* **Mode** : Mode is the most ~~frequency~~ frequently answering occuring value in a set of observations. In other words, mode is the value of a variable which occurs with the highest frequency.

Let us consider the set of data 7, 8, 6, 7, 9, 7, 4.
Since 7 is repeated three times and the other values are not repeated. So, 7 is with highest frequency.

$\therefore$ Mode = 7

Exm:

* 6, 5, 7, 5, 2, 3, 3

since 5 and 3 is repeated two times and the other values are not repeated. so, 5 and 3 are highest frequency.

$\therefore$ Mode = 5, 3 $\rightarrow$ bimodal

* 1, 5, 7, 2, 6, 9, 4 $\longrightarrow$ no mode.

Example - 3.22
The responses of 120 athelets on their preferred color of track suits were as follows:

Preferred color	No. of athlets
White	25
Green	21
Blue	54
yellow	12
Orage	6
Red	2

Here, the most preferred color is blue. This blue is regarded as modal color.

*

Age (in years)	Frequency	
24.5 - 29.5	3	
29.5 - 34.5	9	— Pre
34.5 - 39.5	15	→ modal class.
39.5 - 44.5	12	— Post
44.5 - 49.5	7	
49.5 - 54.5	4	

Mode

$M_o = l_o + h\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) = 34.5 + 5\left(\frac{6}{6+3}\right) = 37.83$

Here, l_o = lower limit of the modal class = 34.5

h = size of the modal class / class width = 5

Δ_1 = Difference in the frequency of modal and Pre-modal class.

= 15 - 9 = 6

Δ_2 = " " - post modal class

= 15 - 12 = 3

Page (2)

Geometric Mean :-

The geometric mean of n non-zero positive values $x_1, x_2, \ldots, x_n$ is defined as the nth positive root of the product of all the values.

Symbolically

$$G = (x_1, \ldots x_n)^{\frac{1}{n}}$$

For two observations

$$G = \sqrt{x_1 \cdot x_2}$$

For three observations

$$G = \sqrt[3]{x_1 \cdot x_2 \cdot x_3}$$

page 3

Scanned with CamScanner

$$G = (x_1 \cdot x_2 \cdots x_n)^{1/n}$$

$$\Rightarrow \log G = \tfrac{1}{n} \log (x_1 \cdot x_2 \cdots x_n)$$

$$= \frac{1}{n} (\log x_1 + \log x_2 + \cdots + \log x_n)$$

$$\Rightarrow \log G = \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

$$G = \text{Antilog}\left[\frac{1}{n} \sum_{i=1}^{n} \log x_i\right]$$

Page 4

* Find the geometric mean —

47, 23, 71, 81, 109, 119, 124

$n = \cancel{8}7$

$$\log G = \frac{1}{n} (\log n_1 + \log n_2 + \cdots + \log n_n)$$

$$\Rightarrow \log G = \frac{1}{7} (\log 47 + \log 23 + \log 71 + \log 81 + \log 109 + \log 119 + \log 124)$$

$$\Rightarrow \log G = \frac{1}{7} (3.83 + 3.1 + 4.2 + 4.3 + 4.6 + 4.7 + 4.8)$$

$$\Rightarrow \log G = \frac{29.56}{7}$$

$$\Rightarrow \log G = 4.2$$

$$\Rightarrow \cancel{\log}\ G = \log^{-1}(4.2)$$

$$\Rightarrow G = 66.68 \text{ (Ans)}.$$

Let a set of observation are $x_1, x_2, \cdots x_n$ with frequency $f_1, f_2, \cdots, f_n$

So, $G = \left(x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n}\right)^{1/n}$

$$\Rightarrow \log G = \frac{1}{n} \log \left(x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n}\right)$$

$$= \frac{1}{n}\left[f_1 \log x_1 + f_2 \log x_2 + \cdots + f_n \log x_n\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} f_i \log x_i$$

$$G = \text{Antilog}\left[\frac{1}{n} \sum_{i=1}^{n} f_i \log x_i\right]$$

<u>Harmonic Mean :-</u> Harmonic mean is the reciprocal of the arithmatic mean of the reciprocal of individual values. For a set of values $x_1, \cdots x_n$, Harmonic mean

$$H = \frac{1}{\frac{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}{n}} = \frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}}$$

$$= \frac{n}{\sum^{n} \frac{1}{\ }}$$

For a set of observations $x_1, x_2, \dots x_n$ with frequency $f_1, f_2, \dots f_n$. Harmonic mean H is

$$H = \frac{1}{\dfrac{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \dots + \frac{f_n}{x_n}}{n}} = \frac{n}{\frac{f_1}{x_1} + \dots + \frac{f_n}{x_n}}$$

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

Comparison of measures of central Tendency.

When to use:

Arithmatic Mean

- Applicable only for quantitative data.
- Mean is not suggested to apply when there are extreme values in the data set.
- When the distribution has open ends, mean cannot be computed.

x	f
5–10	5
11–16	2
>17	4

Here this >17 is an open end

1, 2, 3, 4, 6, 10 mean of this set is 4.33. But if we add an extreme value 100 i.e. 1, 2, 3, 4, 6, 10, 100 then what about mean?

Median:

Median is not influenced by extreme values

* Median can be computed for open end distribution

1, 3, 5, 6, 7 median of this set is 5
1, 3, 5, 6, 100 median of this set is also 5.

Mode

* Mode is not influenced by extreme values.
* Mode can be computed for open end distribution.

page 7

G.M (Geometric)

* G.M is useful in averaging ratios and percentages.
* Geometric mean can not be calculated when one or more of values are zero or negative.

H.M (Harmonic)

* In problem related to time and rates harmonic mean provides better result than any other measure of average.
* Harmonic mean can not be calculated when one or more of the values are zero.

Dispersion:-

Student	Math	Statistics	Physics	English	Average
1	68	30	70	40	52
2	49	50	55	54	52
3	51	52	52	53	52

Dispersion:

Dispersion means how the items or observations are spread or scattered on each side of the center.

Some important measures of dispersion–

1. The Range
2. The Quartile Deviation
3. The mean Deviation
4. The variance
5. The standard deviation.

The range:

Range is defined as the difference between the smallest and largest values in the distribution.

If $x_1, x_2, \ldots x_n$ are the values, the range is,

$$R = \max(x_1, x_2, \ldots x_n) - \min(x_1, x_2 \ldots x_n)$$

If the values are arranged in ascending order such that

$$x_1 < x_2 < \ldots < x_n$$

$$R = x_n - x_1$$

~~Disper~~ **Variance:**

$$x_1, x_2, \ldots, x_n$$

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

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

Scanned with CamScanner

Variance:

The variance of a set of data is defined as the sum of squares of deviations of observations from their mean.

devided by the total number of observations.

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

Suppose that $x_1, x_2, \ldots\ldots x_n$ is a sample of size n. Then variance

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

Where $\bar{x}$ is the sample mean.

$$\left[\bar{x} = \frac{x_1 + x_2 + \cdots x_n}{n}\right]$$

$$\left[a + a + \cdots + a = na\right]$$

$$\left[\sum_{i=1}^{n} a = na \qquad \sum_{i=1}^{n} \bar{x}^2 = n\bar{x}^2\right]$$

$$\left[\bar{x} = \frac{x_1 + x_2 + \cdots x_n}{n} = \frac{\sum x_i}{n}\right]$$

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

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

$$= \frac{1}{n} \left(\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2\right)$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - 2 \cdot \sum_{i=1}^{n} \frac{x_i}{n} \cdot \sum_{i=1}^{n} x_i + n\left(\sum_{i=1}^{n} \frac{x_i}{n}\right)^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - 2\frac{(\sum x_i)^2}{n} + n\frac{(\sum x_i)^2}{n^2}\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - 2\frac{(\sum x_i)^2}{n} + \frac{(\sum x_i)^2}{n}\right]$$

$$s^2 = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n}\right]$$

If there are ~~frequencies associated~~ with x_i. Then $s^2 = \frac{1}{n} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2$

$$= \frac{1}{n}\left[\sum f_i x_i^2 - \frac{(\sum f_i x_i)^2}{n}\right]$$

Standard deviation (SD):-

SD is the most important measure of dispersion.
SD is defined as the positive square root of variance.

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

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

Find variance and Standard Deviation from the following data set.

Age in years	Frequency f_i	x_i	$f_i x_i$	$f_i x_i^2$
24.5 – 29.5	3	27	81	2187
29.5 – 34.5	9	32	288	9216
34.5 – 39.5	15	37	555	20535
39.5 – 44.5	12	42	504	21168
44.5 – 49.5	7	47	329	15463
49.5 – 54.5	4	52	208	10816
	$n = \Sigma f_i = 50$		$\Sigma f_i x_i = 1965$	$\Sigma f_i x_i^2 = 79385$

Only the first two columns will be given on exam. Others are to be calculated.

$$s^2 = \frac{1}{n}\left[\Sigma f_i x_i^2 - \frac{(\Sigma f_i x_i)^2}{n}\right]$$

$$= \frac{1}{50}\left[79385 - \frac{(1965)^2}{50}\right]$$

$$= \frac{1}{50}\left[79385 - 77224.5\right]$$

$$= \frac{1}{50} \times 2160.5$$

$$= 43.21$$

$$s = \sqrt{43.21}$$

$$= 6.57$$

[please cross check all the calculation]