Q: Define frequency distribution. What are the types of frequency distribution? (ALP)

A frequency distribution is a tabular arrangement for classifying data into different groups. There are two types of frequency distribution. These are:

(i) Discrete frequency distribution
(ii) Continues frequency distribution

Q: Define grouped data.

The data presented in the form of frequency distribution is called grouped data.

Q: Define ungrouped data.

When the data is collected from any source and record as it stands is called ungrouped data. For example, the number of students collected from each class of a school $10, 15, 20, 14, 15, 18$.

Q: Define frequency.

The number of times a quantity occurs in the data is called its frequency.

Q: What is cumulative frequency. (ALP)

The total frequency up to an upper-class limit or boundary is called cumulative frequency.

Q: Define class limits. (ALP)

The minimum and the maximum values defined for a class or group are called class limits. The minimum value is called the lower-class limit and the maximum value is called the upper class limit.

Q: Define class boundary.

The real class limits of a class is called class boundary. It is obtained by adding two successive class limits and dividing the sum by 2.

Q: Define midpoint or class mark. (ALP)

For a given class the average of that class is obtained by dividing the sum of lower and upper class limits by 2, is called mid-point or class mark.

Q: What is a histogram? (ALP)

A histogram is a graph of adjacent rectangle constructed on XY-plane. It is a graph of frequency distribution.

Q: Define frequency polygon.

A frequency polygon is many sided closed figure.

Q: Define central tendency and write the names of measures of central tendency. (ALP)

The measures or techniques that are used to determine central value are called measures of central tendency. The measures of central tendency are given by

(i) Arithmetic mean **(ii)** Median **(iii)** Mode
(iv) Geometric mean **(v)** Harmonic mean **(vi)** Quartiles

Q: Define arithmetic mean. (ALP)

Arithmetic mean or simply mean is measure by dividing the sum of all values of the variable by their number of observation. Arithmetic mean is denoted by $\bar{X}$.

$$Arithmetic\ mean\ of\ n\ observations = \frac{Sum\ of\ all\ values\ of\ observation}{no.\ of\ observations}$$

$$\bar{X} = \frac{\sum X}{n}$$

Q: Define deviation.

A deviation is defined as a difference of any value of the variable from any constant. That is

$$D_i = x_i - A$$

Q: Write four properties of arithmetic mean. $(\boldsymbol{ALP})$

(i) Mean of a variable with similar observations say constant k is the constant k itself.

(ii) Mean is affected by change in origin.

(iii) Mean is affected by change in scale.

(iv) Sum of deviations of the variable X from its mean is always zero.

Q: Define median. $(\boldsymbol{ALP})$

Median is the middle most observation in an arranged data set. It divides the data set into two equal parts. $\tilde{X}$ is used to represent median. For grouped data median is given by

$$Median = l + \frac{h}{f}\left\{\frac{n}{2} - c\right\}$$

Q: Define mode. $(\boldsymbol{ALP})$

Mode is defined as the most frequent occurring observation of the variable or data. For grouped data mode is given by

$$Mode = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$

Q: Define geometric mean. $(\boldsymbol{ALP})$

Geometric mean of a variable X is the n^{th} positive root of the product of the $x_1, x_2, x_3, \ldots, x_n$ observations. In symbols we write

$$G.M = (x_1 \cdot x_2 \cdot x_3 \cdot \cdots \cdot x_n)^{1/n}$$

Q: What is meant by harmonic mean? $(\boldsymbol{ALP})$

Harmonic mean refers to the value obtained by reciprocating the mean of the reciprocal of $x_1, x_2, x_3, \ldots, x_n$ observations. For grouped data harmonic mean is given by

$$H.M = \frac{\sum f}{\sum \frac{f}{X}}$$

Q: What is meant by dispersion?

Dispersion means the spread or scatterness of observation in a data set.

Q: Define measure of dispersion.

The measure that are used to determine the degree or extent of variation in a data set are called measure of dispersion.

Q: Define range. $(\boldsymbol{ALP})$

Range measures the extent of variation between two extreme observations of a data set. It is given by the formula

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

Where $X_{max}$is the maximum, highest or largest observation and X_{min} is the minimum, lowest or smallest observation.

Q Define variance.

Variance is define as the mean of the squared deviations of X_i $(i = 1,2,3, \ldots, n)$ observations from their arithmetic mean.

$$Variance\ of\ X = Var(X) = S^2 = \frac{\sum(X - \bar{X})^2}{n}$$

Q: Define standard deviation. $(\boldsymbol{ALP})$

Standard deviation is defined as the positive square root of mean of the squared deviations of X_i $(i = 1,2,3, \ldots, n)$ observation from their arithmetic mean.

$$Standard\ deviation\ of\ X = S.D(X) = S = \sqrt{\frac{\sum(X - \bar{X})^2}{n}}$$

Frequency Distribution

Discrete Frequency Distribution:

Following steps are involved in making of a frequency distribution

(i) Find minimum and maximum observation in data and write the values of variable in variable column from minimum to maximum.

(ii) Record the observation by using tally marks (Vertical bar '|')

(iii) Count the tally marks write down the frequency in frequency column.

Q: Five coins are tossed 20 times and number of head recorded at each toss are 3,4, 2, 3, 3, 5, 2, 2, 2, 1, 1, 2, 1, 4, 2, 2, 3, 3, 4, 2. Make frequency distribution of the number of heads observed.

Solution: Let $X = number\ of\ heads$ and here $X_{min} = 1$ and $X_{max} = 5$

Frequency distribution of number of heads		
X	Tally Marks	Frequency
1	𝍸	3
2	𝍸 𝍸	8
3	𝍸	5
4	𝍸	3
5	𝍸	1

Continuous Frequency Distribution:

The making of continuous frequency distribution involves the following steps

(i) Find the range, where $Range = X_{max} - X_{min}$

(ii) Decide the number of groups (denoted by k) into which the data is to be classified (usually an integer between 5 and 20). Usually it depends upon the range. The larger the range the more number of groups.

(iii) Determine the size of class (denoted by h) by using formula $h = \frac{Range}{k}$

Note: The rule of approximation is relaxed in determining h. For example, $h = 7.1$ or $h = 7.9$ may be taken as 8.

(iv) Start writing the class or groups of the frequency distribution usually starting from the minimum observation and keeping in view the size of class.

(v) Record the observation from the data by using tally marks. (Vertical bar '|')

(vi) Count the number of tally marks and record them in the frequency column for each class.

Q: The following data has been obtain after weighing 40 students of class V. Make the frequency distribution taking class interval size 5. Also find class boundaries, midpoints and less than cumulative frequency distribution. 34, 26, 33, 32, 24, 21, 37, 40, 41, 28, 28, 31, 33, 34, 37, 23, 27, 31, 31, 36, 29, 35, 36, 37, 38, 22, 27, 28, 29, 31, 35, 35, 40, 21, 32, 33, 27, 29, 30, 23

Solution: Let

$$Weights\ of\ students = X$$
$$Minimum\ value = X_{min} = 21$$
$$Maximum\ value = X_{max} = 41$$
$$Size\ of\ class\ interval = h = 5$$

Classes	Tally Marks	Frequency (f)	Midpoints (X)	$C.B$
$20 - 24$	𝍸 |	6	22	$19.5 - 24.5$
$25 - 29$	𝍸 𝍸	10	27	$24.5 - 29.5$
$30 - 34$	𝍸 𝍸 ||	12	32	$29.5 - 34.5$
$35 - 39$	𝍸 ||||	9	37	$34.5 - 39.5$
$40 - 44$	|||	3	42	$39.5 - 44.5$

Less Than Cumulative Frequency Distribution

$C.B$	f	$C.F$	Class Boundaries	F
$14.5 - 19.5$	0	0	Less than 19.5	0
$19.5 - 24.5$	6	$0 + 6 = 6$	Less than 24.5	6
$24.5 - 29.5$	10	$6 + 10 = 16$	Less than 29.5	16
$29.5 - 34.5$	12	$16 + 12 = 28$	Less than 34.5	28
$34.5 - 39.5$	9	$28 + 9 = 37$	Less than 39.5	37
$39.5 - 44.5$	3	$37 + 3 = 40$	Less than 44.5	40

Arithmetic Mean	
Ungrouped Data	**Grouped Data**
Direct Method $\bar{X} = \frac{Sum\ of\ all\ values\ of\ observation}{no.of\ observations}$ $\bar{X} = \frac{\sum X}{n}$	**Direct Method** $\bar{X} = \frac{\sum fX}{\sum f}$
Indirect Method **(i) Shortcut** $\bar{X} = A + \frac{\sum D}{n}$ $D = X - A$, where A is any assumed value of X called assumed or provisional **(ii) Coding Method** $\bar{X} = A + \frac{\sum u}{n} \times h$ $u = \frac{X-A}{h}$, where A is any assumed value of X called assumed or provisional and h is the class interval size for unequal intervals.	**Indirect Method** **(i) Shortcut** $\bar{X} = A + \frac{\sum fD}{\sum f}$ $D = X - A$, where A is any assumed value of X called assumed or provisional and X denotes the midpoint of class or group. **(ii) Coding Method** $\bar{X} = A + \frac{\sum fu}{\sum f} \times h$ $u = \frac{X-A}{h}$, where A is any assumed value of X called assumed or provisional and h is the size of class interval.

Median	
Ungrouped Data	**Grouped Data**
Case 1: When the number of observation is odd $\tilde{X} = \left(\frac{n+1}{2}\right)^{th}\ observation$ **Case 2: When the number of observation is even** $\tilde{X} = \frac{1}{2}\left[\left(\frac{n}{2}\right)^{th} observation + \left(\frac{n+2}{2}\right)^{th}\ observation\right]$	**Discrete Data** $Median = the\ class\ containing\ \left(\frac{n}{2}\right)^{th}\ observation$ **Continuous Data** $Median = l + \frac{h}{f}\left\{\frac{n}{2} - c\right\}$ Where l = *lower class boundary of the median class* h = *class interval size of the median class* f = *frequency of median class* c = *cumulative frequency of the class preceding the median class*

Mode	
Ungrouped Data	**Grouped Data**
Ungrouped Data and Discrete Grouped Data $Mode = the\ most\ frequent\ observation$	**Grouped Data (Continuous)** $$Mode = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h$$ Where l = *lower class boundary of modal class* h = *class interval size of modal class* f_m = *frequency of modal class* f_1 = *frequency of the class preceding the modal class* f_2 = *frequency of the class succeeding the modal class*

Weighted Arithmetic Mean
$$\overline{X_W} = \frac{W_1X_1 + W_2X_2 + W_3X_3 + \cdots + W_nW_n}{W_1 + W_2 + W_3 + \cdots + W_n} = \frac{\sum WX}{\sum W}$$

Geometric Mean	
Ungrouped Data	**Grouped Data**
By Definition $$G.M = (x_1 \cdot x_2 \cdot x_3 \cdot \cdots \cdot x_n)^{1/n}$$ **Logarithmic Formula** $$G.M = Anti\log\left(\frac{\sum \log X}{n}\right)$$	$$G.M = Anti\log\left(\frac{\sum f \log X}{\sum f}\right)$$

Harmonic Mean	
Ungrouped Data	**Grouped Data**
$$H.M = \frac{n}{\sum \frac{1}{X}}$$	$$H.M = \frac{\sum f}{\sum \frac{f}{X}}$$

Range	
Ungrouped Data	**Continuous Grouped Data**
$Range = X_{max} - X_{min}$ Where X_{max} = *the maximum, highest or largest observation* X_{min} = *the minimum, lowest or smallest observation*	*Range= (upper class boundary of last group) − (lower class boundary of first group)*

Variance	
Ungrouped Data	**Grouped Data**
By Definition $Variance\ of\ X = Var(X) = S^2 = \frac{\sum(X-\bar{X})^2}{n}$ **By Formula** $Var(X) = S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2$	**By Definition** $Variance\ of\ X = Var(X) = S^2 = \frac{\sum f(X-\bar{X})^2}{\sum f}$ **By Formula** $Var(X) = S^2 = \frac{\sum fX^2}{\sum f} - \left(\frac{\sum fX}{\sum f}\right)^2$

Standard Deviation	
Ungrouped Data	**Grouped Data**
By Definition $Standard\ deviation\ of\ X = S.D(X) = S = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$ **By Formula** $S.D(X) = S = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$	**By Definition** $Standard\ deviation\ of\ X = S.D(X) = S = \sqrt{\frac{\sum f(X-\bar{X})^2}{\sum f}}$ **By Formula** $S.D(X) = S = \sqrt{\frac{\sum fX^2}{\sum f} - \left(\frac{\sum fX}{\sum f}\right)^2}$