GOVERNMENT ARTS AND SCIENCE COLLEGE

(Affiliated to Manonmaniam Sundaranar University, Tirunelveli.)

KANYAKUMARI-629401

STUDY MATERIAL FOR B.COM

QUANTITATIVE TECHNIQUES

IV - SEMESTER

ACADEMIC YEAR 2022-23

PREPARED

BY,

DEPARTMENT OF COMMERCE

II B. COM (IV SEMESTER) PART III – MAJOR CORE - 7

QUANTITATIVE TECHNIQUES

Unit I: Analytical geometry- Distance between two points in a plane-slope of a straight line – equation of straight line – point of intersection of two lines – applications (1) demand and supply(2) cost-output (3) break-even analysis

Unit II: Matrices – meaning – types – algebra of matrices – addition and subtraction – scalar multiplication – Multiplication of matrices-transpose of a matrix –Determinant – minors and co-factors –inverse of a matrix – solving simultaneous linear equations using matrix method.

Unit III: Measures of Central Tendency – Mean – Median – Mode – Geometric Mean . Measures of Dispersion-Range – Quartile Deviation – Mean Deviation - Standard Deviation – Co-efficient of Variation. Skewness - methods of studying Skewness - Karl Pearson's Co-efficient of Skewness – Bowley's co-efficient of Skewness.

Unit IV: Correlation – meaning – types-scatter diagram – Karl Pearson's Co-efficient of Correlation- Rank correlation – Concurrent deviation method. Regression analysis – uses-Regression line – Regression equations – least square method - deviations taken from actual mean and assumed mean method.

Unit V: Index numbers – meaning – types – its problems – Methods of constructing index numbers – unweighted and weighted indices – Index number tests – Consumer price index numbers - Analysis of time series – Meaning – Importance – Components of time series – Seculartrend, seasonal, cyclical and irregular variations – Measurement of trend - Graphic method-Semi average method – Moving average method – Method of least square.

Text / Reference Books

1. D.S. Sancheti & V.K. Kapoor, Business Mathematics Sultan Chand and Sons, New Delhi.

2. M. Manoharan & C. Elango, Business Mathematics, Palani Paramount Publications,

Palani.

3. Dr. S.P. Gupta, Statistical Method, Sultan Chand & Sons, New Delhi.

4. R.S.N. Pillai & Bhagavathi, Statistics-Theory and Practice, S.S. Chand & Co.

5. M. Wilson, Business Statistics, Himalaya Publishing House, Mumbai.

6. Dr. M. Manoharan, Statistical Methods, Palani Paramount Publications, Palani.

7. G.K. Ranganath, Text book of Business Mathematics, Himalaya Publishing House, Delhi.

8. D.C. Sanchetti & B.M. Agarwal, Business Mathematics, Sultan Chand and Sons, New Delhi.

UNIT 1

ANALYTICAL GEOMETRY

Distance between two points:

Example:

1. Calculate the distance between the points (3,7), (4,8)

Solution:

Let the points be P(3,7) and Q(4,8).

$$PQ = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$= \sqrt{(4-3)^2 + (8-7)^2}$$

$$= \sqrt{1+1} = \sqrt{2}$$

Slope of a Straight line:

If M(x_1, y_1) and N(x_2, y_2) are any two points in a straight line, then the slope of the line is

$$\text{Slope of MN} = \frac{y2-y1}{x_2-x_1}$$

Examples:

2. If the distance between two points (5,a) and (8,4) is $\sqrt{13}$, find the value of a.

Solution:

Let A = (5,a) and B = (8,4)

$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$= \sqrt{(8-5)^2 + (4-a)^2}$$

$13 = 3^2 + (4-a)^2$

$= 9 + 4^2 – 2.4a + a^2$

$= 9 + 16 - 8a + a^2$

$13 = a^2 - 8a + 25$

$a^2 - 8a + 25 - 13 = 0$

$a^2 - 8a + 12 = 0$

$(a-6)(a-2) = 0$

$a = 6$ or 2.

2. Find the slope of the line passing through the points (2,5) and (-6,-2).

Solution:

Let P = (2,5) and Q = (-6,-2)

Slope of PQ = $\frac{y2-y1}{x_2-x_1} = \frac{-2-5}{-6-2} = \frac{-7}{-8}$

Equation for a Slope-Intercept form of a Straight line:

The equation for a straight line which intercepts a co-ordinate at ‘c’ and with slope ‘m’ is, y = mx+c.

Example:

If the slope of a straight line is $^2/_5$ and y intercept is 5, then the equation for the line.

Solution:

m = $^2/_5$ and c = 5

y = mx+c

y = $^2/_5$x + 5

y = 2x-5y+25 = 0.

Equation for the point-slope form of straight line:

The equation for a straight line which passes through a point (x_1, y_1) with slope ‘m’ is $y - y_1 = m(x - x_1)$.

Example:

If a line passes through (3,7) with slope 4/5, find the equation for the line.

Solution:

m = 4/5, $x_1 = 3$, $y_1 = 7$

Equation for the line is $y - y_1 = m(x - x_1)$.

y – 7 = 4/5(x-3)

5(y – 7) = 4(x-3)

5y-35 = 4x-12

4x -5y -12 +35 = 0

4x -5y +23 = 0

Equation for Two – point slope form of a straight line:

If the straight line passes through the points (x_1, y_1) and (x_2, y_2) then the equation for the line is

$$\frac{y-y_1}{y_2-y_1} = \frac{x-x_1}{x_2-x_1}$$

Example:

If a line passes through (2, -5) and (4, 6) find the equation for the line.

Solution:

$(x_1, y_1) = (2, -5)$; $(x_2, y_2) = (4,6)$

Equation for the line is $\frac{y-y1}{y_2-y_1} = \frac{x-x1}{x_2-x_1}$

$$\frac{y-(-5)}{6-(-5)} = \frac{x-2}{4-2}$$

$$\frac{y+5}{6+5} = \frac{x-2}{4-2}$$

$$\frac{y+5}{11} = \frac{x-2}{2}$$

2(y + 5) = 11(x -2)

2y +10 = 11x -22

11x-2y-22-10 = 0

11x – 2y – 32 = 0

11x – 2y = 32

Equation for a straight line which intercepts both thee co-ordinates:

If the straight line intercepts the X-coordinate for a distance of ‘a’ and Y-coordinate for a distance of ‘b’ from the origin, then the equation for the straight line is

$$\frac{x}{a} + \frac{y}{b} = 1$$

Example:

If a straight line intercepts the X axis at 5 and Y axis at 6, find the equation for the straight line.

Solution:

a = 5, b = 6

Equation for the straight line is $\frac{x}{a} + \frac{y}{b} = 1$

$$\frac{x}{5} + \frac{y}{6} = 1$$

$$\frac{6x+5y}{30} = 1$$

6x + 5y = 30

$6x + 5y - 30 = 0.$

Intersection of Two Straight Lines

Problems:

1) Find the point of intersection of the lines $4x + 3y = 0$ and $2x+y+2=0$.

Solution:

$4x+3y=0$ -------- (1)

$2x+y= -2$ --------- (2)

$2 \times 2 \Rightarrow 4x+2y = -4$ ------ (3)

$(1) - (3) \Rightarrow y=6+4 = 10$

$(1) \Rightarrow 4x + 3\times10 = 6$

$4x+30 = 6$

$4x=6-30=-24$

$X=-6$

The point of intersection = (-6,10)

2) Find the equation of the straight line passing through the point of intersection of the line $2x-5y+11=0$; $4x+3y-17=0$ and perpendicular to the line $5x+2y-12=0$.

Solution:

$2x-5y = -11$ -------- (1)

$4x+3y = 17$--------- (2)

$(1)\times2 \Rightarrow 4x-10y = -22$ -----(3)

(3)-(2) $\quad -13y = -39$

$Y = 39/13 = 3$

$(1) \Rightarrow 2x-5\times3 = -11$

$2x = 15-11 = 4$

X=2

The point of intersection is (2,3)

The equation to the line perpendicular to ax+by+c=0 is bx-ay+c_1 =0.

The line perpendicular to 5x+2y-12=0 is 2x-5y+ c_1=0

That is, $2\times2-5\times3+ c_1=0$.

$4-15+ c_1=0$

$-11 =- c_1$

$c_1=11$.

Hence the equation for the perpendicular line is 2x-5y+11=0.

3) The cost of production of a company is represented as 5x-y+75 =0. If the company produces 50 units find the marginal cost, total cost, variable cost per unit, fixed cost and total cost per unit.

Solution:

5x-y+75=0

Y=5x+75 x+50 units

Fixed cost = Rs75

Marginal cost = slope=Rs 5

Variable cost per unit = Rs 5

Total cost = y=5x+75 = $5\times50+75$ = 250+75= Rs325

Total cost per unit = Total cost/ output

= 325/50 = Rs 6.50.

4) The demand for a product is 120 units, when the price is Rs20 per unit. When the price decreased to Rs15 per unit the quantity demanded is 200 units. Find the equation to the demand curve. Assume that the demand curve is in linear form.

Solution:

Let x be the quantity demanded and y be the price.

$X_1 = 120$ $y_1 = 20$

$X_2 = 200$ $y_2 = 15$

The equation for the linear demand curve is

$\frac{y-y_1}{y_2-y_1} = \frac{x-x_1}{x_2-x_1}$

$$\frac{y-20}{15-20} = \frac{x-120}{200-120}$$

$$\frac{y-20}{-5} = \frac{x-120}{80}$$

-5(x-120)=80(y-20)

-5x+600 = 80y-1600

5x+80y=2200 or 5x+80y-2200=0.

Break even Analysis

Break-even level, Xe = c/1-m

5) Find the break-even point if the fixed cost is Rs48000 variable cost is Rs 40000 for a sales of Rs100000. Also calculate the profit if the sales of the company is Rs150000.

Solution:

Variable expenses per rupee of sales = 40000/100000 = 0.4

The total cost equation is y=mx+c

i.e) y=0.4x+48000

Break-even point = Xe=c/1-m =48000/1-0.4 = 48000/0.6 = 80000

Break-even sales is Rs80000.

To calculate profit when sales is Rs150000 the total equation may be applied

i.e) y=0.4x+48000

= 0.4(150000)+48000 = 60000+48000= 108000

The total cost for a sales of Rs 150000 is Rs108000.

Profit=Sales – Total cost

= 150000-108000

= Rs42000.

UNIT 2

Matrices

Definition: A matrix is a rectangular arrangement of numbers in rows and columns. The numbers in a matrix are called its elements.

Types of Matrices:

1. **Rectangular Matrix:** A matrix with any number of rows and columns is called rectangular matrix.
2. **Square Matrix:** A matrix with equal number of rows and columns is called square matrix.
3. **Row Matrix:** A matrix with a single row any number of columns is called a row matrix.
4. **Column Matrix:** A matrix with a single column any number of rows is called a column matrix.
5. **Diagonal Matrix:** A diagonal matrix is a square matrix in which all the elements except those on the leading diagonal are zero.
6. **Scalar Matrix:** A diagonal matrix in which all the diagonal elements are equal is called the scalar matrix.
7. **Unit matrix or Identity Matrix:** A diagonal matrix in which each diagonal element is unit is called a unit matrix or identity matrix.
8. **Zero Matrix or Null Matrix:** A matrix in which every element is zero is called a null matrix or zero matrix.
9. **Triangular Matrix:** In triangular matrix every element above (or below) the leading diagonal is zero. Triangular matrix may be upper triangular or lower triangular.

Operations (or) Algebra of Matrices

1. **Equality of Matrices:** Two matrices are said to be equal if they have the same order and all the corresponding elements are equal.
2. **Addition of Matrices:** The sum of two matrices of the same order is the matrix whose elements are the sum of the corresponding elements of the given matrices.
3. **Subtracting of Matrices:** Subtraction of the matrices is also done in the same manner of addition of matrices.
4. **Multiplication of Matrices:** A matrix may be multiplied by any one number or any other matrix. Multiplication of a matrix by any one number is called a scalar multiplication. One matrix may also be multiplied by other matrix.

Transpose of a Matrix: The matrix obtained by interchanging the rows and columns of a matrix is called transpose of a matrix.

Determinants: Determinant is the association of a unique real number with a square matrix. The determinant of a square matrix A is denoted as $|A|$ or Δ.

Inverse of a Matrix: For a square matrix A, if there exists a square matrix B such that AB = BA = 1,then B is called the inverse of A or reciprocal of A. The inverse of A is denoted by A^{-1}, $A^{-1} = \frac{1}{|A|}(Adjoint\ of\ A)$.

Problems:

1) If A = $\begin{pmatrix} 1 & -1 & 4 \\ 2 & 3 & 2 \end{pmatrix}$, B = $\begin{pmatrix} 1 & 1 & -1 \\ 1 & -2 & 3 \end{pmatrix}$ find 2A-3B.

Solution: 2A = $\begin{pmatrix} 2 & -2 & 8 \\ 4 & 6 & 4 \end{pmatrix}$; 3B = $\begin{pmatrix} 3 & 3 & -3 \\ 3 & -6 & 9 \end{pmatrix}$

2A – 3B = $\begin{pmatrix} -1 & -5 & 11 \\ 1 & 12 & -5 \end{pmatrix}$

2) Solve the following equations

3x + 2y +z = 16; 2x + 3y + 2z = 23; 5x + 2y + 2z = 21

Solution: The three equations can be expressed as

$$\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 5 & 2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 16 \\ 23 \\ 21 \end{pmatrix}$$

$A^{-1} = \frac{1}{|A|}(Adjoint\ of\ A)$.

$$|A| = \begin{vmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 5 & 2 & 2 \end{vmatrix} = 3\begin{vmatrix} 3 & 2 \\ 2 & 2 \end{vmatrix} - 2\begin{vmatrix} 2 & 2 \\ 5 & 2 \end{vmatrix} + 1\begin{vmatrix} 2 & 3 \\ 5 & 2 \end{vmatrix}$$

= 3(6-4) – 2(4-10) +1(4-15)

= 6+12-11 = 7

$$C_{11} = -1^2 \begin{vmatrix} 3 & 2 \\ 2 & 2 \end{vmatrix} = 1(6-4) = 2$$

$$C_{12} = -1^3 \begin{vmatrix} 2 & 2 \\ 5 & 2 \end{vmatrix} = -1(4-10) = 6$$

$$C_{13} = -1^4 \begin{vmatrix} 2 & 3 \\ 5 & 2 \end{vmatrix} = 1(4-15) = -11$$

$$C_{21} = -1^3 \begin{vmatrix} 2 & 1 \\ 2 & 2 \end{vmatrix} = -1(4-2) = -2$$

$$C_{22} = -1^4 \begin{vmatrix} 3 & 1 \\ 5 & 2 \end{vmatrix} = 1(6-5) = 1$$

$$C_{23} = -1^5 \begin{vmatrix} 3 & 2 \\ 5 & 2 \end{vmatrix} = -1(6-10) = 4$$

$$C_{31} = -1^4 \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} = 1(4-3) = 1$$

$$C_{32} = -1^5 \begin{vmatrix} 3 & 1 \\ 2 & 2 \end{vmatrix} = -1(6-2) = -4$$

$$C_{33} = -1^6 \begin{vmatrix} 3 & 2 \\ 2 & 3 \end{vmatrix} = 1(9-4) = 5$$

Co factors of A= $\begin{pmatrix} 2 & 6 & -11 \\ -2 & 1 & 4 \\ 1 & -4 & 5 \end{pmatrix}$

Adj A = $\begin{pmatrix} 2 & -2 & 1 \\ 6 & 1 & -4 \\ -11 & 4 & 5 \end{pmatrix}$

$$A^{-1} = \frac{1}{|A|}(Adjoint\, of\, A) = \frac{1}{7}\begin{pmatrix} 2 & -2 & 1 \\ 6 & 1 & -4 \\ -11 & 4 & 5 \end{pmatrix}$$

$$A^{-1}B = \frac{1}{7}\begin{pmatrix} 2 & -2 & 1 \\ 6 & 1 & -4 \\ -11 & 4 & 5 \end{pmatrix}\begin{pmatrix} 16 \\ 23 \\ 21 \end{pmatrix}$$

$$= \frac{1}{7}\begin{pmatrix} 32 & -46 & 21 \\ 96 & 23 & -84 \\ -176 & 96 & 105 \end{pmatrix}$$

$$= \frac{1}{7}\begin{pmatrix} 7 \\ 35 \\ 21 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ 3 \end{pmatrix}$$

x= 1, y = 5, z = 3

3) Find the matrix X which satisfy the relation 4A -5B+2X = 0 where A = $\begin{pmatrix} 1 & 3 & 2 & 5 \\ 2 & 1 & 3 & -1 \\ 0 & 4 & 2 & 3 \end{pmatrix}$, B = $\begin{pmatrix} 2 & 4 & 3 & 1 \\ 3 & 7 & 4 & -2 \\ 5 & 3 & 1 & 8 \end{pmatrix}$.

Solution:

$$4A = \begin{pmatrix} 4 & 12 & 8 & 20 \\ 8 & 4 & 12 & -4 \\ 0 & 16 & 8 & 12 \end{pmatrix}; \quad 5B = \begin{pmatrix} 10 & 20 & 15 & 5 \\ 15 & 35 & 20 & -10 \\ 25 & 15 & 5 & 40 \end{pmatrix}$$

4A -5B+2X = 0

4A -5B= -2X

$$4A - 5B = \begin{pmatrix} -6 & -8 & -7 & 15 \\ -7 & -31 & -8 & 6 \\ -25 & 1 & 3 & -28 \end{pmatrix} = -2X$$

$$X = \begin{pmatrix} 3 & 4 & \frac{7}{2} & \frac{-15}{2} \\ \frac{7}{2} & \frac{31}{2} & 4 & -3 \\ \frac{25}{2} & \frac{-1}{2} & \frac{-3}{2} & 14 \end{pmatrix}$$

4) Find out the Transpose of the matrix A= $\begin{pmatrix} 1 & 3 \\ 5 & 7 \end{pmatrix}$.

Solution:

Transpose of A = $\begin{pmatrix} 1 & 5 \\ 3 & 7 \end{pmatrix}$

UNIT 3

Measures of Central Tendency

Measure of Central Tendency: Usually when two or more different data sets are to be compared it is necessary to condense the data, but for comparison the condensation of data set into a frequency distribution and visual presentation are not enough. It is then necessary to summarize the data set in a single value. Such a value usually somewhere in the center and represent the entire data set and hence it is called measure of central tendency or averages. Since a measure of central tendency (i.e. an average) indicates the location or the general position of the distribution on the X-axis therefore it is also known as a measure of location or position.

Types of Measure of Central Tendency

1. Arithmetic Mean
2. Geometric Mean
3. Harmonic Mean
4. Mode
5. Median

Arithmetic Mean or Simply Mean: *"A value obtained by dividing the sum of all the observations by the number of observation is called arithmetic Mean"*

$$\text{Mean} = \frac{Sum\ of\ All\ Observation}{Number\ of\ Observation}$$

Methods	Ungrouped data	Grouped data
Direct Method	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{x} = \frac{\sum fx}{n}$; Here $n = \sum f$
Short cut Method	$\bar{x} = A + \frac{\sum D}{n}$	$\bar{x} = A + \frac{\sum fD}{n}$; Here $n = \sum f$
	Where $D = X_i - A$ and A is the provisional or assumed mean.	
Step deviation Method	$\bar{x} = A + \frac{\sum u}{n} \times h$	$\bar{x} = A + \frac{\sum fu}{n} \times h$; Here $n = \sum f$
	Where $u = \frac{X_i - A}{h}$ and h is the common width of the class intervals	

Calculate the arithmetic mean for the following the marks obtained by 9students are given below:

Using formula of arithmetic mean for ungrouped data:

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

45
32
37
46
39
36
41
48
36
$\sum x_i = 360$

$n = 9$

$$\bar{X} = \frac{360}{9}$$
$= 40$

✧ **Calculate the arithmetic mean for the following data given below:**

◆ Using formula of **direct method** of **arithmetic mean** for **grouped data**:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}, \qquad \sum_{i=1}^{n} f_i$$

The weight recorded to the nearest grams of 60 apples picked out at random from aconsignment are given below:

106	107	76	82	109	107	115	93	187	95	123	125
111	92	86	70	126	68	130	129	139	119	115	128
100	186	84	99	113	204	111	141	136	123	90	115
98	110	78	185	162	178	140	152	173	146	158	194
148	90	107	181	131	75	184	104	110	80	118	82

Weight (grams)	Frequency
65----84	09
85----104	10
105----124	17
125----144	10
145----164	05
165----184	04
185----204	05

Solution:

Weight (grams)	Midpoints (x_i)	quency(f_i)	$f_i x_i$
65----84	$(65+84)\ 2$ =74.5	09	9×74.5=670.5
85----104	94.5	10	945.0
105----124	114.5	17	1946.5
125----144	134.5	10	1345.0
145----164	154.5	05	772.5
165----184	174.5	04	698.0
185----204	194.5	05	972.5
		$\sum_{i=1}^{n} f_i =$ 60	$\sum_{i=1}^{n} f_i\ x_i =$ 7350.0

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{7350}{60} = 122.5$$

Using formula of **short cut method** of **arithmetic mean** for **grouped data**:

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i D_i}{\sum_{i=1}^{n} f_i}, \qquad \sum_{i=1}^{n} f_i$$

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	$D_i = X_i - A$ $A = 114.5$	$f_i D_i$
65----84	(65 + 84) 2 = 74.5	09	-40	-360
85----104	94.5	10	-20	-200
105----124	114.5	17	0	0
125----144	134.5	10	20	200
145----164	154.5	05	40	200
165----184	174.5	04	60	240
185----204	194.5	05	80	400
		$\sum_{i=1}^{n} f_i$ = 60		$\sum_{i=1}^{n} f_i D_i$ =480

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i D_i}{\sum_{i=1}^{n} f_i}$$

$= 114.5 + \frac{480}{60}$

= 122.5 grams

Geometric Mean: "The n^{th} root of the product of "*n*" positive values is called geometric mean"

Geometric Mean = $\sqrt[n]{\text{product of "n" positive values}}$

- The sum of squared deviations from the mean is smaller than the sum of squared deviations from any arbitrary value or provisional mean. i.e. $\sum(xi-\overline{x})^2 < \sum(xi-A)^2$

Proof: Taking $\sum(xi-A)^2$

$$\begin{aligned}
\sum(xi-A)^2 &= \sum(xi-A+\overline{x}-\overline{x})^2 \\
&= \sum[(xi-\overline{x})+(\overline{x}-A)]^2 \\
&= \sum[(xi-\overline{x})^2+(\overline{x}-A)^2-2(xi-\overline{x})(\overline{x}-A)] \\
&= \sum(xi-\overline{x})^2+\sum(\overline{x}-A)^2-2\sum(xi-\overline{x})(\overline{x}-A) \\
&= \sum(xi-\overline{x})^2+n(\overline{x}-A)^2-2(\overline{x}-A)\sum(xi-\overline{x}) \\
&= \sum(xi-\overline{x})^2+n(\overline{x}-A)^2 \qquad \{\because \sum(xi-\overline{x})=0\}
\end{aligned}$$

$$\Rightarrow \sum(xi-A)^2 < \sum(xi-\overline{x})^2 \qquad \{\because n(\overline{x}-A)^2 > 0\}$$

Note: If $A=\overline{x}$ Then $\sum(xi-A)^2 = \sum(xi-\overline{x})^2$

- The arithmetic mean is affected by the change of origin and scale i.e. when a constant is added to or subtracted from each value of a variable or if each value of a variable is multiplied or divided by a constant, then arithmetic mean is affected by these changes.

Variable	Mean
Xi	$\bar{X}$
$Xi \pm a$	$\bar{X} \pm a$
$a\,Xi$	$a\,\bar{X}$
$\frac{X_i}{a}$	$\frac{\bar{X}}{a}$

- The sum of deviations from mean is equal to zero. i.e. $\sum(xi - \bar{x}) = 0$

Proof: Sum of Deviation

$$\begin{aligned} &= \sum(xi - \bar{x}) \\ &= \sum xi - \sum \bar{x} \\ &= \sum xi - n\bar{x} \qquad (\because \bar{x} \text{ is constant}) \\ &= \sum xi - n\left(\frac{\sum xi}{n}\right) \qquad \left(\because \bar{x} = \frac{\sum xi}{n}\right) \\ &= \sum xi - \sum xi \\ &= 0 \end{aligned}$$

Ungrouped data	Grouped data
$G = Antilog\left(\frac{\sum logx}{n}\right)$	$G = Antilog\left(\frac{\sum f\ logx}{n}\right)$;Here $n = \sum f$

Numerical example of geometric Mean for both grouped and ungrouped data:

Calculate the geometric mean for the following the marks obtained by 9students are given below:

x_i	45	32	37	46	39	36	41	48	36

Solution: Using formula of geometric mean for ungrouped data:

$n = 9$

x_i	$\log x_i$
45	$\log 45 = 1.65321$
32	1.50515
37	1.56820
46	1.66276
39	1.59106
36	1.55630
41	1.61278
48	1.62124
36	1.55630
	$\sum_{i=1}^{n} \log x_i = 14.38700$

G.M = antilog $\left(\frac{\sum \log x}{n}\right)$ = antilog $\left(\frac{14.38700}{9}\right)$ = antilog (1.59856)=39.68.

➢ Given the following frequency distribution of weights of 60 apples, calculate the geometric mean for grouped data.

Weights	65--84	85--104	105--124	125--144	145--164	165--184	185--204
Frequency	09	10	17	10	05	04	05

Solution:

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	$\log x_i$	$f_i \log x_i$
65----84	$(65+84)/2 = 74.5$	09	1.8722	16.8498
85----104		10	1.9754	19.7540
105----124	94.5	17	2.0589	35.0013
125----144	114.5	10	2.1287	21.2870
145----164	134.5	05	2.1889	10.9445
165----184	154.5	04	2.2418	8.9672
185----204	174.5	05	2.2889	11.4445
	194.5			
		$\sum_{i=1}^{n} f_i = 60$		$\sum_{i=1}^{n} f_i \log x_i =$ 124.2483

G.M=antilog $\frac{\sum f \log x}{n}$=antilog $\frac{124.2483}{60}$=antilog 2.0708 = 117.7

Median: *"when the observation are arranged in ascending or descending order,then a value, that divides a distribution into equal parts, is called median"*

Median in case of Ungrouped Data	
In this case we first arrange the observations in **increasing** or **decreasing** order then we use the following formulae for Median:	
If "n" is odd	$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$
If "n" is even	$Median = \frac{size\ of \left\{\left(\frac{n}{2}\right) th + \left(\frac{n}{2}+1\right) th\right\} observation}{2}$

- **Numerical example of median for both grouped and ungrouped data:**

If "n" is odd	$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$

- **Calculate the median for the following the marks obtained by 9 students aregiven below:**

x_i	45	32	37	46	39	36	41	48	36

Solution: Arrange the data in ascending order

32, 36, 36, 37, 39, 41, 45, 46, 48.

n=9, n is odd

Median = Size of $(\frac{n+1}{2})^{th}$ observation

Median = Size of $(\frac{9+1}{2})^{th}$ observation

Median = Size of $(5)^{th}$ observation

Median = 39.

If "n" is even	$Median = \dfrac{size\ of \left\{\left(\dfrac{n}{2}\right)th + \left(\dfrac{n}{2}+1\right)th\right\} observation}{2}$

- **Calculate the median for the following the marks obtained by 10 studentsare given below:**

x_i	45	32	37	46	39	36	41	48	36	50

Solution: Arrange the data in ascending order

32, 36, 36, 37, 39, 41, 45, 46, 48, 50.

Median = Size of $\dfrac{(\frac{n}{2})^{th} + (\frac{n}{2}+1)^{th}}{2}$ observation

Median = Size of $\dfrac{(\frac{10}{2})^{th} + (\frac{10}{2}+1)^{th}}{2}$ observation

Median = Size of $\dfrac{(5^{th}+6^{th})}{2}$ observation

Median = $\dfrac{39+41}{2}$ = 40.

- The number of values above the median balances (equals) the number of values below the median i.e. 50% of the data falls above and below the median.

Median in case of Discrete Grouped Data
In case of discrete grouped data, first we find the cumulative frequencies and then use the following formula for Median: $$Median = size\ of \left(\frac{n+1}{2}\right)th\ observation$$ Here $n = \sum f$

Numerical examples: The following distribution relates to the number ofassistants in 50 retail establishments.

No.of assistant	0	1	2	3	4	5	6	7	8	9
f	3	4	6	7	10	6	5	5	3	1

Solution:

No. of assistants	f_i	Cumulative frequency ($c.f$)
0	3	3
1	4	7
2	6	13
3	7	20
4	10	30
5	6	36
6	5	41
7	5	46
8	3	49
9	1	50
	$\sum_{i=1}^{n} f_i = 50$	

n = 50 "n" is even

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

Median = Size of $\frac{\left(\frac{50}{2}\right)^{th} + \left(\frac{50}{2}+1\right)^{th}}{2}$ observation

Median = Size of $\frac{(25^{th} + 26^{th})}{2}$ observation

Median = $\frac{4+4}{2}$ = 4.

Median in case of continuous Grouped Data

In continuous grouped data, when we are finding median, we first construct the class boundaries if the classes are discontinuous. Then we find cumulative frequencies and then we use the following two steps:

- First we determine the median class using n/2.
- When the median class is determined, then the following formula is used to find the value of median. i.e.

$$Median = l + \frac{h}{f}\left(\frac{n}{2} - C\right); \quad \text{Here } n = \sum f$$

Where l = lower class boundary of the median class
h = width of the median class
f = frequency of the median class
C = cumulative frequency of the class preceding the median class.

Example: Find the median, for the distribution of examination marksgiven below:

Marks	30--39	40--49	50--59	60--69	70--79	80--89	90--99
umber of students	08	87	190	304	211	85	20

Solution:

Median = $l + \frac{h}{f}\left(\frac{n}{2} - C\right)$ here n=$\sum f_i$

Class boundaries	Midpoints x_i	Frequency $\sum f_i$	Cumulative frequency
29.5---39.5	34.5	8	8
39.5---49.5	44.5	87	95
49.5---59.5	54.5	190	285
59.5---69.5	64.5	304	589
69.5---79.5	74.5	211	800
79.5---89.5	84.5	85	885
89.5---99.5	94.5	20	905
		$\sum f_i = 905$	

$\frac{n}{2} = \frac{905}{2}$ =452.5th student which corresponds to marks in the class 59.5---69.5. Therefore, Median = $l + \frac{h}{f}(\frac{n}{2} - C)$ =59.5+ $\frac{10}{304}(\frac{905}{2} - 285) = 59.5 + \frac{10}{304}(452.5 - 285)$

$= 59.5 + \frac{1675}{304} = 65$ marks.

Mode in case of Ungrouped Data: "**A value that occurs most frequently in a data is called mode" or**

"if two or more values occur the same number of times but most frequently than the

other values, the there is more than one whole"

"If two or more values occur the same number of times but most frequently thanthe other values, then there is more than one mode"

- The data having one mode is called uni-modal distribution.
- The data having two modes is called bi-modal distribution.
- The data having more than two modes is called multi-modal distribution.

Mode in case of Discrete Grouped Data: **"A value which has the largest frequencyin a set of data is called mode"**

Mode in case of Continuous Grouped Data: In case of continuous grouped data, mode would lie in the class that carries the highest frequency. This class is called the modal class. The formula used to compute the value of mode, is given below:

Examples of Mode for ungrouped and grouped data

Calculate Mode for ungrouped data

x_i : 2, 3, 8, 4, 6, 3, 2, 5, 3.

Mode = 3 (Answer).

➢ Calculate Mode in discrete grouped data

No. of assistants	f_i
0	3
1	4
2	6
3	7
4	10
5	6
6	5
7	5

8	3
9	1
	$\sum f_i = 50$

Mode = 4

Measures of Dispersion

As the name suggests, the measure of dispersion shows the scatterings of the data. It tells the variation of the data from one another and gives a clear idea about the distribution of the data. The measure of dispersion shows the homogeneity or the heterogeneity of the distribution of the observations.

Range: Range refers to the difference between each series' minimum and maximum values. The range offers us a good indication of how dispersed the data is, but we need other measures of variability to discover the dispersion of data from central tendency measurements. A range is the most common and easily understandable measure of dispersion. It is the difference between two extreme observations of the data set. If X_{max} and X_{min} are the two extreme observations then

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

Merits of Range

- It is the simplest of the measure of dispersion
- Easy to calculate
- Easy to understand
- Independent of change of origin

Demerits of Range

- It is based on two extreme observations. Hence, get affected by fluctuations
- A range is not a reliable measure of dispersion
- Dependent on change of scale

Quartile Deviation

The quartiles divide a data set into quarters. The first quartile, (Q_1) is the middle number between the smallest number and the median of the data. The second quartile, (Q_2) is the median of the data set. The third quartile, (Q_3) is the middle number between the median and the largest number.

Quartile deviation or semi-inter-quartile deviation is

$$Q = \frac{1}{2} \times (Q_3 - Q1)$$

Merits of Quartile Deviation

- All the drawbacks of Range are overcome by quartile deviation
- It uses half of the data
- Independent of change of origin
- The best measure of dispersion for open-end classification

Demerits of Quartile Deviation

- It ignores 50% of the data
- Dependent on change of scale
- Not a reliable measure of dispersion

Mean Deviation

Mean deviation is the arithmetic mean of the absolute deviations of the observations from a measure of central tendency. If $x_1, x_2, \dots, x_n$ are the set of observation, then the mean deviation of x about the average A (mean, median, or mode) is

$$\text{Mean deviation from average } A = \frac{1}{n} \left[\sum_i |x_i - A|\right]$$

For a grouped frequency, it is calculated as:

$$\text{Mean deviation from average } A = \frac{1}{N} \left[\sum_i f_i |x_i - A|\right], \; N = \sum f_i$$

Here, x_i and f_i are respectively the mid value and the frequency of the i^{th} class interval.

Merits of Mean Deviation

- Based on all observations
- It provides a minimum value when the deviations are taken from the median
- Independent of change of origin

Demerits of Mean Deviation

- Not easily understandable
- Its calculation is not easy and time-consuming
- Dependent on the change of scale
- Ignorance of negative sign creates artificiality and becomes useless for further mathematical treatment

Standard Deviation

A standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by a Greek letter sigma, σ. It is also referred to as root mean square deviation. The standard deviation is given as

$$\sigma = [(\Sigma_i (y_i - \bar{y}) / n]^{½} = [(\Sigma_i y_i^2 / n) - \bar{y}^2]^{½}$$

For a grouped frequency distribution, it is

$$\sigma = [(\Sigma_i f_i (y_i - \bar{y}) / N]^{½} = [(\Sigma_i f_i y_i^2 / n) - \bar{y}^2]^{½}$$

The square of the standard deviation is the **variance**. It is also a measure of dispersion.

$$\sigma^2 = [(\Sigma_i (y_i - \bar{y}) / n]^{½} = [(\Sigma_i y_i^2 / n) - \bar{y}^2]$$

For a grouped frequency distribution, it is

$$\sigma^2 = [(\Sigma_i f_i (y_i - \bar{y}) / N]^{½} = [(\Sigma_i f_i x_i^2 / n) - \bar{y}^2].$$

If instead of a mean, we choose any other arbitrary number, say A, the standard deviation becomes the root mean deviation.

Variance of the Combined Series

If σ_1, σ_2 are two standard deviations of two series of sizes n_1 and n_2 with means $\bar{y}_1$ and $\bar{y}_2$. The variance of the two series of sizes $n_1 + n_2$ is:

$$\sigma^2 = (1/ n_1 + n_2) \div [n_1 (\sigma_1^2 + d_1^2) + n_2 (\sigma_2^2 + d_2^2)]$$

where, $d_1 = \bar{y}_1 - \bar{y}$, $d_2 = \bar{y}_2 - \bar{y}$, and $\bar{y} = (n_1 \bar{y}_1 + n_2 \bar{y}_2) \div (n_1 + n_2)$.

Merits of Standard Deviation

- Squaring the deviations overcomes the drawback of ignoring signs in mean deviations
- Suitable for further mathematical treatment
- Least affected by the fluctuation of the observations
- The standard deviation is zero if all the observations are constant
- Independent of change of origin

Demerits of Standard Deviation

- Not easy to calculate
- Difficult to understand for a layman
- Dependent on the change of scale

Range: Range (R) is the difference between the largest (L) and the smallest value (S) in a distribution. Thus, R = L – S Higher value of range implies higher dispersion and vice-versa.

Quartile Deviation: The presence of even one extremely high or low value in a distribution can reduce the utility of range as a measure of dispersion. Thus, you may need a measure which is not unduly affected by the outliers. In such a situation, if the entire data is divided into four equal parts, each containing 25% of the values, we get the values of quartiles and median. The upper and lower quartiles (Q3 and Q1, respectively) are used to calculate inter-quartile range which is Q3 – Q1. Interquartile range is based upon middle 50% of the values in a distribution and is, therefore, not affected by extreme values. Half of the inter-quartile range is called quartile deviation (Q.D.). Thus: Q.D. is therefore also called Semi Inter Quartile Range.

Calculation of Range and Q.D. for ungrouped data

Example: Calculate range and Q.D. of the following observations: 20, 25, 29, 30, 35, 39, 41, 48, 51, 60 and 70

Solution:

Range is clearly 70 – 20 = 50

For Q.D., we need to calculate values of Q3 and Q1.

Q1 is the size of $\frac{n+1}{4}$th value. n being 11, Q1 is the size of 3rd value. As the values are already arranged in ascending order, it can be seen that Q1, the 3rd value is 29. [What will you do if these values are not in an order?] Similarly, Q3 is size of $\frac{3(n+1)}{4}$th value; i.e. 9th value which is 51. Hence Q3 = 51

Q.D= Q3 – Q1/2= 51- 29/2 = 11.

Example: For the following distribution of marks scored by a class of 40 students, calculate the Range and Q.D.

Class intervals	No. of students
0-10	5
10-20	8
20-40	16
40-60	7
60-90	4
	40

Solution: Range is just the difference between the upper limit of the highest class and the lower limit of the lowest class. So range is 90 – 0 = 90. For Q.D., first calculate cumulative frequencies as follows:

Class Intervals	Frequencies	Cumulative Frequencies
0-10	5	5
10-20	8	13
20-40	16	29
40-60	7	36
60-90	4	40
	N=40	

Q1 is the size of $\frac{n}{4}$th value in a continuous series. Thus, it is the size of the 10th value. The class containing the 10th value is 10–20. Hence, Q1 lies in class 10–20. Now, to calculate the exact value of Q1, the following formula is used:

Q1=L+$\frac{\frac{n}{4}+cf}{f}\times i$, where L=10, c.f=5, i=10,f=8

Q1 = 10+$\frac{10-5}{8}\times 10$=16.25

Similarly, Q3 is the size of $\frac{3n}{4}$th value. i.e, 30th value, which lies in class 40-60.

Q3= $L+\frac{\frac{3n}{4}+cf}{f}\times i$

Q3=$40+\frac{30-29}{7}\times 20$= 42.87

Q.D = 42.87-16.25/2 = 13.31.

Mean Deviation

Calculation of Mean Deviation from Arithmetic Mean for ungrouped data.

Direct Method

Steps: (i) The A.M. of the values is calculated

(ii) Difference between each value and the A.M. is calculated. All differences are considered positive. These are denoted as |d|

(iii) The A.M. of these differences (called deviations) is the Mean Deviation.

i.e. M.D. = $\frac{\Sigma|d|}{n}$

Example: Calculate the mean deviation of the following values; 2, 4, 7, 8 and 9.

The A.M. = $\frac{\Sigma X}{n}$ = 6

X	|d|
2	4
4	2
7	1
8	2
9	3
	12

M.D = 12/5 = 2.4.

Mean Deviation from median for ungrouped data.

Method Using the values in the above Example, M.D. from the Median can be calculated as follows,

(i) Calculate the median which is 7.

(ii) Calculate the absolute deviations from median, denote them as |d|.

(iii) Find the average of these absolute deviations. It is the Mean Deviation.

Example

X	d=|X-MEDIAN|
2 4 7 8 9	5 3 0 1 2
	11

M. D. from Median is thus,

$$M.D_{Median} = \frac{\Sigma|d|}{n} = 11/5 = 2.2$$

Mean Deviation from Mean for Continuous Distribution

Steps:

(i) Calculate the mean of the distribution.

(ii) Calculate the absolute deviations |d| of the class midpoints from the mean.

(iii) Multiply each |d| value with its corresponding frequency to get f|d| values. Sum them up to get Σ f|d|.

(iv) Apply the following formula, M.D = $\frac{\Sigma f|d|}{\Sigma f}$.

Profits of Companies (Rs in lakh) Class intervals	Number of companies
10-20	5
20-30	8
30-50	16
50-70	8
70-80	3

Solution:

C.I	f	m.p	|d|	f|d|
10-20	5	15	25.5	127.5
20-30	8	25	15.5	124.5
30-50	16	35	0.5	8
50-70	8	55	19.5	156
70-80	3	75	34.5	103.5
	40			519

M.D = $\frac{\Sigma f|d|}{\Sigma f}$ = 519/40 = 12.975.

Standard Deviation

Calculation of Standard Deviation for ungrouped data

Four alternative methods are available for the calculation of standard deviation of individual values. All these methods result in the same value of standard deviation.

These are: (i) Actual Mean Method (ii) Assumed Mean Method (iii) Direct Method (iv) Step-Deviation Method

Actual Mean Method:

Example: Suppose you have to calculate the standard deviation of the following values: 5, 10, 25, 30, 50

Solution: First step is to calculate

$\overline{X}$=5+10+25+30+50/5 = 120/5 = 24

X	d=x-$\overline{X}$	d^2
5	-19	361
10	-14	196
25	1	1
30	6	36
50	26	676
		1270

$\sigma = \sqrt{\frac{\Sigma d^2}{n}} = \sqrt{\frac{1270}{5}}$ = 15.937.

Assumed Mean Method

For the same values, deviations may be calculated from any arbitrary value A x such that d = x-$A\overline{X}$. Taking $A\overline{X}$ = 25, the computation of the standard deviation is shown below:

X	d=x-$A\overline{X}$	d^2
5	-20	400
10	-15	225
25	0	0
30	5	25
50	25	625
		1275

Formula for Standard Deviation

$\sigma = \sqrt{\frac{\Sigma d^2}{n} - \frac{(\Sigma d)^2}{n^2}} = \sqrt{\frac{1275}{5} - \frac{(-5)^2}{25}}$ = 15 937

Coefficient of Variation

1. The standard deviation and mean of a data are 6.5 and 12.5 respectively. Find the coefficient of variation.

Solution :

S.D $\sigma = 6.5$ mean $\overline{x} = 12.5$

Coefficient of variation $C.V = \frac{\sigma}{\overline{x}} \times 100$

$$= \frac{6.5}{12.5} \times 100$$

Ans : Coefficient of variation = 52 %

2. The standard deviation and coefficient of variation of a data are 1.2 and 25.6 respectively. Find the value of mean.

Solution :

S.D $\sigma = 12.5$ C.V = 25.6

$$C.V = \frac{\sigma}{\overline{x}} \times 100$$

$$25.6 = \frac{1.2}{\overline{x}} \times 100$$

$$\overline{x} = \frac{1.2}{25.6} \times 100$$

$$= 4.96$$

Answer : mean = 4.69

3. If the mean and coefficient of variation of a data are 15 and 48 respectively, then find the value of standard deviation.

Solution :

$\overline{x} = 15$ C.V = 48

$$C.V = \frac{\sigma}{\overline{x}} \times 100$$

$$48 = \frac{\sigma}{15} \times 100$$

$$\sigma = \frac{48 \times 15}{100} = 7.2$$

Answer : Standard deviation = 7.2

4. Find the coefficient of variation of 24, 26, 33, 37, 29, 31.

Solution :

Arrange in ascending order 24, 26, 29, 31, 33, 37.

$$\overline{x} = \frac{24+26+29+31+33+37}{6} = \frac{180}{6} = 30$$

$\overline{x} = 30$

x_i	$d_i = x_i - \overline{x}$ $= x_i - 30$	d_i^2
24	−6	36
26	−4	16
29	−1	1
31	1	1
33	3	9
37	7	49
	$\Sigma d_i = 0$	$\Sigma d_i^2 = 112$

$$\sigma = \frac{\Sigma d_i^2}{n}$$

$$= \sqrt{\frac{112}{6}} \qquad = \sqrt{18.67}$$

$$= 4.32$$

$$C.V = \frac{\sigma}{\overline{x}} \times 100$$

$$= \frac{4.32}{30} \times 100 = 14.4$$

Answer :

Coefficient of variation = 14.4 %

5. The total marks scored by two students Sathya and Vidhya in 5 subjects are 460 and 480 with standard deviation 4.6 and 2.4 respectively. Who is more consistent in performance?

Solution :

$n = 5$

Vidhya	Sathya
total marks $\sum x = 480$	total marks $\sum x = 460$
S.D $\sigma = 2.4$	S.D $\sigma = 4.6$
$\bar{x} = \frac{\Sigma x}{n} = \frac{480}{5}$	$\bar{x} = \frac{\Sigma x}{n} = \frac{460}{5}$
$\bar{x} = 96$	$\bar{x} = 92$
$C.V = \frac{\sigma}{x} \times 100$	$C.V = \frac{\sigma}{x} \times 100$
$= \frac{2.4}{96} \times 100 = 2.5$	$= \frac{4.6}{92} \times 100 = 5$

C.V of Vidhya < C.V of Sathya.

Answer : Vidhya is more consistent.

6. Find which city is more consistent in temperature changes?

Solution :

Find the coefficient of variation for city A and city B.

<u>City A</u>

$$\bar{x} = \frac{18+20+22+24+26}{5}$$

$$= \frac{110}{5} = 22$$

$\bar{x} = 22$

x_i	$d_i = x_i - \bar{x} = x_i - 22$	d_i^2
18	−4	16
20	−2	4
22	0	0
24	2	4
26	4	16
	$\Sigma d_i = 0$	$\Sigma d_i^2 = 40$

$$\sigma = \sqrt{\frac{\Sigma d_i^2}{n}}$$

$$= \sqrt{\frac{40}{5}} = \sqrt{8}$$

$\sigma = 2.828$

$$C.V = \frac{\sigma}{x} \times 100$$

$$= \frac{2.828}{22} \times 100$$

$= 12.85$ %

<u>City B</u>

$$\bar{x} = \frac{11+14+15+17+18}{5}$$

$$= \frac{75}{5} = 15$$

$\bar{x} = 15$

x_i	$d_i = x_i - \bar{x} = x_i - 22$	d_i^2
18	−4	16
20	−2	4
22	0	0
24	2	4
26	4	16
	$\Sigma d_i = 0$	$\Sigma d_i^2 = 40$

$$\sigma = \sqrt{\frac{\Sigma d_i^2}{n}}$$

$$= \sqrt{\frac{30}{5}} = \sqrt{6}$$

$\sigma = 2.449$

$$C.V = \frac{\sigma}{x} \times 100$$

$$= \frac{2.449}{15} \times 100$$

$= 16.33$ %

C.V of city A < C.V of city B.

Answer : City A is more consistent.

Karl-Pearson's Coefficient of skewness:

According to Karl – Pearson, the absolute measure of skewness = mean – mode. This measure is not suitable for making valid comparison of the skewness in two or more distributions because the unit of measurement may be different in different series. To avoid this difficulty, we use relative measure of skewness, called Karl Pearson's coefficient of skewness given by:

Karl-Pearson's Coefficient of Skewness = Mean − Mode /S.D.

In case of mode is ill- defined, the coefficient can be determined by

Coefficient of skewness = 3(Mean − Median) /S.D.

Bowleys's Coefficient of Skewness

This method is based on quartiles. The formula for calculating coefficient of skewness is given by

Bowley's Coefficient of Skewness = $\frac{Q3-2\ median+Q1}{Q3-Q1}$

Example: For a distribution Karl Pearson's coefficient of skewness is 0.64, standard deviation is 13 and mean is 59.2 Find mode and median.

Solution: We have given S_k = 0.64, σ = 13 and Mean = 59.2

Therefore by using formulae S_k = Mean –Mode/ σ

0.64 =59.2–Mode/13

Mode = 59.20 – 8.32

= 50.88

Mode = 3 Median – 2 Mean

50.88 = 3 Median - 2 (59.2)

Median = 50.88+118.4/3 = 169.28/3 = 56.42

Example: Calculate Karl – Pearson's coefficient of skewness for the following data. 25, 15, 23, 40, 27, 25, 23, 25, 20

Solution:

Size	Deviation from A=25 d	d^2
25	0	0
15	-10	100
23	-2	4
40	15	225
27	2	4
25	0	0
23	-2	4
25	0	0
20	-5	25
N=9	∑d=-2	$\sum d^2 = 362$

Mean = A+∑d/n = 25+(-2)/9 = 25-0.22 = 24.78

$$\sigma = \sqrt{\frac{\sum d^2}{n} - \frac{(\sum d)^2}{n^2}} = \sqrt{\frac{362}{9} - \frac{(-2)^2}{81}} = \sqrt{40.22 - 0.05} = 6.3$$

Mode=25 as the size of item repeats 3 times

Karl Pearson's coefficient of skewness is given by

= Mean – Mode/ S.D

= 24.78-25/6.3 = -0.22/6.3 = -0.03.

Example: Find the coefficient of skewness from the data given below

Size	: 3	4	5	6	7	8	9	10
Frequency	: 7	10	14	35	102	136	43	8

Solution:

Size	f	Deviation from A=6 (d)	d^2	fd	fd^2
3	7	-3	9	-21	63
4	10	-2	4	-20	40
5	14	-1	1	-14	14
6	35	0	0	0	0
7	102	1	1	102	102
8	136	2	4	272	544
9	43	3	9	129	387
10	8	4	16	32	128
	n=355			$\sum fd$=480	$\sum fd^2$=1278

Mean = A+∑d/n = 6+480/355 = 6+1.35 = 7.35

$\sigma = \sqrt{\frac{\sum d^2}{n} - \frac{(\sum d)^2}{n^2}} = \sqrt{\frac{1278}{355} - \frac{(480)^2}{355^2}} = \sqrt{\cancel{3.6 - 1.82}} = 1.33$

Mode = 8

Coefficient of skewness = Mean-Mode/S.D = 7.35-8/1.33 = -.65/1.33 = -0.5.

Example: Find Karl – Pearson' s coefficient of skewness for the given distribution:

X:	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
F:	2	5	7	13	21	16	8	3

Solution:

Mode lies in 20-25 group which contains the maximum frequency

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$$

L=20, f_1=21, f_0=13, f_2=16, C=5

Mode = $20 + \frac{21-13}{42-13-16} \times 5$

$= 20 + \frac{8}{13} \times 5 \quad = 20 + 40/13 = 23.08$

X	Mid point m	f	d’= m-22.5/5	fd’	d'^2	fd'^2
0-5	2.5	2		-8	16	32
5-10	7.5	5	-4	-15	9	45
10-15	12.5	7	-3	-14	4	28
15-20	17.5	13	-2	-13	1	13
20-25	22.5	21	-1	0	0	0
25-30	27.5	16	0	16	1	16
30-35	32.5	8	1	16	4	32
35-40	37.5	3	2	9	9	27
		N=75	3	-9		193

Mean = A+∑fd/n ×C =22.5 + (-9/75) ×5 = 22.5-0.6 = 21.9

$$\sigma = \sqrt{\frac{\sum d^2}{n} - \frac{(\sum d)^2}{n^2}} \times C$$

$$= \sqrt{\frac{193}{75} - \frac{81}{75^2}} \times 5 = \sqrt{\cancel{2.57 - 0.0144}} \times 5 = \sqrt{\cancel{2.5556}} \times 5 = 1.5986 \times 5 = 7.99$$

Karl Pearson’s coefficient of Skewness $= \frac{Mean - Mode}{s.d}$

$$= \frac{21.9 - 23.08}{7.99} = \frac{-1.18}{7.99} = -0.1477$$

Example:

Find the Bowley’ s coefficient of skewness for the following series. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22

Solution:

The given data in order

2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22

Q1 = size of $\frac{n+1}{4}$ th item

= size of $\frac{11+1}{4}$ th item

= size of 3rd item = 6

Q3 = size of $3(\frac{n+1}{4})$ th item

= size of $3(\frac{11+1}{4})$ th item

= size of 9th item = 18

Median = size of $(\frac{n+1}{2})$ th item

= size of $(\frac{11+1}{2})$ th item

= size of 6^{th} item = 12

Bowley's Coefficient of Skewness $= \frac{Q3-2\ median+Q1}{Q3-Q1}$

= 18+6-2×12 / 18-6 = 0.

UNIT 4

CORRELATION

Correlation Coefficient is a statistical concept, which helps in establishing a relation between predicted and actual values obtained in a statistical experiment. The calculated value of the correlation coefficient explains the exactness between the predicted and actual values.

Correlation Coefficient value always lies between -1 to +1. If correlation coefficient value is positive, then there is a similar and identical relation between the two variables. Else it indicates the dissimilarity between the two variables.

The covariance of two variables divided by the product of their standard deviations gives Pearson's correlation coefficient. It is usually represented by ρ (rho).

$\rho\ (X,Y) = \text{cov}\ (X,Y) / \sigma X.\sigma Y.$

Here cov is the covariance. σX is the standard deviation of X and σY is the standard deviation of Y. The given equation for correlation coefficient can be expressed in terms of means and expectations.

$\rho(X,Y)=E(X-\mu x)(Y-\mu y)/\sigma x.\sigma y$

μx and μy are mean of x and mean of y respectively. E is the expectation.

Scatter Diagram

A scatter diagram is a diagram that shows the values of two variables X and Y, along with the way in which these two variables relate to each other. The values of variable X are given along the horizontal axis, with the values of the variable Y given on the vertical axis.

Later, when the regression model is used, one of the variables is defined as an independent variable, and the other is defined as a dependent variable. In regression, the independent variable X is considered to have some effect or influence on the dependent variable Y. Correlation methods are symmetric with respect to the two variables, with no indication of causation or direction of influence being part of the statistical consideration. A scatter diagram is given in the following example. The same example is later used to determine the correlation coefficient.

Pearson Correlation Coefficient Formula

If x & y are the two variables of discussion, then the correlation coefficient can be calculated using the formula

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$$

Here,

n = Number of values or elements

$\sum x$ = Sum of 1st values list

$\sum y$ = Sum of 2nd values list

$\sum xy$ = Sum of the product of 1st and 2nd values

$\sum x^2$ = Sum of squares of 1st values

$\sum y^2$ = Sum of squares of 2nd values

Example: Calculate the Correlation coefficient of given data:

x	12	15	18	21	27
y	2	4	6	8	12

Solution:

Here n = 5

x	12	15	18	21	27
y	2	4	6	8	12
xy	24	60	94	168	324
x^2	144	225	324	441	729
y^2	4	16	36	64	144

$\sum x = 93$

$\sum y = 32$

$\sum xy = 670$

$\sum x^2 = 1863$

$\sum y^2 = 264$

Now, substitute all the values in the below formula.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$$

We have, r = 0.84

Example: Find Karl Pearson's correlation coefficient if N = 0, ∑X=75,∑Y=80,∑X2=130,∑Y2=140 and ∑XY=128.

Solution:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$$

$$\frac{50\times128-(75\times80)}{\sqrt{50\times130-75^2}\sqrt{50\times140-80^2}} = \frac{6400-6000}{\sqrt{6500-5625}\sqrt{7000-6400}} = \frac{400}{\sqrt{875}\sqrt{600}}$$

$$= \frac{400}{724.57} = .55$$

Example: Find Karl Pearson's coefficient of correlation between capital employed and profit obtained from the following data.

Capital Employed (Rs. In Crore)	10	20	30	40	50	60	70	80	90	100
Profit (Rs. In Crore)	2	4	8	5	10	15	14	20	22	50

Solution:

X	Y	X^2	Y^2	XY
10	2	100	4	20
20	4	400	16	80
30	8	900	64	240
40	5	1600	25	200
50	10	2500	100	500
60	15	3600	225	900
70	14	4900	196	980
80	20	6400	400	1600
90	22	8100	484	1980
100	50	10000	2500	5000

550	150	38500	4014	11500

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$$

$$= \frac{10\times11500-(550\times150)}{\sqrt{10\times38500-550^2}\sqrt{10\times4014-150^2}} = \frac{115000-82500}{\sqrt{385000-302500}\sqrt{40140-22500}} = \frac{32500}{\sqrt{82500}\sqrt{17640}}$$

$$= \frac{32500}{38148.3945} = .8519$$

Spearman's Rank Coefficient of Correlation

R = 1 - $\frac{6\sum D^2}{N^3-N}$ where, D =Difference of the ranks between paired items in two series. N = Number of pairs of ranks

Example: Find out spearman's coefficient of correlation between the two kinds of assessment of graduate students' performance in a college.

Name of students	A	B	C	D	E	F	G	H	I
Internal Exam	51	68	73	46	50	65	47	38	60
External Exam	49	72	74	44	58	66	50	30	35

Solution:

Name	Internal Exam	Ranks (R1)	External Exam	Ranks (R2)	D= R1 – R2	D^2
A	51	5	49	6	-1	1
B	68	2	72	2	0	0
C	73	1	74	1	0	0
D	46	8	44	7	1	1
E	50	6	58	4	2	4
F	65	3	66	3	0	0
G	47	7	50	5	2	4
H	38	9	30	9	0	0
I	60	4	35	8	-4	16
						26

R = 1 - $\frac{6\sum D^2}{N^3-N}$ = 1 - $\frac{6\times 26}{9^3-9}$ = 1 - $\frac{156}{729-9}$ = 1- 0.2167 = 0.7833.

Example: Ten competitors in a beauty contest are ranked by three judges in the following order:

1st Judge	1	6	5	10	3	2	4	9	7	8
2nd Judge	3	5	8	4	7	10	2	1	6	9
3rd Judge	6	4	9	8	1	2	3	10	5	7

Use the rank correlation coefficient to determine which pairs of judges has the nearest approach to common tastes in beauty.

Solution:

Rank by 1st Judge (R1)	Rank by 2nd Judge (R2)	Rank by 3rd Judge (R3)	$D^2 = (R1-R2)^2$	$D^2 = (R2-R3)^2$	$D^2 = (R1-R3)^2$
1	3	6	4	9	25
6	5	4	1	1	4
5	8	9	9	1	16
10	4	8	36	16	4
3	7	1	16	36	4
2	10	2	64	64	0
4	2	3	4	1	1
9	1	10	64	81	1
7	6	5	1	1	4
8	9	7	1	4	1
N=10	N=10	N=10	200	214	60

1. 1st Judge and 2nd Judge: R = 1 - $\frac{6\sum D^2}{N^3-N}$ = 1 - $\frac{6\times 200}{10^3-10}$ = 1 - $\frac{1200}{990}$ = 1- 1.2121 = -0.2121.

2. 2nd Judge and 3rd Judge: R = 1 - $\frac{6\sum D^2}{N^3-N}$ = 1 - $\frac{6\times 214}{10^3-10}$ = 1 - $\frac{1284}{990}$ = 1- 1.297 = -0.297.

3. 1st Judge and 2nd Judge: R = 1 - $\frac{6\sum D^2}{N^3-N}$ = 1 - $\frac{6\times 60}{10^3-10}$ = 1 - $\frac{360}{990}$ = 1- 0.3636 = 0.6364.

From the above calculation it can be observed that coefficient of correlation is positive in the judgment of the first and third judges. Therefore, it can be concluded that first and third judges have the nearest approach to common tastes in beauty.

REGRESSION LINES

If we take the case of two variables X and Y, we shall have two regression lines as the regression of X on Y and the regression of Y on X. The regression line of Y on X gives the most probable values of Y for given values of X and the regression line of X on Y gives the most probable values of X for given values of Y.

EXAMPLE

Height of fathers(inches) : 65 63 67 64 68 62 70 66 68 67 69 71

Height of sons(inches) : 68 66 68 65 69 66 68 65 71 67 68 70

The two regression equations corresponding to these variables are:

X = -3.38 + 1.036Y--------(1)

Y = 35.82 + 2.476X ------- (2)

By assuming any values of Y we can find out corresponding values of X from equation(1).

For example, if Y = 65,

X would be -3.38 + 1.036(65) = -3.38 + 67.34 = 63.96

Similarly, if X = 70,

Y would be -3.38 + 1.036(70) = -3.38 + 72.52 = 69.14

REGRESSION EQUATIONS

Regression equations also known as estimating equations, are algebraic expressions of the regression lines. Since there are two regression lines, there are two regression equations – the regression equation of X on Y is used to describe the variations in the values of X for given changes in Y and the regression equation of Y on X is used to describe the variation in the values of Y for given changes in X.

REGRESSION EQUATION OF Y ON X

The regression equation of Y on X is expressed as follows:

Y = a + bX

In this equation 'Y is a dependent variable' and 'X is independent variable'.

If the values of the constants 'a' and 'b' are obtained, the line is completely determined. These values are provided by the method of least squares which states that the line should be drawn through the plotted points in such a manner that the sum of the squares of the deviations of the actual Y values from the computed Y values is the least, or in other words, in order to obtain a line which fits the points best $\sum(Y - Y_c)^2$, should be minimum. Such a line is known as the line of 'best fit'.

A straight line fitted by least squares has the following characteristics:

1. It gives the best fit to the data in the sense that it makes the sum of the squared deviations from the line, $\sum(Y - Y_c)^2$, smaller than they would be from any other straight line. This property accounts for the 'Least Squares'.

2. The deviations above the line equal those below the line, on the average. This means that the total of the positive and negative deviations is zero, or $\sum(Y - Y_c)^2 = 0$.

3. The straight line goes through the overall mean of the data $(\bar{X}, \bar{Y})$.

4. When the data represent a sample from a large population the least squares line is a best estimate of the population regression line.

When a little algebra and differential calculus it can be shown that the following two equations, if solved simultaneously, will yield values of the parameters a and b such that the least squares requirement is fulfilled.

$$\sum Y = \text{Na} + \text{b}\sum X$$

$$\sum XY = \text{a}\sum X + \text{b}\sum X^2$$

These equations are usually called the normal equations.

EXAMPLE

Obtain the two lines of regression from the following data and estimate the blood pressure when age is 50 years. Can we also estimate the blood pressure of a person aged 20 years on the basis of this regression equation? Discuss.

Age (in years)	56	42	72	39	63	47	52	49	40	42	68	60
Blood pressure	127	112	140	118	129	116	130	125	115	120	135	133

Solution:

Let age be denoted by X and blood pressure be Y.

CALCULATION OF REGRESSION EQUATION

Age X	(X-60) d_X	d_X^2	Blood pressure Y	(Y-125) d_Y	d_Y^2	$d_X d_Y$
56	-4	16	127	2	4	-8
42	-18	324	112	-13	169	234
72	12	144	140	15	225	180
39	21	441	118	-7	49	147
63	3	9	129	4	16	12
47	-13	169	116	-9	81	117
52	-8	64	130	5	25	-40

49	-11	121	125	0	0	0
40	-20	400	115	-10	100	200
42	-18	324	120	-5	25	90
68	8	64	135	10	100	80
60	0	0	133	8	64	0
$\sum X$ = 630	$\sum d_X$ = -90	$\sum d_X^2$=2076	$\sum Y$= 1500	$\sum d_Y$ = 0	$\sum d_Y^2$ = 858	$\sum d_X d_Y$=1012

Regression equation of X on Y: $X-\bar{X} = r\frac{\sigma X}{\sigma Y}(Y-\bar{Y})$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{630}{12} = 52.5,\ \bar{Y} = \frac{\Sigma Y}{N} = \frac{1500}{12} = 125$$

$$r\frac{\sigma X}{\sigma Y} = \frac{N\sum d_X d_Y - (\sum d_X)(\sum d_Y)}{N\sum d_Y^2 - \sum(d_Y)^2} = \frac{12(1012)-(-90)(0)}{12(858)-0} = \frac{12144}{10296} = 1.179$$

X-52.5 = 1.179(Y-125)

X-52.5 = 1.179Y – 147.38

X = 1.179Y-94.88 or X = -94.88 + 1.179Y

Regression equation of Y on X: $Y-\bar{Y} = r\frac{\sigma Y}{\sigma X}(X-\bar{X})$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{1500}{12} = 125,\ \bar{X} = \frac{\Sigma X}{N} = \frac{630}{12} = 52.5$$

$$r\frac{\sigma Y}{\sigma X} = \frac{N\sum d_X d_Y - (\sum d_X)(\sum d_Y)}{N\sum d_X^2 - \sum(d_X)^2} = \frac{12(1012)-(-90)(0)}{12(2076)-(-90)^2} = \frac{12144-0}{24912-8100} = \frac{12144}{16812} = 0.722$$

Y-125 = 0.722(X-52.5)

Y-125 = 0.722X – 37.905

Y = 0.722X-37.905 or Y = 87.095+ 0.722X

We can estimate the blood pressure of a person aged 50 years from the regression equation of Y on X:

For X = 50 Y=87.095+0.722(50) = 87.095+36.1 = 123.195 = 123.2.

We can estimate the blood pressure of a person aged 20 years from the regression equation of Y on X:

For X = 20 Y=87.095+0.722(20) = 87.095+14.44 = 101.535 = 101.5.

DEVIATIONS TAKEN FROM ARITHMETIC MEANS OF X AND Y

Regression equation of X on Y: $X-\bar{X} = r\frac{\sigma X}{\sigma Y}(Y-\bar{Y})$

$\bar{X}$ is the mean of X series; $\bar{Y}$ is the mean of Y series.

$r\frac{\sigma X}{\sigma Y}$ is known as the regression coefficient of X on Y.

The regression coefficient of X on Y is denoted by the symbol b_{XY} or b_1.

It measures the change in X corresponding to a unit change in Y. When deviations are taken from the means of X and Y, the regression coefficient of X on Y is obtained as follows:

$$b_{XY} \text{ or } r\frac{\sigma_X}{\sigma_Y} = \frac{\sum XY}{\sum Y^2}$$

Regression equation of Y on X: $Y-\bar{Y} = r\frac{\sigma_Y}{\sigma_X}(X-\bar{X})$

$r\frac{\sigma_Y}{\sigma_X}$ is known as the regression coefficient of Y on X.

The regression coefficient of Y on X is denoted by the symbol b_{YX} or b_2.

It measures the change in Y corresponding to a unit change in X. When deviations are taken from the means of X and Y, the regression coefficient of Y on X is obtained as follows:

$$b_{YX} \text{ or } r\frac{\sigma_Y}{\sigma_X} = \frac{\sum XY}{\sum X^2}$$

EXAMPLE

Fit a regression line Y = a + bX by the method of least squares.

Income(X)	41	65	50	57	96	94	110	30	79	65
Expenditure(Y)	44	60	39	51	80	68	84	34	55	48

Solution:

The least square regression equation of Y on X is given by:

$$Y-\bar{Y} = b_{YX}(X-\bar{X})$$

FITTING REGRESSION EQUATION OF Y ON X

Income X	(X-68) d_X	d_X^2	Expenditure Y	(Y-56) d_Y	d_Y^2	d_Xd_Y
41	-27	729	44	-12	144	324
65	-3	9	60	4	16	-12
50	-18	324	39	-17	289	306
57	-11	121	51	-5	25	55
96	28	784	80	24	576	672
94	26	676	68	12	144	312
110	42	1764	84	28	784	1176
30	-38	1444	34	-22	484	836
79	11	121	55	-1	1	-11
65	-3	9	48	-8	64	24

$\sum X = 687$	$\sum d_X = 7$	$\sum d_X^2 = 5981$	$\sum Y = 563$	$\sum d_Y = 3$	$\sum d_Y^2 = 2527$	$\sum d_X d_Y = 3682$

$\bar{X} = \frac{\sum X}{N} = \frac{687}{10} = 68.7$; $\bar{Y} = \frac{\sum Y}{N} = \frac{563}{10} = 56.3$

$$b_{YX} = \frac{N\sum d_X d_Y - \sum d_X \sum d_Y}{N\sum d_X^2 - (\sum d_X)^2} = \frac{10(3682)-(7)(3)}{10(5981)-7^2} = \frac{36820-21}{59810-49} = \frac{36799}{59761} = 0.616$$

Substituting these values in the above equation, we get

Y-56.3 = 0.616(X-68.7)

Y = 0.616X + 13.98 or Y = 13.98 + 0.616X

This is the required least square regression equation of Y on X.

UNIT 5

INDEX NUMBERS

INTRODUCTION

Historically, the first index was constructed in 1764 to compare the Italian price index in 1750 with a price level 1500. Though originally developed for measuring the effect of change in prices, index numbers have today become one of the most widely used statistical devices and there is hardly any field where they are not used. Newspapers headline the fact that prices are going up or down, that industrial production is rising or falling, that imports are increasing or decreasing, that crimes are rising in a particular period compared to the previous period as disclosed by index numbers. They are used to feel the pulse of the economy and they have come to be used as indicators of inflationary or deflationary tendencies. In fact, they are described as ‘barometers of economic activity’, that is, if one wants to get an idea as to what is happening to an economy, he should look to important indices like the index number of industrial production, agricultural production, business activity, etc.

Some prominent definitions of Index numbers are given below:

1. “Index numbers are devices for measuring differences in the magnitude of a group of related variable”

2. “An index number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographic location and other characteristics such as income, profession, etc.

3. “In its simplest form an index number is the ratio of two index numbers expressed as a percent. An index number is a statistical measure – a measure designed to show changes in one variable or in a group of related variables over time, or with respect to geographic location, or in terms of some other characteristics”.

4. "In its simplest form, an index number is nothing more than a relative number, or a relative which expresses the relationship between two figures, where one of the figures is used as a base".

5. "Generally speaking, index numbers measure the size or magnitude of some object at a particular point in time as a percentage of some base or reference object in the past".

CLASSIFICATION OF INDEX NUMBERS

Index numbers may be classified in terms of what they measure. In economics and business the classifications are:

(1) price

(2) quantity

(3) value

(4) Special purpose

METHODS OF CONSTRUCTING INDEX NUMBERS

A large number of formulae have been devised for constructing index numbers. Broadly speaking, they can be grouped under two heads:

(a) Unweighted indices; and

(b) Weighted indices.

The following chart illustrates the various methods:

Index Number

(1) Unweighted

(i) Simple Aggregative

(ii) Simple Average of Relatives

(2) Weighted

(i) Weighted Aggregative

(ii) Weighted Average of Relatives

UNWEIGHTED INDEX NUMBERS

(i) **SIMPLE AGGREGATIVE METHOD**

This is the simplest method of constructing index numbers. When this method is used to construct a price index the total of current year prices for the various commodities in question is divided by the total of base year prices and the quotient is multiplied by 100.

EXAMPLE

For the data given below, calculate the index number by taking:

(i) 2009 as the base year;

(ii) 2016 as the base year; and

(iii) 2009 to 2011 as the base period.

Year	Price of Commodity X	Year	Price of Commodity X
2009	4	2014	10
2010	5	2015	9
2011	6	2016	10
2012	7	2017	11
2013	8		

Solution:

(i) INDEX NUMBERS TAKING 2009 AS THE BASE YEAR

Year	Price of Commodity X	Index Numbers (2009=100)	Year	Price of Commodity X	Index Numbers (2009=100)
2009	4	100	2014	10	$\frac{10}{4}$x100=250
2010	5	$\frac{5}{4}$x100=125	2015	9	$\frac{9}{4}$x100=225
2011	6	$\frac{6}{4}$x100=150	2016	10	$\frac{10}{4}$x100=250
2012	7	$\frac{7}{4}$x100=175	2017	11	$\frac{11}{4}$x100=275
2013	8	$\frac{8}{4}$x100=200			

(ii) INDEX NUMBERS TAKING 2016 AS THE BASE YEAR

Year	Price of Commodity X	Index Numbers (2016=100)	Year	Price of Commodity X	Index Numbers (2016=100)
2009	4	$\frac{4}{10}$x100=40	2014	10	$\frac{10}{10}$x100=100
2010	5	$\frac{5}{10}$x100=50	2015	9	$\frac{9}{10}$x100=90
2011	6	$\frac{6}{10}$x100=60	2016	10	$\frac{10}{10}$x100=100
2012	7	$\frac{7}{10}$x100=70	2017	11	$\frac{11}{10}$x100=110
2013	8	$\frac{8}{10}$x100=80			

(iii) INDEX NUMBERS TAKING 2009 TO 2011 AS THE BASE PERIOD

When 2009 to 2011 is to be taken as a base, it means we have to take an average of 2009, 2010 and 2011.

$$\text{Average} = \frac{4+5+6}{3} = 5$$

Hence 2010 will be taken as 100.

Year	Price of Commodity 'X'	Index Numbers (2010 = 100)	Year	Price of Commodity 'X'	Index Numbers (2010 = 100)
2009	4	$\frac{4}{5}$ x 100 =80	2014	10	$\frac{10}{5}$ x 100 =200
2010	5	$\frac{5}{5}$ x 100 =100	2015	9	$\frac{9}{5}$ x 100 =180
2011	6	$\frac{6}{5}$ x 100 =120	2016	10	$\frac{10}{5}$ x 100 =200
2012	7	$\frac{7}{5}$ x 100 =140	2017	11	$\frac{11}{5}$ x 100 =220
2013	8	$\frac{8}{5}$ x 100 =160			

(ii) SIMPLE AVERAGE OF PRICE RELATIVES METHOD

EXAMPLE

From the following data construct an index for 2017 taking 2016 as base by the average of relatives method using (a) arithmetic mean and (b) geometric mean for averaged relatives:

Commodity	Price in 2016 (Rs.)	Price in 2017 (Rs.)
A	50	70
B	40	60
C	80	90
D	110	120
E	20	20

SOLUTION

(a) INDEX NUMBERS USING ARITHMETIC MEAN OF PRICE RELATIVES

Commodity	Price in 2016(Rs.) P_0	Price in 2017(Rs.) P_1	Price Relatives $\frac{P_1}{P_0} \times 100$
A	50	70	140.0
B	40	60	150.0
C	80	90	112.5
D	110	120	109.1
E	20	20	100.0
			$\sum \frac{P_1}{P_0} \times 100 = 611.6$

$$P_{01} = \frac{\sum \frac{P1}{P0} \times 100}{N} = \frac{611.6}{5} = 122.32$$

(b) INDEX NUMBERS USING GEOMETRIC MEAN OF PRICE RELATIVES

Commodity	Price in 2016(Rs.) P_0	Price in 2017(Rs.) P_1	Price Relatives P	Log P
A	50	70	140.0	2.1461
B	40	60	150.0	2.1761
C	80	90	112.5	2.0512
D	110	120	109.1	2.0378
E	20	20	100.0	2.0000
				$\sum \log P = 10.4112$

$$P_{01} = \text{Antilog}\left[\frac{\sum \log P}{N}\right] = \text{Antilog}\left[\frac{10.4112}{5}\right] = \text{Antilog } 2.0822 = 120.9$$

WEIGHTED INDEX NUMBERS

(i) WEIGHTED AGGREGATIVE INDICES

- Laspeyres Method
- Paasche Method
- Dorbish and Bowley’s Method
- Fisher’s Ideal Method
- Marshall Edgeworth Method and
- Kelly’s Method

EXAMPLE

Construct index numbers of price from the following data by applying:

- Laspeyres Method
- Paasche Method
- Dorbish and Bowley’s Method
- Fisher’s Ideal Method
- Marshall Edgeworth Method

Commodity	2010		2011	
	Price	Quantity	Price	Quantity
A	20	8	40	6
B	50	10	60	5
C	40	15	50	15
D	20	20	20	25

SOLUTION

CALCULATION OF VARIOUS INDICES

Commodity	2010	2011				
	Price Qty p_0 q_0	Price Qty p_1 q_1	p_1 q_0	p_0 q_0	p_1 q_1	p_0 q_1
A	20 8	40 6	320	160	240	120
B	50 10	60 5	600	500	300	250
C	40 15	50 15	750	600	750	600
D	20 20	20 25	400	400	500	500
			$\sum p_1q_0 = 2070$	$\sum p_0q_0 = 1660$	$\sum p_1q_1 = 1790$	$\sum p_0q_1 = 1470$

1. Laspeyre's Method: $P_{01} = \frac{\sum p_1q_0}{\sum p_0q_0} \times 100 = 124.70$

2. Paasche's Method: $P_{01} = \frac{\sum p_1q_1}{\sum p_0q_1} \times 100 = 121.77$

3. Bowley's Method: $P_{01} = \frac{\frac{\sum p1q0}{\sum p0q0} + \frac{\sum p1q1}{\sum p0q1}}{2} \times 100 = 123.5$

4. Fisher's Ideal Method: $P_{01} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} + \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100 = 123.5$

5. Marshall Edgeworth Method: $P_{01} = \frac{\sum p_1(q_0+q_1)}{\sum p_0(q_0+q_1)} \times 100 = 123.32$

TESTS OF ADEQUACY OF INDEX NUMBER FORMULAE

The following tests are suggested for choosing an appropriate index.

- Unit Test
- Time Reversal Test
- Factor Reversal Test
- Circular Test

1. UNIT TEST

The unit test requires that the formula for constructing an index should be independent of the units in which, or for which, prices and quantities are quoted. Except for the simple aggregative index all other formulae discussed in this chapter satisfy this test.

2. TIME REVERSAL TEST

Prof. Irving Fisher has made a careful study of the various proposals for computing index numbers and has suggested various tests to be applied to any formula to indicate whether or not it is satisfactory. The two most important of these he calls the time reversal test and the factor reversal test.

Time reversal test is a test to determine whether a given method will work both ways in time, forward and backward. In the words of Fisher, ‘ The test is that the formula for calculating the index number should be such that it will give the same ratio between one point of comparison and the other, no matter which of the two is taken as base’.

3. FACTOR REVERSAL TEST

Another test suggested by Fisher is known as Factor reversal test. It holds that the product of a price index and the quantity index should be usual to the corresponding value index. In the words of Fisher, ‘ Just as each formula should permit the interchange of the two times without giving inconsistent results, so it ought to permit interchanging the prices and quantities without giving inconsistent result, i.e., the two results multiplied together should give the true value ratio’.

4. CIRCULAR TEST

This test is just an extension of the time reversal test. The test requires that if an index is constructed for the year a on base year b, and for the year b on base year c, we ought to get the same result as if we calculated direct an index for a on base year c without going through b as an intermediary.

THE CHAIN INDEX NUMBERS

In constructing a chain index following steps are desirable:

(i) Express the figures for each year as percentages of the preceding year. The results so obtained are called link relatives.

(ii) Chain together these percentages by successive multiplication to form a chain index. Chain index of any year is the average link relative of that year multiplied by chain index of previous year divided by 100. In the form of formula:

Chain index for current year =

$$\frac{Average\ link\ relative\ of\ current\ year \times Chain\ index\ of\ previous\ year}{100}$$

CONSTRUCTION OF A CONSUMER PRICE INDEX

The following are the steps in constructing a consumer price index:

(i) Decision about the class of people for whom the index is meant: It is absolutely essential to decide clearly the class of people for whom the index is meant. That is, whether it relates to industrial workers, teachers, officers etc. The scope of the index must be clearly defined.

(ii) Conducting Family Budget Enquiry: Once the scope of the index is clearly defined the next step is to conduct a family budget enquiry covering the population group for whom the index is to be designed. The object of conducting a family budget enquiry is to determine the amount that an average family of the group included in the index spends on different items of consumption.

(iii) Obtaining Price Quotations: The collection of retail prices is a very important and , at the same time, very tedious and difficult task because such prices may vary from place to place, shop to shop and person to person.

ANALYSIS OF TIME SERIES

1. A time series is a set of statistical observations arranged in chronological order.
2. A time series consists of statistical data which are collected, recorded observed over Successive increments.
3. A time series may be defined as a collection of magnitudes belonging to different time periods, of some variable or composite of variables, such as production of steel, per capita income, gross national product, price of tobacco or index of industrial production.
4. When quantitative data are arranged in the order of their occurrence, the resulting statistical series is called a time series.
5. Time series analysis is used to detect patterns of change in statistical information over regular intervals of time. We project these patterns to arrive at an estimate for the future.

COMPONENTS OF TIME SERIES

Components or elements of time series are:
(1) Secular Trend

(2) Seasonal Variations

(3) Cyclical Variations

(4) Irregular Variations

(1) Secular Trend

The secular trend refers to the general tendency of the data to grow or decline over a long period of time, one may be interested in finding out as to what constitutes a long period of time. If we are studying the figures of sales of a firm for 2016 and 2017 and we find that in 2017 the sales have gone up, this increase cannot be called as secular trend because this is too short a period of time to conclude that the series are showing an increasing tendency. On the other hand, if we put a strong germicide into bacterial culture, and count the number of organisms still alive after each 10 seconds for 8 minutes, these 40 observations showing a general pattern would be called secular movement.

The study of secular trends allows us to describe historical patterns. The study of secular trends permit us to project past patterns or trend into the future. The trend component can be eliminated from the time series.

(2) Seasonal Variations

Seasonal variations are those periodic movements in business activity which occur regularly every year and have their origin in the nature of the year itself. The factors that cause seasonal variations are:

(i) Climate and Weather conditions: The most important factor causing seasonal variations is the climate. Changes in the climate and weather conditions such as rainfall, humidity,heat,etc., act on different products and industries differently. For example, during winter there is greater demand for woolen clothes, hot drinks, etc., whereas in summer cotton clothes, cold drinks have a greater sale.

(ii) Customs, traditions and habits: Though nature is primarily responsible for seasonal variations in time series, customs, traditions and habits also have their impact. For example, on certain occasions like Deepavali, Dussehrah, Christmas, etc., there is a big demand for sweets and also there is a large demand for cash before the festivals because people want money for shopping and gifts.

The study and measurement of seasonal patterns constitute a very important part of analysis of a time series.

(3) Cyclical Variations

The term cycle refers to the recurrent variations in time series that usually last longer than a year and it can be as many as 15 or 20 years. These variations are regular neither in amplitude nor in length. The study of cyclical variations is extremely useful in framing suitable policies for stabilizing the level of business activity, that is, for avoiding periods of booms and depressions as both are bad for an economy, particularly depression which brings about a complete disaster and shatters the economy.

(4) Irregular Variations

Irregular Variations also called erratic, accidental, random, refer to such variations in business activity which do not repeat in a definite pattern. Irregular Variations are caused by such isolated special occurrences as floods, earthquakes, strikes and wars. Sudden changes in demand or very rapid technological progress may also be included in this category.

MEASUREMENT OF TREND

The various methods that can be used for determining trend are:

- Graphic Method
- Semi-average Method
- Moving average Method
- Method of Least Squares

(1) Graphic Method

This is the simplest method of studying trend. The procedure of obtaining a straight line trend by this method is given below:

1. Plot the time series on a graph.

2. Examine carefully the direction of the trend based on the plotted informations.

3. Draw a straight line which will best fit to the data according to personal judgement. The line now shows the direction of the trend.

(2) Method of Semi-Averages

When this method is used, the given data is divided into two parts, preferably with the same number of years. For example, if we are given data from 2000 to 2017, over a period of 18 years, the two equal parts will be each nine years, from 2000 to 2008 and from 2009 to 2017. In case of odd number of years like 9.13,17, etc., two equal parts can be made simply by omitting the middle year. For example, if data are given for 19 years from 1999 to 2017, the two equal parts would be from 1999 to 2007 and from 2009 to 2017, the middle year 2008 will be omitted.

After the data have been divided into two parts, an average of each part is obtained. We thus get two points. Each point is plotted at the mid-point of the class interval covered by the respective part and then the two points are joined by a straight line which gives us the required trend line. The line can be extended downwards or upwards to get intermediate values or to predict future values.