statistic that measures the deviation in the middle of the data. Quartile deviation is also referred to as the semi interquartile range and is half of the difference between the third quartile and the first quartile value.

The coefficient of quartile deviation is a measure of the spread or variability of a set of data. It is calculated by taking the difference between the upper and lower quartiles, squaring it, and then taking the square root. It is expressed as a percentage, making it easy to compare data sets.

~Quartile Deviation

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

12 15 20 28 30 40 50

n = number of observations =7

Q_1= Size of $\left(\frac{(n+1)}{4}\right)^{th}$ value = Size of $\left(\frac{7+1}{4}\right)^{th}$ value = Size of 2^{nd} value = 15

Hence $Q_1 = 15$

Q_3 = Size of $\left(\frac{3(n+1)}{4}\right)^{th}$ value = Size of $\left(\frac{3\times 8}{4}\right)^{th}$ value = Size of 6^{th} value = 40

Hence $Q_3 = 40$

$$QD = \frac{1}{2}(Q_3 - Q_1) = \frac{40-15}{2} = 12.5$$

$$\text{Coefficient of } QD = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{40-15}{40+15} = \frac{25}{55} = 0.455$$

10	20	36	92	95	40	50	56	60	70
92	88	80	70	72	70	36	40	36	40
92	40	50	50	56	60	70	60	60	88

Compute **all** the measures of central tendency, dispersion and quartile position of the following ungrouped sample data. Then organize and graph the ungrouped data using **an** appropriate chart and table.

70	17	20	38	9	37	14	99	14	53	35
41	59	83	91	74	40	89	53	49	74	15
35	47	3	58	62	75	100	39	14	62	16
28	30	99	66	11	65	15	6	72	6	82
29	98	76	47	56	100	20	87	30	18	46

Table 1: Sample Data

Sr. No.	Absolute Measure	Relative Measure
1	Range = L - S	Coeffficient of Range $= \frac{L - S}{L + S}$
2	Quartile Deviation $Q.D. = \frac{Q_3 - Q_1}{2}$	Coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$
3	Mean Deviation $M.D. = \frac{\sum f_i (x_i - x^{-})}{\sum f_i}$	Coefficient of M.D. $= \frac{M.D.}{Mean}$
4	Standard Deviation $S.D. = \sigma = \sqrt{\frac{\sum f_i (x_i - x^{-})^2}{N}}$	Coefficient of S.D. $= \frac{S.D.}{Mean} = \sigma / x^{-}$
5	Variance $\sigma^2 = \frac{\sum f_i (x_i - x^{-})^2}{\sum f_i}$	Coefficient of Variance $= \frac{\sigma}{x} \times 100$

Application of different measures of dispersion in nutrition and food science

Nutritional Composition Analysis:

- Variance and standard deviation are commonly used to assess the variability in the nutritional content of food products.
- Coefficient of variation (CV) is useful for comparing the relative variability of nutrient levels in different foods.

Sensory Analysis:

- In sensory evaluation of food products, measures of dispersion are employed to assess the variability in the sensory attributes like taste, texture, aroma, and appearance.
- Range and standard deviation are used to describe the spread of sensory scores among panelists for a particular attribute.
- Coefficient of variation (CV) can be applied to compare the sensory variability of different food products or versions within the same product line.

Quality Control and Assurance:

- In food manufacturing and processing, measures of dispersion are utilized to monitor the consistency and quality of food products.
- Range, variance, and standard deviation are helpful in evaluating batch-to-batch variations in product attributes like nutritional content and sensory characteristics.
- Coefficient of variation (CV) is used to assess process stability and identify variations that may affect product quality.

Dietary Assessment and Planning:

- Variability in nutrient intake across individuals can be analyzed using measures of dispersion like standard deviation and coefficient of variation.

Thank You

Any Question?