DIFFERENTIAL EQUATIONS 6

Many physical problems, when formulated in mathematical forms, lead to **differential equations.** Differential equations enter naturally as models for many phenomena in economics, commerce, engineering etc. Many of these phenomena are complex in nature and very difficult to understand. But when they are described by differential equations, it is easy to analyse them. For example, if the rate of change of cost for x outputs is directly proportional to the cost, then this phenomenon is described by the differential equation,

$\frac{dC}{dx} = k\,C$, where C is the cost and k is constant. The solution of this differential equation is

$$C = C_0\, e^{kx} \text{ where } C = C_0 \text{ when } x = 0.$$

6.1 FORMATION OF DIFFERENTIAL EQUATIONS

A **Differential Equation** is one which involves one or more independent variables, a dependent variable and one or more of their differential coefficients.

There are two types of differential equations:

(i) **Ordinary differential equations** involving only one independent variable and derivatives of the dependent variable with respect to the independent variable.

(ii) **Partial differential equations** which involve more than one independent variable and partial derivatives of the dependent variable with respect to the independent variables.

The following are a few examples for differential equations:

1) $\left(\frac{dy}{dx}\right)^2 - 3\frac{dy}{dx} + 2y = e^x$ 2) $\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 3y = 0$

3) $\left\{1+\left(\frac{dy}{dx}\right)^2\right\}^{\frac{3}{2}} = k\frac{d^2y}{dx^2}$ 4) $x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} = 0$

5) $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$ 6) $\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = x + y$

(1), (2) and (3) are ordinary differential equations and

(4), (5) and (6) are partial differential equations.

In this chapter we shall study ordinary differential equations only.

6.1.1 Order and Degree of a Differential Equation

The order of the derivative of the highest order present in a differential equation is called the **order** of the differential equation.

For example, consider the differential equation

$$x^2\left(\frac{d^2y}{dx^2}\right)^3 + 3\left(\frac{d^3y}{dx^3}\right)^2 + 7\frac{dy}{dx} - 4y = 0$$

The orders of $\frac{d^3y}{dx^3}$, $\frac{d^2y}{dx^2}$ and $\frac{dy}{dx}$ are 3, 2 and 1 respectively. So the highest order is 3. Thus the order of the differential equation is 3.

The degree of the derivative of the highest order present in a differential equation is called the **degree** of the differential equation. Here the differential coefficients should be free from the radicals and fractional exponents.

Thus the degree of

$$x^2\left(\frac{d^2y}{dx^2}\right)^3 + 3\left(\frac{d^3y}{dx^3}\right)^2 + 7\frac{dy}{dx} - 4y = 0 \quad \text{is } 2$$

Example 1

Write down the order and degree of the following differential equations.

(i) $\left(\frac{dy}{dx}\right)^3 - 4\left(\frac{dy}{dx}\right) + y = 3e^x$ **(ii)** $\left(\frac{d^2y}{dx^2}\right)^3 + 7\left(\frac{dy}{dx}\right)^4 = 3\sin x$

(iii) $\frac{d^2x}{dy^2} + a^2x = 0$ **(iv)** $\left(\frac{dy}{dx}\right)^2 - 3\frac{d^3y}{dx^3} + 7\frac{d^2y}{dx^2} + 4\left(\frac{dy}{dx}\right) - \log x = 0$

(v) $\sqrt{1+\left(\frac{dy}{dx}\right)^2} = 4x$ **(vi)** $\left[1+\left(\frac{dy}{dx}\right)^2\right]^{\frac{2}{3}} = \frac{d^2y}{dx^2}$

(vii) $\frac{d^2y}{dx^2} - \sqrt{\frac{dy}{dx}} = 0$ **(viii)** $\sqrt{1+x^2} = \frac{dy}{dx}$

Solution :

The order and the degree respectively are,

(i) 1 ; 3	(ii) 2 ; 3	(iii) 2 ; 1	(iv) 3 ; 1
(v) 1 ; 2	(vi) 2 ; 3	(vii) 2 ; 2	(viii) 1 ; 1

Note

Before ascertaining the order and degree in (v), (vi) & (vii) we made the differential coefficients free from radicals and fractional exponents.

6.1.2 Family of curves

Sometimes a family of curves can be represented by a single equation. In such a case the equation contains an arbitrary constant c. By assigning different values for c, we get a family of curves. In this case c is called the **parameter** or **arbitrary constant** of the family.

Examples

(i) $y = mx$ represents the equation of a family of straight lines through the origin , where m is the parameter.

(ii) $x^2 + y^2 = a^2$ represents the equation of family of concentric circles having the origin as centre, where a is the parameter.

(iii) $y = mx + c$ represents the equation of a family of straight lines in a plane, where m and c are parameters.

6.1.3 Formation of Ordinary Differential Equation

Consider the equation $y = mx + \lambda$ ---------(1)
where m is a constant and λ is the parameter.

This represents one parameter family of parallel straight lines having same slope m.

Differentiating (1) with respect to x, we get, $\frac{dy}{dx} = m$

This is the differential equation representing the above family of straight lines.

Similarly for the equation $y = Ae^{5x}$, we form the differential equation $\frac{dy}{dx} = 5y$ by eliminating the arbitrary constant A.

The above functions represent one-parameter families. Each family has a differential equation. To obtain this differential equation differentiate the equation of the family with respect to x, treating the parameter as a constant. If the derived equation is free from parameter then the derived equation is the differential equation of the family.

Note

(i) The differential equation of a two parameter family is obtained by differentiating the equation of the family twice and by eliminating the parameters.

(ii) In general, the order of the differential equation to be formed is equal to the number of arbitrary constants present in the equation of the family of curves.

Example 2

Form the differential equation of the family of curves $y = A \cos 5x + B \sin 5x$ where A and B are parameters.

Solution :

Given $y = A \cos 5x + B \sin 5x$

$$\frac{dy}{dx} = -5A \sin 5x + 5B \cos 5x$$

$$\frac{d^2y}{dx^2} = -25\ (\text{A} \cos 5x) - 25\ (\text{B} \sin 5x) = -25y$$

$$\therefore \quad \frac{d^2y}{dx^2} + 25y = 0.$$

Example 3

Form the differential equation of the family of curves $y = ae^{3x} + be^{x}$ where a and b are parameters.

Solution :

$$y = ae^{3x} + be^{x} \quad \text{------------(1)}$$

$$\frac{dy}{dx} = 3ae^{3x} + be^{x} \quad \text{------------(2)}$$

$$\frac{d^2y}{dx^2} = 9ae^{3x} + be^{x} \quad \text{------------(3)}$$

$$(2) - (1) \Rightarrow \frac{dy}{dx} - y = 2ae^{3x} \quad \text{------------(4)}$$

$$(3) - (2) \Rightarrow \frac{d^2y}{dx^2} - \frac{dy}{dx} = 6ae^{3x} = 3\left(\frac{dy}{dx} - y\right) \quad \text{[using (4)]}$$

$$\Rightarrow \frac{d^2y}{dx^2} - 4\frac{dy}{dx} + 3\text{y} = 0$$

Example 4

Find the differential equation of a family of curves given by $y = a \cos (mx + b)$, a and b being arbitrary constants.

Solution :

$$y = a \cos (mx + b) \quad \text{------------(1)}$$

$$\frac{dy}{dx} = -ma \sin (mx + b)$$

$$\frac{d^2y}{dx^2} = -m^2a \cos (mx + b) = -m^2y \quad \text{[using (1)]}$$

$\therefore \quad \frac{d^2y}{dx^2} + m^2y = 0$ is the required differential equation.

Example 5

Find the differential equation by eliminating the arbitrary constants *a* and *b* from *y* = *a* tan *x* + *b* sec *x*.

Solution :

$y = a \tan x + b \sec x$

Multiplying both sides by $\cos x$ we get,

$y \cos x = a \sin x + b$

Differentiating with respect to x we get

$$y(-\sin x) + \frac{dy}{dx}\cos x = a \cos x$$

$$\Rightarrow \quad -y \tan x + \frac{dy}{dx} = a \qquad \text{-----------(1)}$$

Differentiating (1) with respect to x, we get

$$\frac{d^2y}{dx^2} - \frac{dy}{dx}\tan x - y \sec^2 x = 0$$

EXERCISE 6.1

1) Find the order and degree of the following :

(i) $x^2\frac{d^2y}{dx^2} - 3\frac{dy}{dx} + y = \cos x$ (ii) $\frac{d^3y}{dx^3} - 3\left(\frac{d^2y}{dx^2}\right)^2 + 5\frac{dy}{dx} = 0$

(iii) $\frac{d^2y}{dx^2} - \sqrt{\frac{dy}{dx}} = 0$ (iv) $\left(1 + \frac{d^2y}{dx^2}\right)^{\frac{1}{2}} = \frac{dy}{dx}$

(v) $\left(1 + \frac{dy}{dx}\right)^{\frac{1}{3}} = \frac{d^2y}{dx^2}$ (vi) $\sqrt{1 + \frac{d^2y}{dx^2}} = x\frac{dy}{dx}$

(vii) $\left(\frac{d^2y}{dx^2}\right)^{\frac{3}{2}} = \left(\frac{dy}{dx}\right)^2$ (viii) $3\frac{d^2y}{dx^2} + 5\left(\frac{dy}{dx}\right)^3 - 3y = e^x$

(ix) $\frac{d^2y}{dx^2} = 0$ (x) $\left(\frac{d^2y}{dx^2} + 1\right)^{\frac{2}{3}} = \left(\frac{dy}{dx}\right)^{\frac{1}{3}}$

2) Find the differential equation of the following

(i) $y = mx$ (ii) $y = cx - c + c^2$

(iii) $y = mx + \frac{a}{m}$, where m is arbitrary constant
(iv) $y = mx + c$ where m and c are arbitrary constants.

3) Form the differential equation of family of rectangular hyperbolas whose asymptotes are the coordinate axes.

4) Find the differential equation of all circles $x^2 + y^2 + 2gx = 0$ which pass through the origin and whose centres are on the x-axis.

5) Form the differential equation of $y^2 = 4a(x + a)$, where a is the parameter.

6) Find the differential equation of the family of curves $y = ae^{2x} + be^{3x}$ where a and b are parameters.

7) Form the differential equation for $y = a \cos 3x + b \sin 3x$ where a and b are parameters.

8) Form the diffrential equation of $y = ae^{bx}$ where a and b are the arbitrary constants.

9) Find the differential equation for the family of concentric circles $x^2 + y^2 = a^2$, a is the paramter.

6.2 FIRST ORDER DIFFERENTIAL EQUATIONS

6.2.1 Solution of a differential equation

A **solution** of a differential equation is an explicit or implicit relation between the variables which satisfies the given differential equation and does not contain any derivatives.

If the solution of a differential equation contains as many arbitrary constants of integration as its order, then the solution is said to be the **general solution** of the differential equation.

The solution obtained from the general solution by assigning particular values for the arbitrary constants, is said to be a **particular solution** of the differential equation.

For example,

Differential equation	General solutuion	Particular solution
(i) $\frac{dy}{dx} = \sec^2 x$	$y = \tan x + c$ (c is arbitrary constant)	$y = \tan x - 5$
(ii) $\frac{dy}{dx} = x^2 + 2x$	$y = \frac{x^3}{3} + x^2 + c$	$y = \frac{x^3}{3} + x^2 + 8$
(iii) $\frac{d^2y}{dx^2} - 9y = 0$	$y = Ae^{3x} + Be^{-3x}$	$y = 5e^{3x} - 7e^{-3x}$

6.2.2 Variables Separable

If it is possible to re-arrange the terms of the first order and first degree differential equation in two groups, each containing only one variable, the variables are said to be separable.

When variables are separated, the differentail equation takes the form $f(x)\,dx + g(y)\,dy = 0$ in which $f(x)$ is a function of x only and $g(y)$ is a function of y only.

Then the general solution is

$$\int f(x)\,dx + \int g(y)\,dy = c \qquad (c \text{ is a constant of integration})$$

For example, consider $x\frac{dy}{dx} - y = 0$

$$x\frac{dy}{dx} = y \quad \Rightarrow \quad \frac{dy}{y} = \frac{dx}{x} \qquad \text{(separating the variables)}$$

$\Rightarrow \quad \int \frac{dy}{y} = \int \frac{dx}{x} + k$ where k is a constant of integration.

$\Rightarrow \quad \log y = \log x + k.$

The value of k varies from $-\infty$ to ∞.

This general solution can be expressed in a more convenient form by assuming the constant of integration to be $\log c$. This is possible because $\log c$ also can take all values between $-\infty$ and ∞ as k does. By this assumption, the general solution takes the form

$$\log y - \log x = \log c \;\Rightarrow\; \log\left(\frac{y}{x}\right) = \log c$$

i.e. $\frac{y}{x} = c \quad \Rightarrow \quad y = cx$

which is an elegant form of the solution of the differential equation.

Note

(i) When y is absent, the general form of first order linear differential equation reduces to $\frac{dy}{dx} = f(x)$ and therefore the solution is $y = \int f(x)\, dx + c$

(ii) When x is absent , it reduces to $\frac{dy}{dx} = g(y)$

and in this case, the solution is $\int \frac{dy}{g(y)} = \int dx + c$

Example 6

Solve the differential equation $xdy + ydx = 0$

Solution :

$xdy + ydx = 0$, dividing by xy we get

$$\frac{dy}{y} + \frac{dx}{x} = 0. \quad \text{Then } \int \frac{dy}{y} + \int \frac{dx}{x} = c_1$$

$\therefore$ $\log y + \log x = \log c \quad \Rightarrow \quad xy = c$

Note

(i) $xdy + ydx = 0 \quad \Rightarrow d(xy) = 0 \Rightarrow xy = c$, a constant.

(ii) $d(\frac{x}{y}) = \frac{ydx - xdy}{y^2} \quad \therefore \int \frac{ydx - xdy}{y^2} = \int d\,(\frac{x}{y}) + c = \frac{x}{y} + c$

Example 7

Solve $\frac{dy}{dx} = e^{3x+y}$

Solution :

$$\frac{dy}{dx} = e^{3x} e^{y} \quad \Rightarrow \quad \frac{dy}{e^{y}} = e^{3x}\, dx$$

$$\int e^{-y}\, dy = \int e^{3x}\, dx + c$$

$\Rightarrow$ $-e^{-y} = \frac{e^{3x}}{3} + c \quad \Rightarrow \quad \frac{e^{3x}}{3} + e^{-y} = c$

Example 8

Solve $(x^2 - ay)\,dx = (ax - y^2)dy$

Solution :

Writing the equation as

$$x^2dx + y^2dy = a(xdy + ydx)$$

$\Rightarrow \quad x^2dx + y^2dy = a\, d(xy)$

$\therefore \quad \int x^2\, dx + \int y^2\, dy = a\int d\ (xy) + c$

$\Rightarrow \quad \dfrac{x^3}{3} + \dfrac{y^3}{3} = a(xy) + c$

Hence the general solution is $x^3 + y^3 = 3axy + c$

Example 9

Solve $y\,(1+x^2)^{\frac{1}{2}}\, dy + x\sqrt{1+y^2}\, dx = 0$

Solution :

$y\sqrt{1+x^2}\, dy + x\sqrt{1+y^2}\, dx = 0$ [dividing by $\sqrt{1+x^2}\ \sqrt{1+y^2}$]

$\Rightarrow \quad \dfrac{y}{\sqrt{1+y^2}}\,dy + \dfrac{x}{\sqrt{1+x^2}}\,dx = 0$

$\therefore \quad \int \dfrac{y}{\sqrt{1+y^2}}\,dy + \int \dfrac{x}{\sqrt{1+x^2}}\,dx = c_1$

$\therefore \quad \dfrac{1}{2}\int t^{-\frac{1}{2}}\, dt + \dfrac{1}{2}\int u^{-\frac{1}{2}}\, du = c$

Put $1+y^2 = t$
$2ydy = dt$
put $1+x^2 = u$
$2xdx = du$

i.e. $\quad t^{\frac{1}{2}} + u^{\frac{1}{2}} = c$ or $\sqrt{1+y^2} + \sqrt{1+x^2} = c$

Note : This problem can also be solved by using

$$\int [\, f(x)]^n\, f'(x)\, dx = \frac{[\, f(x)]^{n+1}}{n+1}$$

Example 10

Solve $(\sin x + \cos x)\, dy + (\cos x - \sin x)\, dx = 0$

Solution :

The given equation can be written as

$$dy + \frac{\cos x - \sin x}{\sin x + \cos x} dx = 0$$

$$\Rightarrow \int dy + \int \frac{\cos x - \sin x}{\sin x + \cos x} dx = c$$

$$\Rightarrow \quad y + \log(\sin x + \cos x) = c$$

Example 11

Solve $\boldsymbol{x \frac{dy}{dx} + \cos y = 0}$**, given** $\boldsymbol{y = \frac{\text{ð}}{4}}$ **when** $\boldsymbol{x = \sqrt{2}}$

Solution :

$x\, dy = -\cos y\, dx$

$\therefore \quad \int \sec y\, dy = -\int \frac{dx}{x} + k$, where k is a constant of integration.

$\log(\sec y + \tan y) + \log x = \log c$, where $k = \log c$

or $\quad x(\sec y + \tan y) = c$.

When $x = \sqrt{2}$, $y = \frac{\pi}{4}$, we have

$\sqrt{2}\,(\sec\frac{\pi}{4} + \tan\frac{\pi}{4}) = c$ or $c = \sqrt{2}\,(\sqrt{2} + 1) = 2 + \sqrt{2}$

$\therefore$ The particular solution is $x\,(\sec y + \tan y) = 2 + \sqrt{2}$

Example 12

The marginal cost function for producing *x* units is MC = 23+16*x* - 3*x*² and the total cost for producing 1 unit is Rs.40. Find the total cost function and the average cost function.

Solution :

Let C(x) be the total cost function where x is the number of units of output. Then

$$\frac{dC}{dx} = MC = 23 + 16x - 3x^2$$

$\therefore \int \frac{dC}{dx}\,dx = \int (23+16x - 3x^2)dx + k$

$C = 23x + 8x^2 - x^3 + k$, where k is a constant

At $x = 1$, $C(x) = 40$ (given)

$23(1) + 8(1)^2 - 1^3 + k = 40 \Rightarrow k = 10$

$\therefore$ Total cost function $C(x) = 23x + 8x^2 - x^3 + 10$

$$\text{Average cost function} = \frac{\text{Total cost function}}{x}$$

$$= \frac{23x + 8x^2 - x^3 + 10}{x}$$

$$\text{Average cost function} = 23 + 8x - x^2 + \frac{10}{x}$$

Example 13

What is the general form of the demand equation which has a constant elasticity of –1 ?

Solution :

Let x be the quantity demanded at price p. Then the elasticity is given by

$$\eta_d = \frac{-p}{x}\frac{dx}{dp}$$

Given $\frac{-p}{x}\frac{dx}{dp} = -1 \Rightarrow \frac{dx}{x} = \frac{dp}{p} \Rightarrow \int \frac{dx}{x} = \int \frac{dp}{p} + \log k$

$\Rightarrow \log x = \log p + \log k$, where k is a constant.

$\Rightarrow \log x = \log kp \Rightarrow x = kp \Rightarrow p = \frac{1}{k}x$

i.e. $p = cx$, where $c = \frac{1}{k}$ is a constant

Example 14

The relationship between the cost of operating a warehouse and the number of units of items stored in it is given by $\frac{dC}{dx} = ax + b$, where C is the monthly cost of operating the warehouse and x is the number of units of items in storage. Find C as a function of x if $C = C_0$ when $x = 0$.

Solution :

Given $\dfrac{dC}{dx} = ax + b \quad \therefore \; dC = (ax + b)\, dx$

$\int dC = \int (ax+b)\, dx + k$, ($k$ is a constant)

$\Rightarrow \quad C = \dfrac{ax^2}{2} + bx + k,$

when $x = 0$, $C = C_0$ $\therefore$ (1) $\Rightarrow$ $C_0 = \dfrac{a}{2}(0) + b(0) + k$

$\Rightarrow \; k = C_0$

Hence the cost function is given by

$$C = \frac{a}{2}x^2 + bx + C_0$$

Example 15

The slope of a curve at any point is the reciprocal of twice the ordinate of the point. The curve also passes through the point (4, 3). Find the equation of the curve.

Solution :

Slope of the curve at any point P(x, y) is the slope of the tangent at P(x, y)

$\therefore \; \dfrac{dy}{dx} = \dfrac{1}{2y} \quad \Rightarrow \; 2y\,dy = dx$

$\int 2y\, dy = \int dx + c \quad \Rightarrow \quad y^2 = x + c$

Since the curve passes through (4, 3), we have

$9 = 4 + c \quad \Rightarrow \; c = 5$

$\therefore$ Equation of the curve is $y^2 = x + 5$

EXERCISE 6.2

1) Solve (i) $\dfrac{dy}{dx} + \sqrt{\dfrac{1-y^2}{1-x^2}} = 0$ (ii) $\dfrac{dy}{dx} = \dfrac{1+y^2}{1+x^2}$

(iii) $\dfrac{dy}{dx} = \dfrac{y+2}{x-1}$ (iv) $x\sqrt{1+y^2} + y\sqrt{1+x^2}\,\dfrac{dy}{dx} = 0$

2) Solve (i) $\frac{dy}{dx} = e^{2x-y} + x^3 e^{-y}$ (ii) $(1-e^x)\sec^2 y\, dy + 3e^x \tan y\, dx = 0$

3) Solve (i) $\frac{dy}{dx} = 2xy + 2ax$ (ii) $x(y^2 + 1)\, dx + y(x^2 + 1)\, dy = 0$

(iii) $(x^2 - yx^2)\frac{dy}{dx} + y^2 + xy^2 = 0$

4) Solve (i) $xdy + ydx + 4\sqrt{1-x^2y^2}\, dx = 0$ (ii) $ydx - xdy + 3x^2y^2e^{x^3}dx = 0$

5) Solve (i) $\frac{dy}{dx} = \frac{y^2+4y+5}{x^2-2x+2}$ (ii) $\frac{dy}{dx} + \frac{y^2+y+1}{x^2+x+1} = 0$

6) Find the equation of the curve whose slope at the point (x, y) is $3x^2 + 2$, if the curve passes through the point (1, -1)

7) The gradient of the curve at any point (x, y) is proportional to its abscissa. Find the equation of the curve if it passes through the points (0, 0) and (1, 1)

8) Solve : $\sin^{-1}x\, dy + \frac{y}{\sqrt{1-x^2}}\, dx = 0$, given that $y = 2$ when $x = \frac{1}{2}$

9) What is the general form of the demand equation which has an elasticity of $-n$?

10) What is the general form of the demand equation which has an elasticity of $-\frac{1}{2}$?

11) The marginal cost function for producing x units is $MC = e^{3x+7}$. Find the total cost function and the average cost function, given that the cost is zero when there is no production.

6.2.3 Homogeneous differential equations

A differential equation in x and y is said to be **homogeneous** if it can be defined in the form

$$\frac{dy}{dx} = \frac{f(x,y)}{g(x,y)}$$ where $f(x, y)$ and $g(x, y)$ are

homogeneous functions of the same degree in x and y.

$$\frac{dy}{dx} = \frac{xy}{x^2+y^2}, \quad \frac{dy}{dx} = \frac{x^2+y^2}{2xy}, \quad \frac{dy}{dx} = \frac{x^2y}{x^3+y^3}$$

and $\quad \dfrac{dy}{dx} = \dfrac{\sqrt{x^2 - y^2} + y}{x}$

are some examples of first order homogeneous differential equations.

6.2.4 Solving first order homogeneous differential equations

If we put $y = vx$ then $\dfrac{dy}{dx} = v + x\dfrac{dv}{dx}$ and the differential equation reduces to variables sepaerable form. The solution is got by replacing $\dfrac{y}{x}$ for v after the integration is over.

Example 16

Solve the differential equation $(x^2 + y^2)dx = 2xydy$

Solution :

The given differential equation can be written as

$$\frac{dy}{dx} = \frac{x^2 + y^2}{2xy} \qquad \text{-------------} \ (1)$$

This is a homogeneous differential equation

Put $y = vx \quad \therefore \ \dfrac{dy}{dx} = v + x\dfrac{dv}{dx}$ ------------- (2)

Substituting (2) in (1) we get,

$$v + x\frac{dv}{dx} = \frac{x^2 + v^2x^2}{2x(vx)} = \frac{1+v^2}{2v}$$

$$x\frac{dv}{dx} = \frac{1+v^2}{2v} - v \ \Rightarrow \ x\frac{dv}{dx} = \frac{1-v^2}{2v}$$

Now, separating the variables,

$$\frac{2v}{1-v^2}dv = \frac{dx}{x} \quad \text{or} \quad \int\frac{-2v}{1-v^2} = \int\frac{-dx}{x} + c_1$$

$\log(1 - v^2) = -\log x + \log c \quad [\int\dfrac{f'(x)}{f(x)}dx = \log f(x)]$

or $\quad \log(1 - v^2) + \log x = \log c \ \Rightarrow \ (1 - v^2)\,x = c$

Replacing v by $\dfrac{y}{x}$, we get

$$\left(1-\frac{y^2}{x^2}\right)x = c \quad \text{or} \quad x^2 - y^2 = cx$$

Example 17

Solve : $(x^3 + y^3)dx = (x^2y + xy^2)\, dy$

Solution :

The given equation can be written as

$$\frac{dy}{dx} = \frac{x^3+y^3}{x^2y+xy^2} \qquad \text{-------------- (1)}$$

Put $y = vx \quad \therefore \quad \frac{dy}{dx} = v + x\frac{dv}{dx}$

$$\Rightarrow \quad v + x\frac{dv}{dx} = \frac{1+v^3}{v+v^2}$$

$$\Rightarrow \quad x\frac{dv}{dx} = \frac{1+v^3}{v+v^2} - v = \frac{1-v^2}{v(v+1)} = \frac{(1-v)(1+v)}{v(v+1)}$$

$$\int\frac{v}{1-v}\,dv = \int\frac{1}{x}\,dx + c$$

$$\Rightarrow \quad \int\frac{-v}{1-v}\,dv = -\int\frac{1}{x}\,dx + c \quad \text{or} \quad \int\frac{(1-v)-1}{1-v}\,dv = -\int\frac{1}{x}\,dx + c$$

$$\Rightarrow \quad \int\left(1+\frac{(-1)}{1-v}\right)dv = -\int\frac{1}{x}\,dx + c$$

$$\therefore \quad v + \log(1-v) = -\log x + c$$

Replacing v by $\frac{y}{x}$, we get $\frac{y}{x} + \log(x - y) = c$

Example 18

Solve $x\frac{dy}{dx} = y - \sqrt{x^2+y^2}$

Solution :

Now, $\frac{dy}{dx} = \frac{y-\sqrt{x^2+y^2}}{x}$ ------------(1)

Put $y = vx \quad \therefore \quad \frac{dy}{dx} = v + x\frac{dv}{dx}$

(1) $\Rightarrow$ $v + x\dfrac{dv}{dx} = \dfrac{vx - \sqrt{x^2 + v^2x^2}}{x} = v - \sqrt{1+v^2}$

$\therefore$ $x\dfrac{dv}{dx} = -\sqrt{1+v^2}$ or $= \dfrac{dv}{\sqrt{1+v^2}} = -\dfrac{dx}{x}$

$\Rightarrow$ $\displaystyle\int \frac{dv}{\sqrt{1+v^2}} = -\int \frac{dx}{x} + c_1$

$\Rightarrow$ $\log (v + \sqrt{1+v^2}) = -\log x + \log c$

$\log x\, (v + \sqrt{1+v^2}) = \log c$

or $x\,(v + \sqrt{1+v^2}) = c$

i.e. $x\left[\dfrac{y}{x} + \sqrt{1+\dfrac{y^2}{x^2}}\right] = c$ or $y + \sqrt{x^2+y^2} = c$

Example 19

Solve $(x + y)\, dy + (x - y)dx = 0$

Solution :

The equation is $\dfrac{dy}{dx} = -\left(\dfrac{x-y}{x+y}\right)$ ------------ (1)

Put $y = vx$ $\therefore$ $\dfrac{dy}{dx} = v + x\dfrac{dv}{dx}$

we get $v + x\dfrac{dv}{dx} = -\dfrac{x - vx}{x + vx}$ or $v + x\dfrac{dv}{dx} = -\dfrac{1-v}{1+v}$

i.e. $x\dfrac{dv}{dx} = -\left(\dfrac{1-v}{1+v} + v\right)$ or $x\dfrac{dv}{dx} = \dfrac{-(1 - v + v + v^2)}{1+v}$

$\therefore$ $\dfrac{1+v}{1+v^2}dv = -\dfrac{1}{x}dx$ or

$\displaystyle\int \frac{dv}{1+v^2}\,dv + \frac{1}{2}\int \frac{2v}{1+v^2}\,dv = \int -\frac{1}{x}\,dx + c$

$\tan^{-1}v + \dfrac{1}{2}\log (1 + v^2) = -\log x + c$

i.e. $\tan^{-1}\left(\dfrac{y}{x}\right) + \dfrac{1}{2}\log\left(\dfrac{x^2+y^2}{x^2}\right) = -\log x + c$

$$\tan^{-1}\left(\frac{y}{x}\right) + \frac{1}{2}\log(x^2+y^2) - \frac{1}{2}\log x^2 = -\log x + c$$

i.e. $$\tan^{-1}\left(\frac{y}{x}\right) + \frac{1}{2}\log(x^2+y^2) = c$$

Example 20

The net profit p and quantity x satisfy the differential equation $\frac{dp}{dx} = \frac{2p^3 - x^3}{3xp^2}$.

Find the relationship between the net profit and demand given that $p = 20$ when $x = 10$.

Solution :

$$\frac{dp}{dx} = \frac{2p^3 - x^3}{3xp^2} \qquad \text{-------------(1)}$$

is a differential equation in x and p of homogeneous type

Put $p = vx$ $\therefore$ $\frac{dp}{dx} = v + x\frac{dv}{dx}$

(1) $\Rightarrow$ $v + x\frac{dv}{dx} = \frac{2v^3-1}{3v^2}$ $\Rightarrow$ $x\frac{dv}{dx} = \frac{2v^3-1}{3v^2} - v$

$$\Rightarrow \quad x\frac{dv}{dx} = -\left[\frac{1+v^3}{3v^2}\right]$$

$$\frac{3v^2}{1+v^3}dv = -\frac{dx}{x} \quad \therefore \quad \int\frac{3v^2}{1+v^3}dv = -\int\frac{dx}{x} = k$$

$\Rightarrow$ $\log(1 + v^3) = -\log x + \log k$, where k is a constant

$$\log(1 + v^3) = \log\frac{k}{x} \quad \text{i.e.} \quad 1 + v^3 = \frac{k}{x}$$

Replacing v by $\frac{p}{x}$, we get

$$\Rightarrow \quad x^3 + p^3 = kx^2$$

But when $x = 10$, it is given that $p = 20$

$\therefore$ $(10)^3 + (20)^3 = k(10)^2$ $\Rightarrow k = 90$ $\therefore$ $x^3+p^3 = 90x^2$

$p^3 = x^2(90 - x)$ is the required relationship.

Example 21

The rate of increase in the cost C of ordering and holding as the size q of the order increases is given by the differential equation

$$\frac{d\mathrm{C}}{dq} = \frac{\mathrm{C}^2 + 2\mathrm{C}q}{q^2}$$ **. Find the relationship between C and q if C = 1 when q = 1.**

Solution :

$$\frac{d\mathrm{C}}{dq} = \frac{\mathrm{C}^2 + 2\mathrm{C}q}{q^2} \qquad \text{------------(1)}$$

This is a homogeneous equation in C and q

Put $\mathrm{C} = vq \quad \therefore \quad \frac{d\mathrm{C}}{dq} = v + q\frac{dv}{dq}$

$(1) \Rightarrow v + q\frac{dv}{dq} = \frac{v^2q^2 + 2vq^2}{q^2} = v^2 + 2v$

$\Rightarrow \quad q\frac{dv}{dq} = v^2 + v = v(v + 1) \quad \Rightarrow \quad \frac{dv}{v(v+1)} = \frac{dq}{q}$

$\Rightarrow \quad \int \frac{(v+1)-v}{v(v+1)}\,dv = \int \frac{dq}{q} + k$, k is a constant

$\Rightarrow \quad \int \frac{dv}{v} - \int \frac{dv}{v+1} = \int \frac{dq}{q} + \log k,$

$\Rightarrow \quad \log v - \log(v + 1) = \log q + \log k$

$\Rightarrow \quad \log \frac{v}{v+1} = \log qk \quad$ or $\quad \frac{v}{v+1} = kq$

Replacing v by $\frac{\mathrm{C}}{q}$ we get, $\mathrm{C} = kq(\mathrm{C} + q)$

when C = 1 and q = 1

$\mathrm{C} = kq(\mathrm{C} + q) \quad \Rightarrow \quad k = \frac{1}{2}$

$\therefore \quad \mathrm{C} = \frac{q(\mathrm{C}+q)}{2}$ is the relation between C and q

Example 22

The total cost of production y and the level of output x are related to the marginal cost of production by the equation $(6x^2 + 2y^2)\, dx - (x^2 + 4xy)\, dy = 0$. What is the relation between total cost and output if $y = 2$ when $x = 1$?

Solution :

Given $(6x^2 + 2y^2)\, dx = (x^2 + 4xy)\, dy$

$$\therefore \frac{dy}{dx} = \frac{6x^2 + 2y^2}{x^2 + 4xy} \quad \text{------------(1)}$$

is a homogeneous equation in x and y.

Put $y = vx \therefore \frac{dy}{dx} = v + x\frac{dv}{dx}$

$$(1) \Rightarrow \quad v + x\frac{dv}{dx} = \frac{6x^2 + 2y^2}{x^2 + 4xy} \quad \text{or} \quad \frac{1+4v}{6-v-2v^2}\, dv = \frac{1}{x}\, dx$$

$$\therefore \quad -\int \frac{-1-4v}{6-v-2v^2}\, dv = \int \frac{1}{x}\, dx + k, \text{ where k is a constant}$$

$$\Rightarrow \quad -\log(6-v-2v^2) = \log x + \log k = \log kx$$

$$\Rightarrow \quad \frac{1}{6-v-2v^2} = kx$$

$$\Rightarrow \quad x = c(6x^2 - xy - 2y^2) \text{ where } c = \frac{1}{k} \text{ and } v = \frac{y}{x}$$

when $x = 1$ and $y = 2$, $\quad 1 = c(6 - 2 - 8) \Rightarrow c = -\frac{1}{4}$

$$\Rightarrow \quad 4x = (2y^2 + xy - 6x^2)$$

EXERCISE 6.3

1) Solve the following differential equations

(i) $\frac{dy}{dx} = \frac{y}{x} - \frac{y^2}{x^2}$ (ii) $2\frac{dy}{dx} = \frac{y}{x} - \frac{y^2}{x^2}$

(iii) $\frac{dy}{dx} = \frac{xy - 2y^2}{x^2 - 3xy}$ (iv) $x(y - x)\frac{dy}{dx} = y^2$

(v) $\frac{dy}{dx} = \frac{y^2 - 2xy}{x^2 - 2xy}$ (vi) $\frac{dy}{dx} = \frac{xy}{x^2 - y^2}$

(vii) $(x + y)^2\, dx = 2x^2\, dy$ (viii) $x\dfrac{dy}{dx} = y + \sqrt{x^2 + y^2}$

2) The rate of increase in the cost C of ordering and holding as the size q of the order increases is given by the differential eqation $\dfrac{dC}{dq} = \dfrac{C^2 + q^2}{2Cq}$. Find the relatinship between C and q if C = 4 when $q = 2$.

3) The total cost of production y and the level of output x are related to the marginal cost of production by the equation $\dfrac{dy}{dx} = \dfrac{24x^2 - y^2}{xy}$. What is the total cost function if $y = 4$ when $x = 2$?

6.2.5 First order linear differential equation

A first order differential equation is said to be **linear** when the dependent variable and its derivatives occur only in first degree and no product of these occur.

An equation of the form $\dfrac{dy}{dx} + Py = Q$,

where P and Q are functions of x only, is called a **first order linear differential equation.**

For example,

(i) $\dfrac{dy}{dx} + 3y = x^3$; here $P = 3$, $Q = x^3$

(ii) $\dfrac{dy}{dx} + y \tan x = \cos x$, $P = \tan x$, $Q = \cos x$

(iii) $\dfrac{dy}{dx} x - 3y = xe^x$, $P = -\dfrac{3}{x}$, $Q = e^x$

(iv) $(1 + x^2)\dfrac{dy}{dx} + xy = (1+x^2)^3$, $P = \dfrac{x}{1+x^2}$, $Q = (1 + x^2)^2$

are first order linear differential equations.

6.2.6 Integrating factor (I.F)

A given differential equation may not be integrable as such. But it may become integrable when it is multiplied by a function.

Such a function is called the **integrating factor (I.F).** Hence an integrating factor is one which changes a differential equation into one which is directly integrable.

Let us show that $e^{\int P dx}$ is the integrating factor

for $\frac{dy}{dx} + Py = Q$ ---------(1)

where P and Q are function of x.

Now, $\frac{d}{dx}(ye^{\int P dx}) = \frac{dy}{dx} e^{\int P dx} + y\frac{d}{dx}(e^{\int P dx})$

$$= \frac{dy}{dx} e^{\int P dx} + ye^{\int P dx} \frac{d}{dx}\int P dx$$

$$= \frac{dy}{dx} e^{\int P dx} + ye^{\int P dx} P = (\frac{dy}{dx} + Py)e^{\int P dx}$$

When (1) is multiplied by $e^{\int P dx}$,

it becomes $(\frac{dy}{dx} + Py)\ e^{\int P dx} = Qe^{\int P dx}$

$\Rightarrow \frac{d}{dx}(ye^{\int P dx}) = Q\ e^{\int P dx}$

Integrating this, we have

$ye^{\int P dx} = \int Q e^{\int P dx}\ dx + c$ -------------(2)

So $e^{\int p dx}$ is the integrating factor of the differential equation.

Note

(i) $e^{\log f(x)} = f(x)$ when $f(x) > 0$

(ii) If Q = 0 in $\frac{dy}{dx} + Py = Q$, then the general solution is y (I.F) = c, where c is a constant.

(iii) For the differential equation $\frac{dx}{dy} + Px = Q$ where P and Q are functions of y alone, the **(I.F)** is $e^{\int P dy}$ and the solution is

$$x \text{ (I.F)} = \int Q \text{ (I.F) } dy + c$$

Example 23

Solve the equation $(1 - x^2)\dfrac{dy}{dx} - xy = 1$

Solution :

The given equation is $(1-x^2)\dfrac{dy}{dx} - xy = 1$

$$\Rightarrow \frac{dy}{dx} - \frac{x}{1-x^2}\, y = \frac{1}{1-x^2}$$

This is of the form $\dfrac{dy}{dx} + \text{P}y = \text{Q}$,

where $\text{P} = \dfrac{-x}{1-x^2}$; $\text{Q} = \dfrac{1}{1-x^2}$

$$\text{I.F} = e^{\int \text{P}dx} = e^{\int \frac{-x}{1-x^2}dx} = \sqrt{1-x^2}$$

The general solution is,

$$y\,(\text{I.F}) = \int \text{Q}\,(\text{I.F})dx + c$$

$$y\sqrt{1-x^2} = \int \frac{1}{1-x^2}\sqrt{1-x^2}\,dx + c$$

$$= \int \frac{dx}{\sqrt{1-x^2}} + c$$

$$y\sqrt{1-x^2} = \sin^{-1}x + c$$

Example 24

Solve $\dfrac{dy}{dx} + ay = e^x$ **(where** $a \neq -1$**)**

Solution :

The given equation is of the form $\dfrac{dy}{dx} + \text{P}y = \text{Q}$

Here $\text{P} = a$; $\text{Q} = e^x$

$$\therefore \quad \text{I.F} = e^{\int \text{P}dx} = e^{ax}$$

The general solution is

$$y\,(\text{I.F}) = \int \text{Q}\,(\text{I.F})dx + c$$

$$\Rightarrow \quad y\, e^{ax} = \int e^{x}\, e^{ax}\, dx + c = \int e^{(a+1)x}\, dx + c$$

$$y\, e^{ax} = \frac{e^{(a+1)x}}{a+1} + c$$

Example 25

Solve $\cos x \dfrac{dy}{dx} + y \sin x = 1$

Solution :

The given equation can be reduced to

$$\frac{dy}{dx} + y\frac{\sin x}{\cos x} = \frac{1}{\cos x} \quad \text{or} \quad \frac{dy}{dx} + y \tan x = \sec x$$

Here P = $\tan x$; Q = $\sec x$

$$\text{I.F} = e^{\int \tan x\, dx} = e^{\log \sec x} = \sec x$$

The general solution is

$$y\,(\text{I.F}) = \int \text{Q}\,(\text{I. F})\, dx + c$$

$$y \sec x = \int \sec^2 x\, dx + c$$

$$\therefore \quad y \sec x = \tan x + c$$

Example 26

A bank pays interest by treating the annual interest as the instantaneous rate of change of the principal. A man invests Rs.50,000 in the bank deposit which accrues interest, 6.5% per year compounded continuously. How much will he get after 10 years? (Given : $e^{.65}$ =1.9155)

Solution :

Let P(t) denotes the amount of money in the account at time t. Then the differential equation governing the growth of money is

$$\frac{dP}{dt} = \frac{6.5}{100}P = 0.065P \Rightarrow \int \frac{dP}{P} = \int (0.065)\, dt + c$$

$$\log_e P = 0.065t + c \quad \therefore \quad P = e^{0.065t}\, e^{c}$$

$$P = c_1\ e^{0.065t} \qquad \text{-------------}(1)$$

At $t = 0$, P = 50000.

$(1) \Rightarrow\ 50000 = c_1\, e^0$ or $c_1 = 50000$

$\therefore \quad P = 50000\ e^{0.065t}$

At $t = 10$, $P = 50000\, e^{0.065 \times 10} = 50000\ e^{0.65}$

$= 50000 \times (1.9155) = \text{Rs.}95{,}775.$

Example 27

Solve $\dfrac{dy}{dx} + y \cos x = \dfrac{1}{2}\sin 2x$

Solution :

Here $P = \cos x$; $Q = \dfrac{1}{2}\sin 2x$

$\int P\, dx = \int \cos x\, dx = \sin x$

$\text{I.F} = e^{\int P dx} = e^{\sin x}$

The general solution is

$$
\begin{aligned}
y\ (\text{I.F}) &= \int Q\,(\text{I.F})\, dx + c \\
&= \int \frac{1}{2}\sin 2x.\ e^{\sin x}\, dx + c \\
&= \int \sin x \cos x.\ e^{\sin x}\, dx + c \\
&= \int t\, e^t\, dt + c = e^t\,(t-1) + c \\
&= e^{\sin x}\,(\sin x - 1) + c
\end{aligned}
$$

Let $\sin x = t$,
then $\cos x\, dx = dt$

Example 28

A manufacturing company has found that the cost C of operating and maintaining the equipment is related to the length m of intervals between overhauls by the equation

$m^2 \dfrac{dC}{dm} + 2mC = 2$ and $C = 4$ when $m = 2$. Find the relationship between C and m.

Solution :

Given $m^2 \dfrac{dC}{dm} + 2mC = 2$ or $\dfrac{dC}{dm} + \dfrac{2C}{m} = \dfrac{2}{m^2}$

This is a first order linear differential equation of the form

$\dfrac{dy}{dx} + Py = Q$, where $P = \dfrac{2}{m}$; $Q = \dfrac{2}{m^2}$

$I.F = e^{\int P dm} = e^{\int \frac{2}{m} dm} = e^{\log m^2} = m^2$

General solution is

$C\,(I.F) = \int Q\,(I.F)\, dm + k$ where k is a constant

$Cm^2 = \int \dfrac{2}{m^2} m^2\, dm + k$

$Cm^2 = 2m + k$

When $C = 4$ and $m = 2$, we have

$16 = 4 + k \Rightarrow k = 12$

∴ The relationship between C and m is

$Cm^2 = 2m + 12 = 2(m + 6)$

Example 29

Equipment maintenance and operating costs C are related to the overhaul interval *x* by the equation

$$x^2 \frac{dC}{dx} - 10xC = -10 \text{ with } C = C_0 \text{ when } x = x_0.$$

Find C as a function of *x*.

Solution :

$x^2 \dfrac{dC}{dx} - 10xC = -10$ or $\dfrac{dC}{dx} - \dfrac{10C}{x} = -\dfrac{10}{x^2}$

This is a first order linear differential equation.

$P = -\dfrac{10}{x}$ and $Q = -\dfrac{10}{x^2}$

$\int P\, dx = \int -\dfrac{10}{x} dx = -10 \log x = \log\left(\dfrac{1}{x^{10}}\right)$

$$\text{I.F} = e^{\int P dx} = e^{\log\left(\frac{1}{x^{10}}\right)} = \frac{1}{x^{10}}$$

General solution is

$$C(\text{I.F}) = \int Q\,(\text{I.F})\, dx + k, \text{ where } k \text{ is a constant.}$$

$$\frac{C}{x^{10}} = \int \frac{-10}{x^2}\left(\frac{1}{x^{10}}\right) dx + k \quad \text{or} \quad \frac{C}{x^{10}} = \frac{10}{11}\left(\frac{1}{x^{11}}\right) + k$$

when $C = C_0 \quad x = x_0$

$$\frac{C_0}{x_0^{10}} = \frac{10}{11}\left(\frac{1}{x_0^{11}}\right) + k \Rightarrow k = \frac{C}{x_0^{10}} - \frac{10}{11x_0^{11}}$$

$\therefore$ The solution is

$$\frac{C}{x^{10}} = \frac{10}{11}\left(\frac{1}{x^{11}}\right) + \left[\frac{C}{x_0^{10}} - \frac{10}{11x_0^{11}}\right]$$

$$\Rightarrow \quad \frac{C}{x^{10}} - \frac{C}{x_0^{10}} = \frac{10}{11}\left(\frac{1}{x^{11}} - \frac{1}{x_0^{11}}\right)$$

EXERCISE 6.4

1) Solve the following differential equations

(i) $\frac{dy}{dx} + y \cot x = \operatorname{cosec} x$

(ii) $\frac{dy}{dx} - \sin 2x = y \cot x$

(iii) $\frac{dy}{dx} + y \cot x = \sin 2x$

(iv) $\frac{dy}{dx} + y \cot x = 4x \operatorname{cosec} x$, if $y = 0$ when $x = \frac{\pi}{2}$

(v) $\frac{dy}{dx} - 3y \cot x = \sin 2x$ and if $y = 2$ when $x = \frac{\pi}{2}$

(vi) $x\frac{dy}{dx} - 3y = x^2$

(vii) $\frac{dy}{dx} + \frac{2xy}{1+x^2} = \frac{1}{(1+x^2)^2}$ given that $y = 0$ when $x = 1$

(viii) $\frac{dy}{dx} - y \tan x = e^x \sec x$

(ix) $\log x \frac{dy}{dx} + \frac{y}{x} = \sin 2x$

2) A man plans to invest some amount in a small saving scheme with a guaranteed compound interest compounded continuously at the rate of 12 percent for 5 years. How much should he invest if he wants an amount of Rs.25000 at the end of 5 year period. ($e^{-0.6} = 0.5488$)

3) Equipment maintenance and operating cost C are related to the overhaul interval x by the equation $x^2\frac{dC}{dx} - (b-1)Cx = -ba$, where a, b are constants and $C = C_0$ when $x = x_0$. Find the relationship between C and x.

4) The change in the cost of ordering and holding C as quantity q is given by $\frac{dC}{dq} = a - \frac{C}{q}$ where a is a constant.

Find C as a function of q if $C = C_0$ when $q = q_0$

6.3 SECOND ORDER LINEAR DIFFERENTIAL EQUATIONS WITH CONSTANT COEFFICIENTS

The general form of linear and second order differential equation with constant coefficients is

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x).$$

We shall consider the cases where

(i) $f(x) = 0$ and $f(x) = Ke^{\lambda x}$

For example,

(i) $3\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 6y = 0$ (or) $3y'' - 5y' + 6y = 0$

(ii) $\frac{d^2y}{dx^2} - 4\frac{dy}{dx} + 3y = e^{5x}$ (or) $(D^2 - 4D + 3)y = e^{5x}$

(iii) $\frac{d^2y}{dx^2} + \frac{dy}{dx} - y = 7$ (or) $(D^2 + D - 1)y = 7$

are second order linear differential equations.

6.3.1 Auxiliary equations and Complementary functions

For the differential equation, $a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x)$, $am^2 + bm + c = 0$ is said to be the **auxiliary equation**. This is a quadratic equation in m. According to the nature of the roots m_1 and m_2 of auxiliary equation we write the complementary function (C.F) as follows.

Nature of roots	Complementary function
(i) Real and unequal ($m_1 \neq m_2$)	$Ae^{m_1x} + Be^{m_2x}$
(ii) Real and equal ($m_1 = m_2 = m$ say)	$(Ax + B)\, e^{mx}$
(iii) Complex roots ($\alpha \pm i\beta$)	$e^{\alpha x}(A\cos \beta x + B\sin \beta x)$
(In all the cases, A and B are arbitrary constants)	

6.3.2 Particular Integral (P.I)

Consider $(aD^2 + bD + c)y = e^{\lambda x}$

Let $f(D) = aD^2 + bD + c$

Case 1 : If $f(\lambda) \neq 0$ then λ is not a root of the auxiliary equation $f(m) = 0$.

Rule: P.I $= \frac{1}{f(D)}e^{\lambda x} = \frac{1}{f(\lambda)}e^{\lambda x}$.

Case 2 : If $f(\lambda) = 0$, λ satisfies the auxiliary equation $f(m) = 0$. Then we proceed as follows.

(i) Let the auxiliary equation have two distinct roots m_1 and m_2 and let $\lambda = m_1$.

Then $f(m) = a(m - m_1)(m - m_2) = a(m - \lambda)(m - m_2)$

Rule : P.I $= \frac{1}{a(D-\lambda)(D-m_2)}e^{\lambda x} = \frac{1}{a(\lambda - m_2)}\, xe^{\lambda x}$

(ii) Let the auxiliary equation have two equal roots each equal to λ. i.e. $m_1 = m_2 = \lambda$.

$\therefore$ $f(m) = a\,(m - \lambda)^2$

Rule : $\therefore$ P.I $= \dfrac{1}{a(D-\lambda)^2} e^{\lambda x} = \dfrac{1}{a}\dfrac{x^2}{2!} e^{\lambda x}$

6.3.3 The General solution

The general solution of a second order linear differential equation is y = Complementary function (C.F) + Particular integral (P.I)

Example 30

Solve $3\dfrac{d^2y}{dx^2} - 5\dfrac{dy}{dx} + 2y = 0$

Solution :

The auxiliary equation is $3m^2 - 5m + 2 = 0$

$\Rightarrow$ $(3m - 2)(m - 1) = 0$

The roots are $m_1 = \dfrac{2}{3}$ and $m_2 = 1$ (Real and distinct)

$\therefore$ The complementary function is

$$\text{C.F} = Ae^{\frac{2}{3}x} + Be^{x}$$

The general solution is

$$y = Ae^{\frac{2}{3}x} + Be^{x}$$

Example 31

Solve $(16D^2 - 24D + 9)\,y = 0$

Solution :

The auxiliary equation is $16m^2 - 24m + 9 = 0$

$$(4m - 3)^2 = 0 \Rightarrow m = \frac{3}{4}, \frac{3}{4}$$

The roots are real and equal

$\therefore$ The C.F is $(Ax + B)e^{\frac{3}{4}x}$

The general solution is $y = (Ax + B)e^{\frac{3}{4}x}$

Example 32

Solve $(D^2 - 6D + 25)\,y = 0$

Solution :

The auxiliary equation is $m^2 - 6m + 25 = 0$

$$\Rightarrow \quad m = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{6 \pm \sqrt{36 - 100}}{2} = \frac{6 \pm 8i}{2} = 3 \pm 4i$$

The roots are complex and is of the form

$\alpha \pm i\beta$ with $\alpha = 3$ and $\beta = 4$

C.F $= e^{\alpha x}$ (A cos βx + B sin βx)

$= e^{3x}$ (A cos $4x$ + B sin $4x$)

The general solution is

$y = e^{3x}$ (A cos $4x$ + B sin $4x$)

Example 33

Solve $\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 6y = e^{5x}$

Solution :

The auxiliary equation is $m^2 - 5m + 6 = 0 \Rightarrow m = 3, 2$

$\therefore$ Complementary function C. F = $Ae^{3x} + Be^{2x}$

$$\text{P. I} = \frac{1}{D^2 - 5D + 6} e^{5x} = \frac{1}{6} e^{5x}$$

$\therefore$ The general solution is

y = C.F + P. I

$$y = Ae^{3x} + Be^{2x} + \frac{e^{5x}}{6}$$

Example 34

Solve $\dfrac{d^2y}{dx^2} + 4\dfrac{dy}{dx} + 4y = 2e^{-3x}$

Solution :

The auxiliary equation is $m^2 + 4m + 4 = 0 \Rightarrow m = -2, -2$

$\therefore$ Complementary function is C. F = $(Ax + B)e^{-2x}$

$$\text{P. I} = \frac{1}{D^2 + 4D + 4} 2e^{-3x}$$

$$= \frac{1}{(-3)^2 + 4(-3) + 4} 2e^{-3x} = 2e^{-3x}$$

$\therefore$ The general solution is

y = C.F + P. I

$y = (Ax + B)\, e^{-2x} + 2e^{-3x}$

Example 35

Solve $\dfrac{d^2y}{dx^2} - 2\dfrac{dy}{dx} + 4y = 5 + 3e^{-x}$

Solution :

The auxiliary equation is $m^2 - 2m + 4 = 0$

$$\Rightarrow \quad m = \frac{2 \pm \sqrt{4-16}}{2} = \frac{2 \pm i2\sqrt{3}}{2} = 1 \pm i\sqrt{3}$$

C.F = e^x (A cos $\sqrt{3}\,x$ + B sin $\sqrt{3}\,x$)

$$\text{P. I}_1 = \frac{1}{D^2 - 2D + 4} 5\, e^{0x} = \frac{1}{4} 5\, e^{0x} = \frac{5}{4}$$

$$\text{P.I}_2 = \frac{1}{D^2 - 2D + 4} 3\, e^{-x}$$

$$= \frac{1}{(-1)^2 - 2(-1) + 4} 3e^{-x} = \frac{3e^{-x}}{7}$$

$\therefore$ The general solution is

y = C.F + P. I_1 + P.I_2

$y = e^x$ (A cos $\sqrt{3}\,x$ + B sin $\sqrt{3}\,x$) + $\dfrac{5}{4}$ + $\dfrac{3}{7} e^{-x}$

Example 36

Solve $(4D^2 - 8D + 3)y = e^{\frac{1}{2}x}$

Solution :

The auxiliary equation is $4m^2 - 8m + 3 = 0$

$m_1 = \frac{3}{2}, \; m_2 = \frac{1}{2}$

C.F $= Ae^{\frac{3}{2}x} + Be^{\frac{1}{2}x}$

P. I $= \frac{1}{4D^2 - 8D + 3} e^{\frac{1}{2}x} = \frac{1}{4(D - \frac{3}{2})(D - \frac{1}{2})} e^{\frac{1}{2}x}$

$= \frac{1}{4(\frac{1}{2} - \frac{3}{2})(D - \frac{1}{2})} e^{\frac{1}{2}x} = \frac{x}{-4} e^{\frac{1}{2}x}$

∴ The general solution is

$y = \text{C.F.} + \text{P. I}$

$y = Ae^{\frac{3}{2}x} + Be^{\frac{1}{2}x} - \frac{x}{4} e^{\frac{x}{2}}$

Example 37

Solve : $(D^2 + 10D + 25)y = \frac{5}{2} + e^{-5x}$

Solution :

The auxiliary equation is $m^2 + 10m + 25 = 0$

$\Rightarrow \quad (m + 5)^2 = 0$

$\Rightarrow \quad m = -5, \; -5$

∴ C.F $= (Ax + B)\, e^{-5x}$

P. $I_1 = \frac{1}{D^2 + 10D + 25} \frac{5}{2} e^{0x} = \frac{1}{25}\left(\frac{5}{2}\right) = \frac{1}{10}$

P.$I_2 = \frac{1}{D^2 + 10D + 25} e^{-5x} = \frac{1}{(D + 5)^2} e^{-5x}$

$= \frac{x^2}{2!} e^{-5x} = \frac{x^2}{2} (e^{-5x})$

∴ The general solution is

y = C.F + P. I_1 + P.I_2

$$y = (Ax + B)\,e^{-5x} + \frac{1}{10} + \frac{x^2}{2}\,e^{-5x}$$

Example 38

Suppose that the quantity demanded

$Q_d = 42 - 4p - 4\frac{dp}{dt} + \frac{d^2p}{dt^2}$ and quantity supplied $Q_s = -6 + 8p$ where p is the price. Find the equilibrium price for market clearance.

Solution :

For market clearance, the required condition is $Q_d = Q_s$.

$$\Rightarrow \quad 42 - 4p - 4\frac{dp}{dt} + \frac{d^2p}{dt^2} = -6 + 8p$$

$$\Rightarrow \quad 48 - 12p - 4\frac{dp}{dt} + \frac{d^2p}{dt^2} = 0$$

$$\Rightarrow \quad \frac{d^2p}{dt^2} - 4\frac{dp}{dt} - 12p = -48$$

The auxiliary equation is $m^2 - 4m - 12 = 0$

$\Rightarrow$ $m = 6\,,\ -2$

C.F. $= Ae^{6t} + Be^{-2t}$

$$\text{P. I} = \frac{1}{D^2 - 4D - 12}(-48)\,e^{0t} = \frac{1}{-12}(-48) = 4$$

$\therefore$ The general solution is

p = C.F. + P. I

$p = Ae^{6t} + Be^{-2t} + 4$

EXERCISE 6.5

1) Solve :

(i) $\frac{d^2y}{dx^2} - 10\frac{dy}{dx} + 24y = 0$ (ii) $\frac{d^2y}{dx^2} + \frac{dy}{dx} = 0$

(iii) $\frac{d^2y}{dx^2} + 4y = 0$ (iv) $\frac{d^2y}{dx^2} + 4\frac{dy}{dx} + 4y = 0$

2) Solve :

(i) $(3D^2 + 7D - 6)y = 0$ (ii) $(4D^2 - 12D + 9)y = 0$

(iii) $(3D^2 - D + 1)y = 0$

3) Solve :

(i) $(D^2 - 13D + 12)\, y = e^{-2x} + 5e^x$

(ii) $(D^2 - 5D + 6)\, y = e^{-x} + 3e^{-2x}$

(iii) $(D^2 - 14D + 49)\, y = 3 + e^{7x}$

(iv) $(15D^2 - 2D - 1)\, y = e^{\frac{x}{3}}$

4) Suppose that $Q_d = 30 - 5P + 2\frac{dP}{dt} + \frac{d^2P}{dt^2}$ and $Q_s = 6 + 3P$. Find the equilibrium price for market clearance.

EXERCISE 6.6

Choose the correct answer

1) The differential equation of straight lines passing through the origin is

(a) $x\frac{dy}{dx} = y$ (b) $\frac{dy}{dx} = \frac{x}{y}$ (c) $\frac{dy}{dx} = 0$ (d) $x\frac{dy}{dx} = \frac{1}{y}$

2) The degree and order of the differential equation

$\frac{d^2y}{dx^2} - 6\sqrt{\frac{dy}{dx}} = 0$ are

(a) 2 and 1 (b) 1 and 2 (c) 2 and 2 (d) 1 and 1

3) The order and degree of the differential equation

$\left(\frac{dy}{dx}\right)^2 - 3\frac{d^3y}{dx^3} + 7\frac{d^2y}{dx^2} + \frac{dy}{dx} = x + \log x$ are

(a) 1 and 3 (b) 3 and 1 (c) 2 and 3 (d) 3 and 2

4) The order and degree of $\left[1 + \left(\frac{dy}{dx}\right)^2\right]^{\frac{2}{3}} = \frac{d^2y}{dx^2}$ are

(a) 3 and 2 (b) 2 and 3 (c) 3 and 3 (d) 2 and 2

5) The solution of $x\ dy + y\ dx = 0$ is

(a) $x + y = c$ (b) $x^2 + y^2 = c$ (c) $xy = c$ (d) $y = cx$

6) The solution of $x\ dx + y\ dy = 0$ is

(a) $x^2 + y^2 = c$ (b) $\frac{x}{y} = c$ (c) $x^2 - y^2 = c$ (d) $xy = c$

7) The solution of $\frac{dy}{dx} = e^{x-y}$ is

(a) $e^y e^x = c$ (b) $y = \log ce^x$

(c) $y = \log(e^x + c)$ (d) $e^{x+y} = c$

8) The solution of $\frac{dp}{dt} = ke^{-t}$ (k is a constant) is

(a) $c - \frac{k}{e^t} = p$ (b) $p = ke^t + c$

(c) $t = \log \frac{c-p}{k}$ (d) $t = \log_c p$

9) In the differential equation $(x^2 - y^2)\ dy = 2xy\ dx$, if we make the subsititution $y = vx$ then the equation is transformed into

(a) $\frac{1+v^2}{v+v^3}\ dv = \frac{dx}{x}$ (b) $\frac{1-v^2}{v(1+v^2)}\ dv = \frac{dx}{x}$

(c) $\frac{dv}{v^2-1} = \frac{dx}{x}$ (d) $\frac{dv}{1+v^2} = \frac{dx}{x}$

10) When $y = vx$ the differential equation $x\frac{dy}{dx} = y + \sqrt{x^2+y^2}$ reduces to

(a) $\frac{dv}{\sqrt{v^2-1}} = \frac{dx}{x}$ (b) $\frac{vdv}{\sqrt{v^2+1}} = \frac{dx}{x}$

(c) $\frac{dv}{\sqrt{v^2+1}} = \frac{dx}{x}$ (d) $\frac{vdv}{\sqrt{1-v^2}} = \frac{dx}{x}$

11) The solution of the equation of the type $\frac{dy}{dx} + Py = 0$, (P is a function of x) is given by

(a) $y e^{\int P dx} = c$ (b) $y \int P dx = c$

(c) $x e^{\int P dx} = y$ (d) $y = cx$

12) The solution of the equation of the type $\frac{dx}{dy}$ + Px = Q (P and Q are functions of y) is

(a) $y = \int Q\, e^{\int P dx}\, dy + c$ (b) $y\, e^{\int P dx} = \int Q\, e^{\int P dx}\, dx + c$

(c) $x\, e^{\int P dy} = \int Q\, e^{\int P dy}\, dy + c$ (d) $x\, e^{\int P dy} = \int Q\, e^{\int P dx}\, dx + c$

13) The integrating factor of $x\frac{dy}{dx} - y = e^x$ is

(a) $\log x$ (b) $e^{\frac{-1}{x}}$ (c) $\frac{1}{x}$ (d) $\frac{-1}{x}$

14) The integrating factor of $(1 + x^2)\frac{dy}{dx} + xy = (1 + x^2)^3$ is

(a) $\sqrt{1+x^2}$ (b) $\log(1 + x^2)$ (c) $e^{\tan^{-1}x}$ (d) $\log^{(\tan^{-1}x)}$

15) The integrating factor of $\frac{dy}{dx} + \frac{2y}{x} = x^3$ is

(a) $2\log x$ (b) e^{x^2} (c) $3\log(x^2)$ (d) x^2

16) The complementary function of the differential equation $(D^2 - D)\, y = e^x$ is

(a) $A + B\, e^x$ (b) $(Ax + B)e^x$ (c) $A + Be^{-x}$ (d) $(A+Bx)e^{-x}$

17) The complementary function of the differential equation $(D^2 - 2D + 1)y = e^{2x}$ is

(a) $Ae^x + Be^{-x}$ (b) $A + Be^x$ (c) $(Ax + B)e^x$ (d) $A+Be^{-x}$

18) The particular integral of the differential equation $\frac{d^2y}{dx^2} - 5\frac{dy}{dx} + 6y = e^{5x}$ is

(a) $\frac{e^{5x}}{6}$ (b) $\frac{xe^{5x}}{2!}$ (c) $6e^{5x}$ (d) $\frac{e^{5x}}{25}$

19) The particular integral of the differential equation $\frac{d^2y}{dx^2} - 6\frac{dy}{dx} + 9y = e^{3x}$ is

(a) $\frac{e^{3x}}{2!}$ (b) $\frac{x^2e^{3x}}{2!}$ (c) $\frac{xe^{3x}}{2!}$ (d) $9e^{3x}$

20) The solution of $\frac{d^2y}{dx^2} - y = 0$ is

(a) $(A + B)e^x$ (b) $(Ax + B)e^{-x}$ (c) $Ae^x + \frac{B}{e^x}$ (d) $(A+Bx)e^{-x}$

INTERPOLATION AND FITTING A STRAIGHT LINE 7

7.1 INTERPOLATION

Interpolation is the art of reading between the lines in a table. It means insertion or filling up intermediate values of a function from a given set of values of the function. The following table represents the population of a town in the decennial census.

Year	:	1910	1920	1930	1940	1950
Population (in thousands)	:	12	15	20	27	39

Then the process of finding the population for the year 1914, 1923, 1939, 1947 etc. with the help of the above data is called **interpolation.** The process of finding the population for the year 1955, 1960 etc. is known as extrapolation.

The following assumptions are to be kept in mind for interpolation :

(i) The value of functions should be either in increasing order or in decreasing order.

(ii) The rise or fall in the values should be uniform. In other words that there are no sudden jumps or falls in the value of function during the period under consideration.

The following methods are used in interpolation :

1) Graphic method, 2) Algebraic method

7.1.1 Graphic method of interpolation

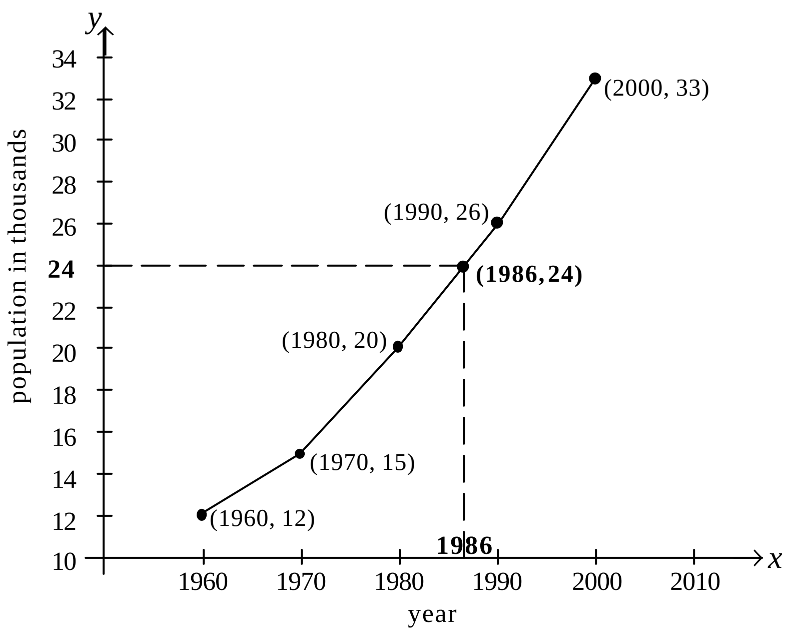
Let $y = f(x)$, then we can plot a graph between different values of x and corresponding values of y. From the graph we can find the value of y for given x.

Example 1

From the following data, estimate the population for the year 1986 graphically.

Year	**:**	**1960**	**1970**	**1980**	**1990**	**2000**
Population (in thousands)	**:**	**12**	**15**	**20**	**26**	**33**

Solution :



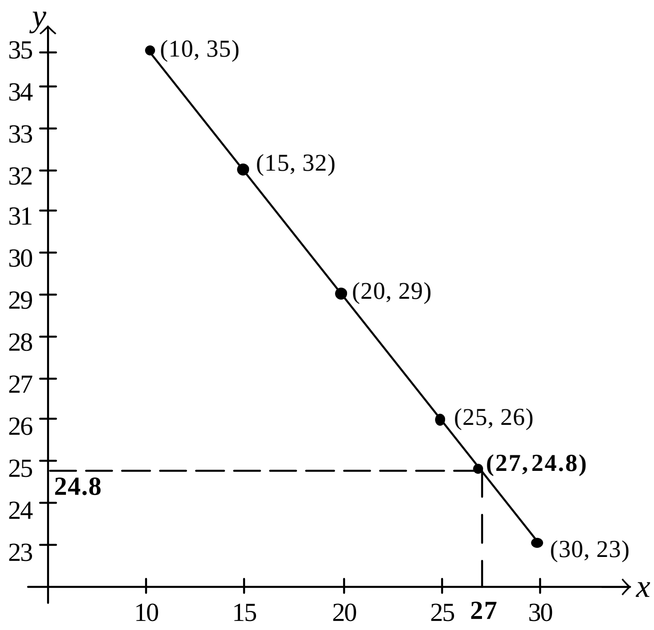
From the graph, it is found that the population for 1986 was 24 thousands

Example 2

Using graphic method, find the value of *y* when *x* = 27, from the following data.

x	**:**	**10**	**15**	**20**	**25**	**30**
y	**:**	**35**	**32**	**29**	**26**	**23**

Solution :



The value of y when $x = 27$ is 24.8

7.1.2 Algebraic methods of interpolation

The mathematical methods of interpolation are many. Of these we are going to study the following methods:

(i) Finite differences

(ii) Gregory-Newton's formula

(iii) Lagrange's formula

7.1.3 Finite differences

Consider the arguments $x_0, x_1, x_2, \ldots x_n$ and the entries $y_0, y_1, y_2, \ldots, y_n$. Here $y = f(x)$ is a function used in interpolation.

Let us assume that the x-values are in the increasing order and equally spaced with a space-length h.

Then the values of x may be taken to be x_0, $x_0 + h$, $x_0 + 2h$, ... $x_0 + nh$ and the function assumes the values $f(x_0)$, $f(x_0+h)$, $f(x_0 + 2h)$, ..., $f(x_0 + nh)$

Forward difference operator

For any value of x, the forward difference operator Δ(delta) is defined by

$$\Delta f(x) = f(x+h) - f(x).$$

In particular, $\Delta y_0 = \Delta f(x_0) = f(x_0+h) - f(x_0) = y_1 - y_0$

$\Delta f(x)$, $\Delta[f(x+h)]$, $\Delta[f(x+2h)]$, ... are the first order differences of $f(x)$.

Consider $\Delta^2 f(x) = \Delta[\Delta\{f(x)\}]$

$$= \Delta[f(x+h) - f(x)]$$

$$= \Delta[f(x+h)] - \Delta[f(x)]$$

$$= [f(x+2h) - f(x+h)] - [f(x+h) - f(x)]$$

$$= f(x+2h) - 2f(x+h) + f(x).$$

$\Delta^2 f(x)$, $\Delta^2 [f(x+h)]$, $\Delta^2 [f(x+2h)]$ ... are the second order differences of $f(x)$.

In a similar manner, the higher order differences $\Delta^3 f(x)$, $\Delta^4 f(x)$,...$\Delta^n f(x)$, ... are all defined.

Backward difference operator

For any value of x, the backward difference operator ∇(nabla) is defined by

$$\nabla f(x) = f(x) - f(x - h)$$

In particular, $\nabla y_n = \nabla f(x_n) = f(x_n) - f(x_n - h) = y_n - y_{n-1}$

$\nabla f(x)$, $\nabla[f(x+h)]$, $\nabla[f(x+2h)]$, ... are the first order differences of $f(x)$.

Consider $\nabla^2 f(x) = \nabla[\nabla\{f(x)\}] = \nabla[f(x) - f(x-h)]$

$$= \nabla[\mathrm{f}(x)] - \nabla [f(x - h)]$$

$$= f(x) - 2f(x - h) + f(x-2h)$$

$\nabla^2 f(x)$, $\nabla^2 [f(x+h)]$, $\nabla^2 [f(x+2h)]$ ... are the second order differences of $f(x)$.

In a similar manner the higher order backward differences $\nabla^3 f(x)$, $\nabla^4 f(x)$,...$\nabla^n f(x)$, ... are all defined.

Shifting operator

For any value of x, the shifting operator E is defined by

$$E[f(x)] = f(x+h)$$

In particular, $E(y_0) = E[f(x_0)] = f(x_0+h) = y_1$

Further, $E^2 [f(x)] = E[E\{f(x)] = E[f(x+h)] = f(x+2h)$

Similarly $E^3[f(x)] = f(x+3h)$

In general $E^n [f(x)] = f(x+nh)$

The relation between D and E

We have $\Delta f(x) = f(x+h) - f(x)$

$= E\, f(x) - f(x)$

$\Delta f(x) = (E - 1)\, f(x)$

$\Rightarrow \quad \Delta = E - 1$

i.e. $\quad E = 1 + \Delta$

Results

1) The differences of constant function are zero.

2) If $f(x)$ is a polynomial of the n^{th} degree in x, then the n^{th} difference of $f(x)$ is constant and $\Delta^{n+1} f(x) = 0$.

Example 3

Find the missing term from the following data.

x	**:**	**1**	**2**	**3**	**4**
f(x)	**:**	**100**	**--**	**126**	**157**

Solution :

Since three values of $f(x)$ are given, we assume that the polynomial is of degree two.

Hence third order differences are zeros.

$\Rightarrow \quad \Delta^3 [f(x_0)] = 0$

or $\quad \Delta^3(y_0) = 0$

$\therefore \quad (E - 1)^3 y_0 = 0 \qquad (\Delta = E - 1)$

$(E^3 - 3E^2 + 3E - 1) y_0 = 0$

$\Rightarrow \quad y_3 - 3y_2 + 3y_1 - y_0 = 0$

$157 - 3(126) + 3y_1 - 100 = 0$

$\therefore \quad y_1 = 107$

i.e. the missing term is 107

Example 4

Estimate the production for 1962 and 1965 from the following data.

Year :	**1961**	**1962**	**1963**	**1964**	**1965**	**1966**	**1967**
Production: (in tons)	**200**	**--**	**260**	**306**	**--**	**390**	**430**

Solution :

Since five values of $f(x)$ are given, we assume that polynomial is of degree four.

Hence fifth order diferences are zeros.

$\therefore \quad \Delta^5 [f(x_0)] = 0$

i.e. $\quad \Delta^5 (y_0) = 0$

$\therefore \quad (E - 1)^5 (y_0) = 0$

i.e. $\quad (E^5 - 5E^4 + 10E^3 - 10E^2 + 5E - 1) y_0 = 0$

$y_5 - 5y_4 + 10y_3 - 10y_2 + 5y_1 - y_0 = 0$

$390 - 5y_4 + 10(306) - 10(260) + 5y_1 - 200 = 0$

$\Rightarrow \quad y_1 - y_4 = -130$ --------------(1)

Since fifth order differences are zeros, we also have

$\Delta^5 [f(x_1)] = 0$

i.e. $\Delta^5 (y_1) = 0$

i.e. $(E - 1)^5 y_1 = 0$

$(E^5 - 5E^4 + 10E^3 - 10E^2 + 5E - 1)y_1 = 0$

$y_6 - 5y_5 + 10y_4 - 10y_3 + 5y_2 - y_1 = 0$

$430 - 5(390) + 10y_4 - 10(306) + 5(260) - y_1 = 0$

$\Rightarrow \quad 10y_4 - y_1 = 3280$ ------------(2)

By solving the equations (1) and (2) we get,

$y_1 = 220$ and $y_4 = 350$

$\therefore$ The productions for 1962 and 1965 are 220 tons and 350 tons respectively.

7.1.4 Derivation of Gregory - Newton's forward formula

Let the function $y = f(x)$ be a polynomial of degree n which assumes $(n+1)$ values $f(x_0), f(x_1), f(x_2)\ldots f(x_n)$, where $x_0, x_1, x_2, \ldots x_n$ are in the increasing order and are equally spaced.

Let $x_1 - x_0 = x_2 - x_1 = x_3 - x_2 = \ldots = x_n - x_{n-1} = h$ (a positive quantity)

Here $f(x_0) = y_0$, $f(x_1) = y_1$, ... $f(x_n) = y_n$

Now $f(x)$ can be written as,

$$f(x) = a_0 + a_1 (x - x_0) + a_2(x-x_0)(x-x_1) + \ldots$$
$$+ a_n(x-x_0)(x-x_1)\ldots(x-x_{n-1}) \quad \text{----------------(1)}$$

When $x = x_0$, (1) implies

$f(x_0) = a_0$ or $a_0 = y_0$

When $x = x_1$, (1) $\Rightarrow$

$f(x_1) = a_0 + a_1 (x_1 - x_0)$

i.e. $y_1 = y_0 + a_1 h$

$$\therefore \quad a_1 = \frac{y_1 - y_0}{h} \quad \Rightarrow \quad a_1 = \frac{\Delta y_0}{h}$$

When $x = x_2$, (1) $\Rightarrow$

$f(x_2) = a_0 + a_1(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1)$

$$y_2 = y_0 + \frac{\Delta y_0}{h}(2h) + a_2\,(2h)\,(h)$$

$$2h^2 a_2 = y_2 - y_0 - 2\Delta y_0$$

$$= y_2 - y_0 - 2(y_1 - y_0)$$

$$= y_2 - 2y_1 + y_0 = \Delta^2 y_0$$

$$\therefore \quad a_2 = \frac{\Delta^2 y_0}{2!\,h^2}$$

In the same way we can obtain

$$a_3 = \frac{\Delta^3 y_0}{3!\,h^3}, \quad a_4 = \frac{\Delta^4 y_0}{4!\,h^4}, \ldots, a_n = \frac{\Delta^n y_0}{n!\,h^n}$$

substituting the values of a_0, a_1, ..., a_n in (1) we get

$$f(x) = y_0 + \frac{\Delta y_0}{h}\,(x - x_0) + \frac{\Delta^2 y_0}{2!\,h^2}(x - x_0)\,(x - x_1) + \ldots$$

$$+ \frac{\Delta^n y_0}{n!\,h^n}(x - x_0)\,(x - x_1)\,\ldots\,(x - x_{n-1}) \quad \text{---------(2)}$$

Denoting $\frac{x - x_0}{h}$ by u, we get

$$x - x_0 = hu$$

$$x - x_1 = (x - x_0) - (x_1 - x_0) = hu - h = h(u-1)$$

$$x - x_2 = (x - x_0) - (x_2 - x_0) = hu - 2h = h(u-2)$$

$$x - x_3 = h\,(u - 3)$$

In general

$$x - x_{n-1} = h\{u - (n-1)\}$$

Thus (2) becomes,

$$f(x) = y_0 + \frac{u}{1!}\Delta y_0 + \frac{u(u-1)}{2!}\Delta^2 y_0 + \ldots$$

$$+ \frac{u(u-1)(u-2)\ldots(u-\overline{n-1})}{n!}\Delta^n y_0$$

where $u = \frac{x - x_0}{h}$. This is the Gregory-Newton's forward formula.

Example 5

Find y when $x = 0.2$ given that

x	:	0	1	2	3	4
y	:	176	185	194	202	212

Solution :

0.2 lies in the first interval (x_0, x_1) i.e. (0, 1). So we can use Gregory-Newton's forward interpolation formula. Since five values are given, the interpolation formula is

$$y = y_0 + \frac{u}{1!}\Delta y_0 + \frac{u(u-1)}{2!}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3!}\Delta^3 y_0$$

$$+ \frac{u(u-1)(u-2)(u-3)}{4!}\Delta^4 y_0 \text{ where } u = \frac{x - x_0}{h}$$

Here $h = 1$, $x_0 = 0$ and $x = 0.2$

$$\therefore u = \frac{0.2 - 0}{1} = 0.2$$

The forward difference table :

x	y	Dy	D²y	D³y	D⁴y
0	**176**				
		9			
1	185		**0**		
		9		**-1**	
2	194		-1		**4**
		8		3	
3	202		2		
		10			
4	212				

$$\therefore \quad y = \mathbf{176} + \frac{0.2}{1!}(\mathbf{9}) + \frac{0.2(0.2-1)}{2!}(\mathbf{0})$$

$$+ \frac{(0.2)(0.2-1)(0.2-2)}{3!}(\mathbf{-1}) + \frac{(0.2)(0.2-1)(0.2-2)(0.2-3)}{4!}(\mathbf{4})$$

$$= 176 + 1.8 - 0.048 - 0.1344$$

$$= 177.6176$$

i.e. when $x = 0.2$, $y = 177.6176$

Example 6

If y_{75} = 2459, y_{80} = 2018, y_{85} = 1180 and y_{90} = 402 find y_{82}.

Solution :

We can write the given data as follows:

x	:	75	80	85	90
y	:	2459	2018	1180	402

82 lies in the interval (80, 85). So we can use Gregory-Newton's forward interpolation formula. Since four values are given, the interpolation formula is

$$y = y_0 + \frac{u}{1!}\Delta y_0 + \frac{u(u-1)}{2!}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3!}\Delta^3 y_0$$

where $u = \frac{x - x_0}{h}$

Here h = 5, x_0 = 75 x = 82

$\therefore\ u = \frac{82-75}{5} = \frac{7}{5} = 1.4$

The forward difference table :

x	y	Dy	D^2y	D^3y
75	**2459**			
		-441		
80	2018		**-397**	
		-838		**457**
85	1180		60	
		-778		
90	402			

$$\therefore\quad y = 2459 + \frac{1.4}{1!}(-441) + \frac{1.4(1.4-1)}{2!}(-397) + \frac{1.4(1.4-1)(1.4-2)}{3!}(457)$$

$= 2459 - 617.4 - 111.6 - 25.592$

y = 1704.408 when x = 82

Example 7

From the following data calculate the value of $e^{1.75}$

x	**:**	**1.7**	**1.8**	**1.9**	**2.0**	**2.1**
e^x	**:**	**5.474**	**6.050**	**6.686**	**7.389**	**8.166**

Solution :

Since five values are given, the interpolation formula is

$$y_x = y_0 + \frac{u}{1!}\Delta y_0 + \frac{u(u-1)}{2!}\Delta^2 y_0 + \frac{u(u-1)(u-2)}{3!}\Delta^3 y_0$$

$$+ \frac{u(u-1)(u-2)(u-3)}{4!}\Delta^4 y_0$$

where $u = \dfrac{x - x_0}{h}$

Here $h = 0.1$, $x_0 = 1.7$ $x = 1.75$

$$\therefore\ u = \frac{1.75 - 1.7}{0.1} = \frac{0.05}{0.1} = 0.5$$

The forward difference table :

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
1.7	**5.474**				
		0.576			
1.8	6.050		**0.060**		
		0.636		**0.007**	
1.9	6.686		0.067		0
		0.703		0.007	
2.0	7.389		0.074		
		0.777			
2.1	8.166				

$$\therefore\ y = 5.474 + \frac{0.5}{1!}(0.576) + \frac{0.5(0.5-1)}{2!}(0.06)$$

$$+ \frac{0.5(0.5-1)(0.5-2)}{3!}(0.007)$$

$$= 5.474 + 0.288 - 0.0075 + 0.0004375$$

$\therefore\ y = 5.7549375$ when $x = 1.75$

Example 8

From the data, find the number of students whose height is between 80cm. and 90cm.

Height in cms x :	**40-60**	**60-80**	**80-100**	**100-120**	**120-140**
No. of students y :	**250**	**120**	**100**	**70**	**50**

Solution :

The difference table

	x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
Below	60	**250**				
			120			
Below	80	370		**-20**		
			100		**-10**	
Below	100	470		-30		**20**
			70		10	
Below	120	540		-20		
			50			
Below	140	590				

Let us calculate the number of students whose height is less than 90cm.

Here $x = 90$ $\quad u = \dfrac{x - x_0}{h} = \dfrac{90-60}{20} = 1.5$

$$y(90) = 250 + (1.5)(120) + \frac{(1.5)(1.5-1)}{2!}(-20)$$

$$+ \frac{(1.5)(1.5-1)(1.5-2)}{3!}(-10) + \frac{(1.5)(1.5-1)(1.5-2)(1.5-3)}{4!}(20)$$

$$= 250 + 180 - 7.5 + 0.625 + 0.46875$$

$$= 423.59 \simeq 424$$

Therefore number of students whose height is between 80cm. and 90cm. is $y(90) - y(80)$

i.e. $424 - 370 = 54$.

Example 9

Find the number of men getting wages between Rs.30 and Rs.35 from the following table

Wages	**x**	**:**	**20-30**	**30-40**	**40-50**	**50-60**
No. of men	**y**	**:**	**9**	**30**	**35**	**42**

Solution :

The difference table

	x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
Under	30	**9**			
			30		
Under	40	39		**5**	
			35		**2**
Under	50	74		7	
			42		
Under	60	116			

Let us calculate the number of men whose wages is less than Rs.35.

For $x = 35$, $u = \dfrac{x - x_0}{h} = \dfrac{35 - 30}{10} = 0.5$

By Newton's forward formula,

$$y(35) = 9 + \frac{(0.5)}{1}(30) + \frac{(0.5)(0.5-1)}{2!}(5) + \frac{(0.5)(0.5-1)(0.5-2)}{3!}(2)$$

$$= 9 + 15 - 0.6 + 0.1$$

$$= 24 \text{ (approximately)}$$

Therefore number of men getting wages between

Rs.30 and Rs.35 is $y(35) - y(30)$ i.e. $24 - 9 = 15$.

7.1.5 Gregory-Newton's backward formula

Let the function $y = f(x)$ be a polynomial of degree n which assumes $(n+1)$ values $f(x_0), f(x_1), f(x_2), \ldots, f(x_n)$ where $x_0, x_1, x_2, \ldots, x_n$ are in the increasing order and are equally spaced.

Let $x_1 - x_0 = x_2 - x_1 = x_3 - x_2 = \ldots x_n - x_{n-1} = h$ (a positive quantity)

Here $f(x)$ can be written as

$$f(x) = a_0 + a_1(x-x_n) + a_2(x-x_n)(x-x_{n-1}) + \dots$$
$$+ a_n(x-x_n)(x-x_{n-1}) \dots (x-x_1) \quad \text{-----------(1)}$$

When $x = x_n$, $(1) \Rightarrow$

$$f(x_n) = a_0 \quad \text{or} \quad a_0 = y_n$$

When $x = x_{n-1}$, $(1) \Rightarrow$

$$f(x_{n-1}) = a_0 + a_1(x_{n-1}-x_n)$$

or $\quad y_{n-1} = y_n + a_1(-h)$

or $\quad a_1 = \dfrac{y_n - y_{n-1}}{h} \Rightarrow a_1 = \dfrac{\nabla y_n}{h}$

When $x = x_{n-2}$, $(1) \Rightarrow$

$$f(x_{n-2}) = a_0 + a_1(x_{n-2} - x_n) + a_2(x_{n-2} - x_n)(x_{n-2} - x_{n-1})$$

$$y_{n-2} = y_n + \frac{\nabla y_n}{h}(-2h) + a_2(-2h)(-h)$$

$$2h^2 a_2 = (y_{n-2} - y_n) + 2\nabla y_n$$
$$= y_{n-2} - y_n + 2(y_n - y_{n-1})$$
$$= y_{n-2} - 2y_{n-1} + y_n = \nabla^2 y_n$$

$$\therefore a_2 = \frac{\nabla^2 y_n}{2!h^2}$$

In the same way we can obtain

$$a_3 = \frac{\nabla^3 y_n}{3!h^3}, \quad a_4 = \frac{\nabla^4 y_n}{4!h^4} \dots a_n = \frac{\nabla^n y_n}{n!}$$

$$\therefore \quad f(x) = y_n + \frac{\nabla y_n}{h}(x-x_n) + \frac{\nabla^2 y_n}{2!h^2}(x-x_n)(x-x_{n-1}) + \dots$$
$$+ \frac{\nabla^n y_n}{n!}(x-x_n)(x-x_{n-1}) \dots (x-x_1) \quad \text{------------(2)}$$

Further, denoting $\dfrac{x-x_n}{h}$ by u, we get

$$x-x_n = h_u$$

$x-x_{n-1} = (x-x_n)\ (x_n-x_{n-1}) = hu + h = h(u+1)$

$x-x_{n-2} = (x-x_n)\ (x_n-x_{n-2}) = hu + 2h = h(u+2)$

$x-x_{n-3} = h(u+3)$

In general

$x-x_{n-k} = h(u+k)$

Thus (2) becomes,

$$f(x) = y_n + \frac{u}{1!}\nabla y_n + \frac{u(u+1)}{2!}\nabla^2 y_n + \ldots$$

$$+ \frac{u(u+1)\ldots\{u+(n-1)\}}{n!}\nabla^n y_n \text{ where } u = \frac{x-x_n}{h}$$

This is the Gregory-Newton's backward formula.

Example 10

Using Gregory-Newton's formula estimate the population of town for the year 1995.

Year	**x**	**:**	**1961**	**1971**	**1981**	**1991**	**2001**
Population	**y**	**:**	**46**	**66**	**81**	**93**	**101**

(in thousands)

Solution :

1995 lies in the interval (1991, 2001). Hence we can use Gregory-Newton's backward interpolation formula. Since five values are given, the interpolation formula is

$$y = y_4 + \frac{u}{1!}\nabla y_4 + \frac{u(u+1)}{2!}\nabla^2 y_4 + \frac{u(u+1)(u+2)}{3!}\nabla^3 y_4$$

$$+ \frac{u(u+1)(u+2)(u+3)}{4!}\nabla^4 y_4 \text{ where } u = \frac{x-x_4}{h}$$

Here $h = 10$, $x_4 = 2001$ $x = 1995$

$$\therefore\ u = \frac{1995-2001}{10} = -0.6$$

The backward difference table :

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
1961	46				
		20			
1971	66		-5		
		15		2	
1981	81		-3		**-3**
		12		**-1**	
1991	93		**-4**		
		8			
2001	**101**				

$$\therefore \quad y = 101 + \frac{(-0.6)}{1!}(8) + \frac{(-0.6)(-0.6+1)}{2!}(-4)$$

$$+ \frac{(-0.6)(-0.6+1)(-0.6+2)}{3!}(-1) + \frac{(-0.6)(-0.6+1)(-0.6+2)(-0.6+3)}{4!}(-3)$$

$$= 101-4.8+0.48+0.056+0.1008 \quad \therefore \quad y = 96.8368$$

i.e. the population for the year 1995 is 96.837 thousands.

Example 11

From the following table, estimate the premium for a policy maturing at the age of 58

Age	**x :**	**40**	**45**	**50**	**55**	**60**
Premium	**y :**	**114.84**	**96.16**	**83.32**	**74.48**	**68.48**

Solution :

Since five values are given, the interpolation formula is

$$y = y_4 + \frac{u}{1!}\nabla y_4 + \ldots + \frac{u(u+1)(u+2)(u+3)}{4!}\nabla^4 y_4$$

where $u = \frac{58-60}{5} = -0.4$

The backward difference table :

x	y	∇y	$\nabla^2 y$	$\nabla^3 y$	$\nabla^4 y$
40	114.84				
		-18.68			
45	96.16		5.84		
		-12.84		-1.84	
50	83.32		4.00		**0.68**
		-8.84		**-1.16**	
55	74.48		**2.84**		
		-6.00			
60	**68.48**				

$\therefore \quad y \quad = 68.48 + \dfrac{(-0.4)}{1!}(-6) + \dfrac{(-0.4)(0.6)}{2}(2.84)$

$\quad + \dfrac{(-0.4)(0.6)(1.6)}{6}(-1.16) + \dfrac{(-0.4)(0.6)(1.6)(2.6)}{24}(0.68)$

$\quad = 68.48 + 2.4 - 0.3408 + 0.07424 - 0.028288$

$\therefore \quad y = 70.5851052$ i.e. $y \simeq 70.59$

$\therefore$ Premium for a policy maturing at the age of 58 is 70.59

Example 12

From the following data, find *y* when *x* = 4.5

x	**:**	**1**	**2**	**3**	**4**	**5**
y	**:**	**1**	**8**	**27**	**64**	**125**

Solution :

Since five values are given, the interpolation formula is

$$y = y_4 + \frac{u}{1!}\nabla y_4 + \ldots + \frac{u(u+1)(u+2)(u+3)}{4!}\nabla^4 y_4$$

where $u = \dfrac{x - x_4}{h}$

Here $u = \dfrac{4.5-5}{1} = -0.5$

The backward difference table :

x	***y***	**Ñ*y***	**Ñ**^{2}***y***	**Ñ**^{3}***y***	**Ñ**^{4}***y***
1	1				
		7			
2	8		12		
		19		6	
3	27		18		**0**
		37		**6**	
4	64		**24**		
		61			
5	**125**				

$\therefore \; y = 125 + \dfrac{(-0.5)}{1}(61) + \dfrac{(-0.5)(0.5)}{2}(24) + \dfrac{(-0.5)(0.5)(1.5)}{6}(6)$

$\therefore \; y = 91.125$ when $x = 4.5$

7.1.6 Lagrange's formula

Let the function $y = f(x)$ be a polynomial of degree n which assumes $(n + 1)$ values $f(x_0)$, $f(x_1)$, $f(x_2)$ ...$f(x_n)$ corresponding to the arguments x_0, x_1, x_2, ... x_n (not necessarily equally spaced).

Here $f(x_0) = y_0$, $f(x_1) = y_1$, ..., $f(x_n) = y_n$.

Then the Lagrange's formula is

$$f(x) = y_0 \frac{(x-x_1)(x-x_2)...(x-x_n)}{(x_0-x_1)(x_0-x_2)...(x_0-x_n)}$$

$$+ y_1 \frac{(x-x_0)(x-x_2)...(x-x_n)}{(x_1-x_0)(x_1-x_2)...(x_1-x_n)}$$

$$+ ... + y_n \frac{(x-x_0)(x-x_1)...(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)...(x_n-x_{n-1})}$$

Example 13

Using Lagrange's formula find the value of y when $x =$ 42 from the following table

x	**:**	**40**	**50**	**60**	**70**
y	**:**	**31**	**73**	**124**	**159**

Solution :

By data we have

$x_0 = 40$, $x_1 = 50$, $x_2 = 60$, $x_3 = 70$ and $x = 42$

$y_0 = 31$, $y_1 = 73$, $y_2 = 124$, $y_3 = 159$

Using Lagrange's formula, we get

$$y = y_0 \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}$$

$$+ y_1 \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}$$

$$+ y_2 \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}$$

$$+\ y_3 \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}$$

$$y(42) = 31\frac{(-8)(-18)(-28)}{(-10)(-20)(-30)} + 73\frac{(2)(-18)(-28)}{(10)(-10)(-20)}$$

$$+124\frac{(2)(-8)(-28)}{(20)(10)(-10)} + 159\frac{(2)(-8)(-18)}{(30)(20)(10)}$$

$$= 20.832 + 36.792 - 27.776 + 7.632$$

$$y = 37.48$$

Example 14

Using Lagrange's formula find y when $x = 4$ from the following table

x	**:**	**0**	**3**	**5**	**6**	**8**
y	**:**	**276**	**460**	**414**	**343**	**110**

Solution :

Given

$x_0 = 0$, $x_1 = 3$, $x_2 = 5$, $x_3 = 6$, $x_4 = 8$ and $x = 4$

$y_0 = 276$, $y_1 = 460$, $y_2 = 414$, $y_3 = 343$, $y_4 = 110$

Using Lagrange's formula

$$y = y_0\frac{(x-x_1)(x-x_2)(x-x_3)(x-x_4)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)(x_0-x_4)}$$

$$+\ y_1\frac{(x-x_0)(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)(x_1-x_4)}$$

$$+\ y_2\frac{(x-x_0)(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)(x_2-x_4)}$$

$$+\ y_3\frac{(x-x_0)(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)(x_3-x_4)}$$

$$+\ y_4\frac{(x-x_0)(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_0)(x_4-x_1)(x_4-x_2)(x_4-x_3)}$$

$$= 276\frac{(1)(-1)(-2)(-4)}{(-3)(-5)(-6)(-8)} + 460\frac{(4)(-1)(-2)(-4)}{(3)(-2)(-3)(-5)}$$

$$+ 414\frac{(4)(1)(-2)(-4)}{(5)(2)(-1)(-3)} + 343\frac{(4)(1)(-1)(-4)}{(6)(3)(1)(-2)}$$

$$+ 110\frac{(4)(1)(-1)(-2)}{(8)(5)(3)(2)}$$

$$= -3.066 + 163.555 + 441.6 - 152.44 + 3.666$$

$$y = 453.311$$

Example 15

Using Lagrange's formula find $y(11)$ from the following table

x	**:**	**6**	**7**	**10**	**12**
y	**:**	**13**	**14**	**15**	**17**

Solution :

Given

$x_0 = 6$, $x_1 = 7$, $x_2 = 10$, $x_3 = 12$ and $x = 11$

$y_0 = 13$, $y_1 = 14$, $y_2 = 15$, $y_3 = 17$

Using Lagrange's formula

$$= 13\frac{(4)(1)(-1)}{(-1)(-4)(-6)} + 14\frac{(5)(1)(-1)}{(1)(-3)(-5)}$$

$$+ 15\frac{(5)(4)(-1)}{(4)(3)(-2)} + 17\frac{(5)(4)(1)}{(6)(5)(2)}$$

$$= 2.1666 - 4.6666 + 12.5 + 5.6666$$

$$y = 15.6666$$

EXERCISE 7.1

1) Using Graphic method, find the value of y when $x = 42$, from the following data.

x	:	20	30	40	50
y	:	51	43	34	24

2) The population of a town is as follows.

Year	x :	1940	1950	1960	1970	1980	1990
Population (in lakhs)	y :	20	24	29	36	46	50

Estimate the population for the year 1976 graphically

3) From the following data, find f(3)

x	:	1	2	3	4	5
f(x)	:	2	5	-	14	32

4) Find the missing term from the following data.

x	:	0	5	10	15	20	25
y	:	7	11	14	--	24	32

5) From the following data estimate the export for the year 2000

Year	x	:	1999	2000	2001	2002	2003
Export (in tons)	y	:	443	–	369	397	467

6) Using Gregory-Newton's formula, find y when $x = 145$ given that

x	:	140	150	160	170	180
y	:	46	66	81	93	101

7) Using Gregory-Newton's formula, find y(8) from the following data.

x	:	0	5	10	15	20	25
y	:	7	11	14	18	24	32

8) Using Gregory-Newton's formula, calculate the population for the year 1975

Year	:	1961	1971	1981	1991	2001
Population	:	98572	132285	168076	198690	246050

9) From the following data find the area of a circle of diameter 96 by using Gregory-Newton's formula

Diameter	x	:	80	85	90	95	100
Area	y	:	5026	5674	6362	7088	7854

10) Using Gregory-Newton's formula, find y when $x = 85$

x	:	50	60	70	80	90	100
y	:	184	204	226	250	276	304

11) Using Gregory-Newton's formula, find $y(22.4)$

x	:	19	20	21	22	23
y	:	91	100	110	120	131

12) From the following data find $y(25)$ by using Lagrange's formula

x	:	20	30	40	50
y	:	512	439	346	243

13) If $f(0) = 5$, $f(1) = 6$, $f(3) = 50$, $f(4) = 105$, find $f(2)$ by using Lagrange's formula

14) Apply Lagrange's formula to find y when $x = 5$ given that

x	:	1	2	3	4	7
y	:	2	4	8	16	128

7.2 FITTING A STRAIGHT LINE

A commonly occurring problem in many fields is the necessity of studying the relationship between two (or more) variables.

For example the weight of a baby is related to its age ; the price of a commodity is related to its demand ; the maintenance cost of a car is related to its age.

7.2.1 Scatter diagram

This is the simplest method by which we can represent diagramatically a bivariate data. Suppose x and y denote respectively the age and weight of an adult male, then consider a sample of n individuals with ages $x_1, x_2, x_3, \dots x_n$ and the corresponding weights as $y_1, y_2, y_3, \dots y_n$. Plot the points

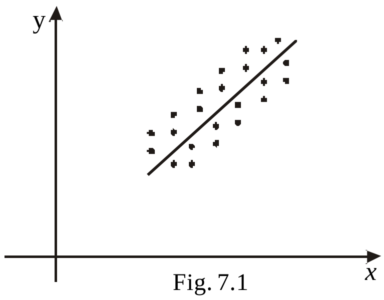


Fig. 7.1

(x_1, y_1), (x_2, y_2), (x_3, y_3), ... (x_n, y_n) on a rectangular co-ordinate system. The resulting set of points in a graph is called a **scatter diagram.**

From the scatter diagram it is often possible to visualize a smooth curve approximating the data. Such a curve is called an approximating curve. In the above figure, the data appears to be well approximated by a straight line and we say that a linear relationship exists between the two variables.

7.2.2 Principle of least squares

Generally more than one curve of a given type will appear to fit a set of data. In constructing lines it is necessary to agree on a definition of a "best fitting line".

Consider the data points (x_1, y_1), (x_2, y_2), (x_3, y_3), ... (x_n, y_n). For a given value of x, say x_1, in general there will be a difference between the value y, and the corresponding value as determined from the curve C (in Fig. 7.2)

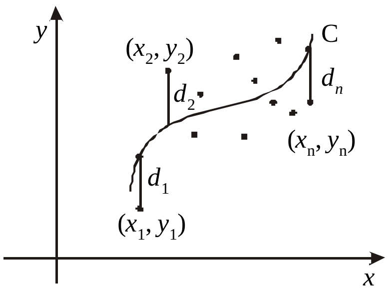


Fig. 7.2

We denote this difference by d_1, which is referred to as a deviation or error. Here d_1 may be positive, negative or zero. Similarly corresponding to the values x_2, x_3, ... x_n we obtain the deviations d_2, d_3, ... d_n.

A measure of the "goodness of fit" of the curve to the set of data is provided by the quantities d_1^2, d_2^2, ... d_n^2.

Of all the curves approximating a given set of data points, the curve having the property that $d_1^2 + d_2^2 + d_3^2 + \ldots + d_n^2$ is a minimum is the best fitting curve. If the approximating curve is a straight line then such a line is called the "line of best fit".

7.2.3 Derivation of normal equations by the principle of least squares.

Let us consider the fitting of a straight line

$$\boldsymbol{y = ax + b} \qquad \text{------------}\mathbf{(1)}$$

to set of n points (x_1, y_1), (x_2, y_2), ... (x_n, y_n).

For the different values of a and b equation (1) represents a family of straight lines. Our aim is to determine a and b so that the line (1) is the line of best fit.

Now a and b are determined by applying principle of least squares.

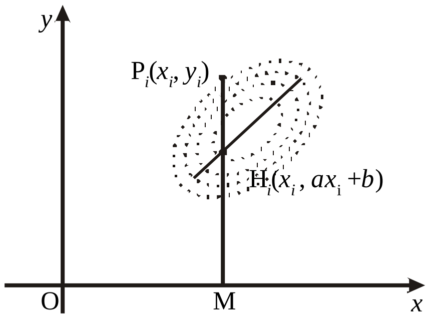


Fig. 7.3

Let P_i (x_i, y_i) be any point in the scatter diagram. Draw P_iM perpendicular to x-axis meeting the line (1) in H_i. The x-coordinate of H_i is x_i. The ordinate of H_i is $ax_i + b$.

$$P_iH_i = P_iM - H_iM$$

$$= y_i - (ax_i + b) \text{ is the deviation for } y_i.$$

According to the principle of least squares, we have to find a and b so that

$$E = \sum_{i=1}^{n} P_iH_i^2 = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2 \text{ is minimum.}$$

For maxima or minima, the partial derivatives of E with respect to a and b should vanish separately.

$$\therefore \quad \frac{\partial E}{\partial a} = 0 \Rightarrow -2\sum_{i=1}^{n} x_i[y_i - (ax_i + b)] = 0$$

$$\boldsymbol{a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i} \qquad \text{-------------}\mathbf{(2)}$$

$$\frac{\partial E}{\partial b} = 0 \Rightarrow -2\sum_{i=1}^{n} [y_i - (ax_i + b)] = 0$$

i.e., $\Sigma y_i - a\Sigma x_i - nb = 0$

$$\Rightarrow \quad a\sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i \qquad \text{-------------(3)}$$

(2) and (3) are known as the **normal equations.** Solving the normal equations we get a and b.

Note

The normal equations for the line of best fit of the form

$y = a + bx$ are

$na + b\Sigma x_i = \Sigma y_i$

$a\Sigma x_i + b\Sigma x_i^2 = \Sigma x_i y_i$

Example 16

Fit a straight line to the following

$\Sigma x = 10$, $\Sigma y = 19$, $\Sigma x^2 = 30$, $\Sigma xy = 53$ and $n = 5$.

Solution :

The line of best fit is $y = ax + b$

$\Sigma y = a\Sigma x + nb$

$\Sigma xy = a\Sigma x^2 + b\Sigma x$

$\Rightarrow \quad 10a + 5b = 19$ ------------(1)

$30a + 10b = 53$ ------------(2)

Solving (1) and (2) we get, $a = 1.5$ and $b = 0.8$

The line of best fit is $y = 1.5x + 0.8$

Example 17

In a straight line of best fit find x-intercept when $\Sigma x = 10$, $\Sigma y=16.9$, $\Sigma x^2 = 30$, $\Sigma xy = 47.4$ and $n = 7$.

Solution :

The line of best fit is $y = ax + b$

The normal equations are

$\Sigma y = a\Sigma x + nb$

$\Sigma xy = a\Sigma x^2 + b\Sigma x$

$\Rightarrow \quad 10a + 7b = 16.9$ -----------(1)

$30a + 10b = 47.4$ -----------(2)

Solving (1) and (2) we get,

$a = 1.48$ and $b = 0.3$

The line of best fit is $y = 1.48x + 0.3$

$\therefore$ The x-intercept of the line of best fit is $-\frac{0.3}{1.48}$

Example 18

Fit a straight line for the following data.

x	**:**	**0**	**1**	**2**	**3**	**4**
y	**:**	**1**	**1**	**3**	**4**	**6**

Solution :

The line of best fit is $y = ax + b$

The normal equations are

$a\Sigma x + nb = \Sigma y$ ------------(1)

$a\Sigma x^2 + b\Sigma x = \Sigma xy$ ------------(2)

Now from the data

x	y	x^2	xy
0	1	0	0
1	1	1	1
2	3	4	6
3	4	9	12
4	6	16	24
10	**15**	**30**	**43**

By substituting these values in (1) and (2) we get,

$10a + 5b = 15$ ------------(3)

$30a + 10b = 43$ ------------(4)

Solving (3) and (4) we get, $a = 1.3$ and $b = 0.4$

The line of best fit is $y = 1.3x + 0.4$.

Example 19

Fit a straight line to the following data:

x :	**4**	**8**	**12**	**16**	**20**	**24**
y :	**7**	**9**	**13**	**17**	**21**	**25**

Solution :

Take the origin at $\frac{12+16}{2} = 14$

Let $u_i = \frac{x_i - 14}{2}$ Here $n = 6$

The line of best fit is $y = au + b$

The normal equations are $a\Sigma u + nb = \Sigma y$ ---------(1)

$a\Sigma u^2 + b\Sigma u = \Sigma uy$ ----------(2)

x	***y***	***u***	***u*²**	***uy***
4	7	-5	25	-35
8	9	-3	9	-27
12	13	-1	1	-13
16	17	1	1	17
20	21	3	9	63
24	25	5	25	125
Total	**92**	**0**	**70**	**130**

On substituting the values in the normal equation (1) and (2)

$a = 1.86$ and $b = 15.33$

The line of best fit is $y = 1.86\left(\frac{x-14}{2}\right)+15.33 = 0.93x + 2.31$

Example 20

Fit a straight line to the following data.

x :	**100**	**200**	**300**	**400**	**500**	**600**
y :	**90.2**	**92.3**	**94.2**	**96.3**	**98.2**	**100.3**

Solution :

Let $u_i = \dfrac{x_i - 350}{50}$ and $v_i = y_i - 94.2$ Here n = 6.

The line of best fit is $v = au + b$

The normal equations are $a\Sigma u + nb = v$ ----------(1)

$a\Sigma u^2 + b\Sigma u = \Sigma uv$----------(2)

x	***y***	***u***	***v***	***u²***	**uv**
100	90.2	-5	-4	25	20
200	92.3	-3	-1.9	9	5.7
300	94.2	-1	0	1	0
400	96.3	1	2.1	1	2.1
500	98.2	3	4	9	12
600	100.3	5	6.1	25	30.5
Total		**0**	**63**	**70**	**70.3**

Substituting the values in (1) and (2) we get

a = 1.0043 and b = 1.05

The line of best fit is $v = 1.0043\,u + 1.05$

$\Rightarrow \quad y = 0.02x + 88.25$

EXERCISE 7.2

1) Define a scatter diagram.

2) State the principle of least squares.

3) Fit the line of best fit if $\Sigma x = 75$, $\Sigma y = 115$, $\Sigma x^2 = 1375$, $\Sigma xy = 1875$, and $n = 6$.

4) In a line of best fit find the slope and the y intercept if $\Sigma x = 10$, $\Sigma y = 25$, $\Sigma x^2 = 30$, $\Sigma xy = 90$ and $n = 5$.

5) Fit a straight line $y = ax + b$ to the following data by the method of least squares.

x	0	1	3	6	8
y	1	3	2	5	4

6) A group of 5 students took tests before and after training and obtained the following scores.

Scores before training	3	4	4	6	8
Scores after training	4	5	6	8	10

Find by the method of least squares the straight line of best fit

7) By the method of least squares find the best fitting straight line to the data given below:

x :	100	120	140	160	180	200
y :	0.45	0.55	0.60	0.70	0.80	0.85

8) Fit a straight line to the data given below. Also estimate the value y at $x = 3.5$

x :	0	1	2	3	4
y :	1	1.8	3.3	4.5	6.3

9) Find by the method of least squares, the line of best fit for the following data.

Depth of water applied (in cm)	x :	0	12	24	36	48
Average yield (tons / acre)	y :	35	55	65	80	90

10) The following data show the advertising expenses (expressed as a percentage of total expenses) and the net operating profits (expressed as a percentage of total sales) in a random sample of six drug stores.

Advertising expenses	0.4	1.0	1.3	1.5	2.0	2.8
Net operating profits	1.90	2.8	2.9	3.6	4.3	5.4

Fit a line of best fit.

11) The following data is the number of hours which ten students studied for English and the scores obtained by them in the examinations.

Hours studied	x :	4	9	10	12	14	22
Test score	y :	31	58	65	68	73	91

(i) Fit a straight line $y = ax + b$

(ii) Predict the score of the student who studied for 17 hours.

EXERCISE 7.3

Choose the correct answer

1) $\Delta f(x)$ =

(a) $f(x+h)$ (b) $f(x)-f(x+h)$

(c) $f(x+h)-f(x)$ (d) $f(x)-f(x-h)$

2) $E^2 f(x)$ =

(a) $f(x+h)$ (b) $f(x+2h)$ (c) $f(2h)$ (d) $f(2x)$

3) E =

(a) $1+\Delta$ (b) $1-\Delta$ (c) $\nabla+1$ (d) $\nabla-1$

4) $\nabla f(x+3h)$ =

(a) $f(x+2h)$ (b) $f(x+3h)-f(x+2h)$

(c) $f(x+3h)$ (d) $f(x+2h)-f(x-3h)$

5) When $h = 1$, $\Delta(x^2)$ =

(a) $2x$ (b) $2x-1$ (c) $2x+1$ (d) 1

6) The normal equations for estimating a and b so that the line $y = ax + b$ may be the line of best fit are

(a) $a\Sigma x_i^2 + b\Sigma x_i = \Sigma x_i y_i$ and $a\Sigma x_i + nb = \Sigma y_i$

(b) $a\Sigma x_i + b\Sigma x_i^2 = \Sigma x_i y_i$ and $a\Sigma x_i^2 + nb = \Sigma y_i$

(c) $a\Sigma x_i + nb = \Sigma x_i y_i$ and $a\Sigma x_i^2 + b\Sigma x_i = \Sigma y_i$

(d) $a\Sigma x_i^2 + nb = \Sigma x_i y_i$ and $a\Sigma x_i + b\Sigma x_i = \Sigma y_i$

7) In a line of best fit $y = 5.8\,(x-1994) + 41.6$ the value of y when $x = 1997$ is

(a) 50 (b) 54 (c) 59 (d) 60

8) Five data relating to x and y are to be fit in a straight line. It is found that $\Sigma x = 0$ and $\Sigma y = 15$. Then the y-intercept of the line of best fit is,

(a) 1 (b) 2 (c) 3 (d) 4

9) The normal equations of fitting a straight line $y = ax + b$ are $10a + 5b = 15$ and $30a + 10b = 43$. The slope of the line of best fit is

(a) 1.2 (b) 1.3 (c) 13 (d) 12

10) The normal equations obtained in fitting a straight line $y = ax + b$ by the method of least squares over n points (x, y) are $4 = 4a + b$ and $\Sigma xy = 120a + 24b$. Then $n =$

(a) 30 (b) 5 (c) 6 (d) 4

PROBABILITY DISTRIBUTIONS 8

8.1 RANDOM VARIABLE AND PROBABILITY FUNCTION

Random variable

A **random variable** is a real valued function defined on a sample space S and taking values in $(-\infty, \infty)$

8.1.1 Discrete Random Variable

A random variable X is said to be discrete if it assumes only a finite or an infinite but countable number of values.

Examples

(i) Consider the experiment of tossing a coin twice. The sample points of this experiment are s_1 = (H, H), s_2 = (H, T), s_3 = (T, H) and s_4 = (T, T).

Random variable X denotes the number of heads obtained in the two tosses.

Then $X(s_1) = 2$ $\quad X(s_2) = 1$

$X(s_3) = 1$ $\quad X(s_4) = 0$

$R_X = \{0, 1, 2\}$

where s is the typical element of the sample space, $X(s)$ represents the real number which the random variable X associates with the outcome s.

R_X, the set of all possible values of X, is called the **range** space X.

(ii) Consider the experiment of rolling a pair of fair dice once. Then sample space

$$S = \{(1, 1)\ (1, 2)\ \ldots\ldots\ldots\ldots(1, 6)$$
$$\vdots \qquad \vdots \qquad \qquad \vdots$$
$$(6, 1)\ (6, 2)\ \ldots\ldots\ldots\ldots(6, 6)\}$$

Let the random variable X denote the sum of the scores on the two dice. Then $R_X = \{2, 3, 4, \ldots\ldots, 12\}$.

(iii) Consider the experiment of tossing of 3 coins simultaneously.

Let the random variable X be Number of heads obtained in this experiment.

Then

$$S = \{HHH, HHT, HTT, TTT, TTH, THH, HTH, THT\}$$

$$R_X = \{0, 1, 2, 3\}$$

(iv) Suppose a random experiment consists of throwing 4 coins and recording the number of heads.

Then $R_X = \{0, 1, 2, 3, 4\}$

The number of printing mistakes in each page of a book and the number of telephone calls received by the telephone operator of a firm, are some other examples of discrete random variable.

8.1.2 Probability function and Probability distribution of a Discrete random variable

Let X be a discrete random variable assuming values $x_1, x_2, x_3 \ldots$ If there exists a function p denoted by $p(x_i) = P[X = x_i]$ such that

(i) $p(x_i) \geq 0$ for $i = 1, 2, \ldots$

(ii) $\sum_i p(x_i) = 1$

then p is called as the **probability function** or **probability mass function (p.m.f)** of X.

The collection of all pairs (x_i, $p(x_i)$) is called the probability distribution of X.

Example 1

Consider the experiment of tossing two coins. Let X be a random variable denoting the number of heads obtained.

X	**:**	**0**	**1**	**2**
p(x_i)	**:**	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Is p(x_i) a p.m.f ?

Solution :

(i) $p(x_i) > 0$ for all i

(ii) $\Sigma p(x_i) = p(0) + p(1) + p(2)$

$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$$

Hence $p(x_i)$ is a p.m.f.

Example 2

Consider the discrete random variable X as the sum of the numbers that appear, when a pair of dice is thrown. The probability distribution of X is

X	**:**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
***p*(*x_i*)**	**:**	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Is p(x_i) a p.m.f?

Solution :

$p(x_i) > 0$ for all i

(ii) $\Sigma p(x_i) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \text{........} + \frac{1}{36} = 1$

Hence $p(x_i)$ is a p.m.f.

8.1.3 Cumulative Distribution function : (c.d.f.)

Let X be a discrete random variable. The function F(x) is said to be the cumulative distribution function (c.d.f.) of the random variable X if

$$F(x) = P(X \leq x)$$

$$= \sum_i p(x_i) \text{ where the sum is taken over } i \text{ such that } x_i \leq x.$$

Remark : $P(a < X \leq b) = F(b) - F(a)$

Example 3

A random variable X has the following probability function :

Values of X, x :	**–2**	**–1**	**0**	**1**	**2**	**3**
$p(x)$:	**0.1**	**k**	**0.2**	**2k**	**0.3**	**k**

(i) Find the value of k

(ii) Construct the c.d.f. of X

Solution :

(i) Since $\sum_i p(x_i) = 1$,

$p(-2) + p(-1) + p(0) + p(1) + p(2) + p(3) = 1$

$0.1 + k + 0.2 + 2k + 0.3 + k = 1$

$0.6 + 4k = 1 \Rightarrow 4k = 1 - 0.6$

$4k = 0.4 \qquad \therefore k = \frac{.4}{4} = 0.1$

Hence the given probability function becomes,

x	:	-2	-1	0	1	2	3
$p(x)$	:	0.1	0.1	0.2	0.2	0.3	0.1

(ii) Cumulative distribution function $F(x) = P(X \leq x)$

x	$F(x) = P(X \leq x)$
-2	$F(-2) = P(X \leq -2) = 0.1$
-1	$F(-1) = P(X \leq -1) = P(X = -2) + P(X = -1)$ $= 0.1 + 0.1 = 0.2$
0	$F(0) = P(X \leq 0) = P(X=-2) + P(X=-1)+ P(X = 0)$ $= 0.1 + 0.1 + 0.2 = 0.4$
1	$F(1) = P(X \leq 1) = 0.6$
2	$F(2) = P(X \leq 2) = 0.9$
3	$F(3) = P(X \leq 3) = 1$

$$
\begin{aligned}
F(x) &= 0 \text{ if } x < -2 \\
&= .1 \text{ if } -2 \leq x < -1 \\
&= .2 \text{ if } -1 \leq x < 0 \\
&= .4 \text{ if } 0 \leq x < 1 \\
&= .6 \text{ if } 1 \leq x < 2 \\
&= 0.9, \text{ if } 2 \leq x < 3 \\
&= 1 \text{ if } x \geq 3
\end{aligned}
$$

Example 4

For the following probability distribution of X

X	**:**	**0**	**1**	**2**	**3**
$\boldsymbol{p(x)}$	**:**	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

Find (i) $P(X \leq 1)$ (ii) $P(X \leq 2)$ (iii) $P(0< X < 2)$

Solution :

(i) $P(X \leq 1) = P(X = 0) + P(X = 1)$

$$= p(0) + p(1)$$

$$= \frac{1}{6} + \frac{1}{2} = \frac{4}{6} = \frac{2}{3}$$

(ii) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= \frac{1}{6} + \frac{1}{2} + \frac{3}{10} = \frac{29}{30}$$

Aliter $P(X \leq 2)$ can also be obtained as

$$P(X \leq 2) = 1 - P(X > 2)$$

$$= 1 - P(X = 3) = 1 - \frac{1}{30} = \frac{29}{30}$$

(iii) $P(0 < X < 2) = P(X = 1) = \frac{1}{2}$

8.1.4 Continuous Random Variable

A random variable X is said to be continuous if it takes a continuum of values. i.e. if it takes all possible values between certain defined limits.

For example,

(i) The amount of rainfall on a rainy day.

(ii) The height of individuals. (iii) The weight of individuals.

8.1.5 Probability function

A function f is said to be the probability density function (p.d.f) of a continuous random variable X if the following conditions are satisfied

(i) $f(x) \geq 0$ for all x (ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

Remark :

(i) The probability that the random variable X lies in the interval (a, b) is given by $P(a < X < b) = \int_a^b f(x)\, dx$.

(ii) $P(X = a) = \int_a^a f(x)\, dx = 0$

(iii) $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$

8.1.6 Continuous Distribution function

If X is a continuous random variable with p.d.f. $f(x)$, then the function $F_X(x) = P(X \leq x)$

$$= \int_{-\infty}^{x} f(t)\, dt$$

is called the distribution function (d.f.) or cumulative distribution function (c.d.f) of the random variable X.

Properties : The cumulative distribution function has the following properties.

(i) $\underset{x \to -\infty}{\mathrm{Lt}} F(x) = 0$ i.e. $F(-\infty) = 0$

(ii) $\underset{x \to \infty}{\mathrm{Lt}} F(x) = 1$ i.e. $F(\infty) = 1$

(iii) Let F be the c.d.f. of a continuous random variable X with p.d.f f. Then $f(x) = \frac{d}{dx} F(x)$ for all x at which F is differentiable.

Example 5

A continuous random variable X has the following p.d.f.

$$f(x) = \begin{cases} k(2-x) & \text{for } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Determine the value of k.

Solution :

If $f(x)$ be the p.d.f., then $\int_{-\infty}^{\infty} f(x)\, dx = 1$

$$\int_{-\infty}^{0} f(x)\, dx + \int_{0}^{2} f(x)\, dx + \int_{2}^{\infty} f(x)\, dx = 1$$

$$\Rightarrow \quad 0 + \int_{0}^{2} f(x)\, dx + 0 = 1$$

$$\Rightarrow \quad \int_{0}^{2} k(2-x)\, dx = 1$$

$$k\left(\int_0^2 2dx - xdx\right) = 1 \qquad \therefore \qquad k = \frac{1}{2}$$

Hence $f(x) = \begin{cases} \frac{1}{2}(2-x) & \text{for } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$

Example 6

Verify that

$$\boldsymbol{f(x) = \begin{cases} 3x^2 & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}}$$

is a p.d.f and evaluate the following probabilities

(i) P(X $\le \frac{1}{3}$) **(ii) P($\frac{1}{3} \le$ X $\le \frac{1}{2}$)**

Solution :

Clearly $f(x) \geq 0$ for all x and hence one of the conditions for p.d.f is satisfied.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 f(x)\,dx = \int_0^1 3x^2\,dx = 1$$

$\therefore$ The other condition for p.d.f is also satisfied.

Hence the given function is a p.d.f

(i) P(X $\le \frac{1}{3}$) $= \int_{-\infty}^{\frac{1}{3}} f(x)\,dx$ $\qquad$ $P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$

$$= \int_0^{\frac{1}{3}} 3x^2\,dx = \frac{1}{27}$$

(ii) P($\frac{1}{3} \le$ X $\le \frac{1}{2}$) $= \int_{\frac{1}{3}}^{\frac{1}{2}} f(x)\,dx$

$$= \int_{\frac{1}{3}}^{\frac{1}{2}} 3x^2\,dx = \frac{1}{8} - \frac{1}{27} = \frac{19}{216}$$

Example 7

Given the p.d.f of a continuous random variable X as follows

$$\boldsymbol{f(x)} = \begin{cases} \boldsymbol{kx(1-x)} & \textbf{for } \boldsymbol{0 < x < 1} \\ \boldsymbol{0} & \textbf{otherwise} \end{cases}$$

Find *k* and c.d.f

Solution :

If X is a continuous random variable with p.d.f $f(x)$ then

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\int_0^1 k\,x\,(1-x)\,dx = 1$$

$$k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 1 \qquad \therefore\ k = 6$$

Hence the given p.d.f becomes,

$$f(x) = \begin{cases} 6x(1-x) & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

To find c.d.f $F(x)$

$F(x) = 0 \qquad \text{for } x \le 0$

$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$

$= \int_0^x 6x(1-x)dx = 3x^2 - 2x^3 \qquad \text{for } 0 < x < 1$

$F(x) = 1 \qquad \text{for } x \ge 1$

$\therefore$ The c.d.f of X is as follows.

$F(x) = 0 \qquad \text{for } x \le 0$

$= 3x^2 - 2x^3 \qquad \text{for } 0 < x < 1.$

$= 1 \qquad \text{for } x \ge 1$

Example 8

Suppose that the life in hours of a certain part of radio tube is a continuous random variable X with p.d.f is given by

$$\mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{100}}{x^2}, & \textbf{when } x \geq 100 \\ \mathbf{0} & \textbf{elsewhere} \end{cases}$$

(i) **What is the probability that all of three such tubes in a given radio set will have to be replaced during the first of 150 hours of operation?**

(ii) **What is the probability that none of three of the original tubes will have to be replaced during that first 150 hours of operation?**

Solution :

(i) A tube in a radio set will have to be replaced during the first 150 hours if its life is < 150 hours. Hence, the required probability '*p*' that a tube is replaced during the first 150 hours is,

$$p = P(X \leq 150) = \int_{100}^{150} f(x)\,dx$$

$$= \int_{100}^{150} \frac{100}{x^2}\,dx = \frac{1}{3}$$

∴ The probability that all three of the original tubes will have to replaced during the first 150 hours $= p^3 = \left(\frac{1}{3}\right)^3 = \frac{1}{27}$

(ii) The probability that a tube is not replaced during the first 150 hours of operation is given by

$$P(X > 150) = 1 - P(X \leq 150) = 1 - \frac{1}{3} = \frac{2}{3}$$

∴ the probability that none of the three tubes will be replaced during the 150 hours of operation $= \left(\frac{2}{3}\right)^3 = \frac{8}{27}$

EXERCISE 8.1

1) Which of the following set of functions define a probability space on S = $[x_1, x_2, x_3]$?

(i) $p(x_1) = \frac{1}{3}$ $p(x_2) = \frac{1}{2}$ $p(x_3) = \frac{1}{4}$

(ii) $p(x_1) = \frac{1}{3}$ $p(x_2) = \frac{1}{6}$ $p(x_3) = \frac{1}{2}$

(iii) $p(x_1) = 0$ $p(x_2) = \frac{1}{3}$ $p(x_3) = \frac{2}{3}$

(iv) $p(x_1) = p(x_2) = \frac{2}{3}$ $p(x_3) = \frac{1}{3}$

2) Consider the experiment of throwing a single die. The random variable X represents the score on the upper face and assumes the values as follows:

X	:	1	2	3	4	5	6
$p(x_i)$	:	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Is $p(x_i)$ a p.m.f?

3) A random variable X has the following probability distribution.

Values of X, x :	0	1	2	3	4	5	6	7	8
$p(x)$:	a	$3a$	5a	$7a$	$9a$	$11a$	$13a$	$15a$	$17a$

(i) Determine the value of a

(ii) Find P(X < 3), P(X > 3) and P(0< X < 5)

4) The following function is a probability mass function - Verify.

$$p(x) = \begin{cases} \frac{1}{3} & \text{for } x=1 \\ \frac{2}{3} & \text{for } x=2 \\ 0 & \text{otherwise} \end{cases}$$

Hence find the c.d.f

5) Find k if the following function is a probability mass function.

$$p(x) = \begin{cases} \frac{k}{6} & \text{for } x = 0 \\ \frac{k}{3} & \text{for } x = 2 \\ \frac{k}{2} & \text{for } x = 4 \\ 0 & \text{otherwise} \end{cases}$$

6) A random variable X has the following probability distribution

values of X, x :	-2	0	5
$p(x)$:	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Evaluate the following probabilities

(a) $P(X \leq 0)$ (b) $P(X < 0)$ (c) $P(0 \leq X \leq 10)$

7) A random variable X has the following probability function

Values of X, x :	0	1	2	3
$p(x)$:	$\frac{1}{16}$	$\frac{3}{8}$	k	$\frac{5}{16}$

(i) Find the value of k (ii) Construct the c.d.f. of X

8) A continuous random variable has the following p.d.f

$f(x) = kx^2, \quad 0 \leq x \leq 10$

$= 0$ otherwise.

Determine k and evaluate (i) $P(.2 \leq X \leq 0.5)$ (ii) $P(X \leq 3)$

9) If the function $f(x)$ is defined by

$f(x) = ce^{-x}, \quad 0 \leq x < \infty$. Find the value of c.

10) Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} ax, & 0 < x \leq 1 \\ a, & 1 \leq x \leq 2 \\ -ax + 3a, & 2 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

(i) Determine the constant a

(ii) Compute $P(X \leq 1.5)$

11) Let X be the life length of a certain type of light bulbs in hours. Determine 'a' so that the function

$f(x) = \frac{a}{x^2}, \quad 1000 \leq x \leq 2000$

$= 0$ otherwise.

may be the probability density function.

12) The kms. X in thousands which car owners get with a certain kind of tyre is a random variable having p.d.f.

$f(x) = \frac{1}{20} e^{-\frac{x}{20}}$, for $x > 0$

$= 0$ for $x \leq 0$

Find the probabilities that one of these tyres will last

(i) atmost 10,000 kms

(ii) anywhere from 16,000 to 24,000 kms

(iii) atleast 30,000 kms.

8.2 MATHEMATICAL EXPECTATION

The concept of Mathematical expectation plays a vital role in statistics. Expected value of a random variable is a weighted average of all the possible outcomes of an experiment.

If X is a discrete random value which can assume values $x_1, x_2, \ldots x_n$ with respective probabilities $p(x_i) = P[X = x_i]$; $i = 1, 2 \ldots n$ then its **mathematical expectation** is defined as

$$E(X) = \sum_{i=1}^{n} x_i \, p(x_i), \qquad \left(\text{Here } \sum_{i=1}^{n} p(x_i) = 1\right)$$

If X is a continuous random variable with probability density function $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

Note

E(X) is also known as the mean of the random variable X.

Properties

1) $E(c) = c$ where c is constant

2) $E(X + Y) = E(X) + E(Y)$

3) $E(aX + b) = aE(X) + b$ where a and b are constants.

4) $E(XY) = E(X)\,E(Y)$ if X and Y are independent

Note

The above properties holds good for both discrete and continous random variables.

Variance

Let X be a random variable. Then the **Variance of X**, denoted by Var(X) or σ^2_x is

$$Var(X) = \sigma^2_x = E[X - E(X)]^2$$
$$= E(X^2) - [E(X)]^2$$

The positive square root of Var(X) is called the **Standard Deviation** of X and is denoted by σ_x.

Example 9

A multinational bank is concerned about the waiting time (in minutes) of its customer before they would use ATM for their transaction. A study of a random sample of 500 customers reveals the following probability distribution.

X	**:**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
***p*(*x*)**	**:**	**.20**	**.18**	**.16**	**.12**	**.10**	**.09**	**.08**	**.04**	**.03**

Calculate the expected value of waiting time, X, of the customer

Solution :

Let X denote the waiting time (in minutes) per customer.

X	:	0	1	2	3	4	5	6	7	8
p(*x*)	:	.20	.18	.16	.12	.10	.09	.08	.04	.03

Then $E(X) = \Sigma x \, p(x)$

$$= (0 \times .2) + (1 \times 0.18) + ... + (8 \times 0.03) = 2.71$$

The expected value of X is equal to 2.71 minutes. Thus the average waiting time of a customer before getting access to ATM is 2.71 minutes.

Example 10

Find the expected value of the number of heads appearing when two fair coins are tossed.

Solvtion :

Let X be the random variable denoting the number of heads.

Possible values of X	:	0	1	2
Probabilities $p(x_i)$	:	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

The Expected value of X is

$$E(X) = x_1\, p(x_1) + x_2\, p(x_2) + x_3\, p(x_3)$$

$$= 0\left(\frac{1}{4}\right) + 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{4}\right) = 1$$

Therefore, the expected number of heads appearing in the experiment of tossing 2 fair coins is 1.

Example 11

The probability that a man fishing at a particular place will catch 1, 2, 3, 4 fish are 0.4, 0.3, 0.2 and 0.1 respectively. What is the expected number of fish caught?

Solution :

Possible values of X	:	1	2	3	4
Probabilities $p(x_i)$	:	0.4	0.3	0.2	0.1

$\therefore$ $E(X) = \sum_i x_i\, p(x_i)$

$$= x_1\, p(x_1) + x_2\, p(x_2) + x_2\, p(x_3) + x_4\, p(x_4)$$

$$= 1\,(.4) + 2(.3) + 3(.2) + 4(.1)$$

$$= .4 + .6 + .6 + .4 = 2$$

Example 12

A person receives a sum of rupees equal to the square of the number that appears on the face when a balance die is tossed. How much money can he expect to receive?

Solution :

Random variable X: as square of the number that can appear on the face of a die. Thus

possible values of	X	:	1^2	2^2	3^2	4^2	5^2	6^2
probabilities	$p(x_i)$	:	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The Expected amount that he receives,

$$E(X) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + \ldots + 6^2\left(\frac{1}{6}\right)$$

$$= \text{Rs.}\frac{91}{6}$$

Example 13

A player tosses two fair coins. He wins Rs.5 if two heads appear, Rs. 2 if 1 head appears and Rs.1 if no head occurs. Find his expected amount of gain.

Solution :

Consider the experiment of tossing two fair coins. There are four sample points in the sample space of this experiment.

i.e. S = {HH, HT, TH, TT}

Let X be the random variable denoting the amount that a player wins associated with the sample point.

Thus,

Possible values of X (Rs.)		:	5	2	1
Probabilities	$p(x_i)$	:	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$E(X) = 5\left(\frac{1}{4}\right) + 2\left(\frac{1}{2}\right) + 1\left(\frac{1}{4}\right)$$

$$= \frac{5}{4} + 1 + \frac{1}{4} = \frac{10}{4} = \frac{5}{2}$$

$$= \text{Rs. } 2.50$$

Hence expected amount of winning is Rs.2.50

Example 14

A random variable X has the probability function as follows :

values of X	**:**	**-1**	**0**	**1**
probability	**:**	**0.2**	**0.3**	**0.5**

Evaluate (i) E(3X +1) (ii) $E(X^2)$ (iii) Var(X)

Solution :

X	:	–1	0	1
$p(x_i)$	:	0.2	0.3	0.5

(i) E(3X+1)= 3E(X) + 1

Now E(X) = –1 x 0.2 + 0 x 0.3 + 1 x 0.5

= –1 x 0.2 + 0 + 0.5 = 0.3

E(3X + 1) = 3(0.3) + 1 = 1.9

(ii) $E(X^2) = \Sigma x^2 \, p(x)$

$= (-1)^2 \times 0.2 + (0)^2 \times 0.3 + (1)^2 \times 0.5$

= 0.2 + 0 + 0.5 = 0.7

(iii) Var(X) $= E(X^2) - [E(X)]^2$

$= .7 - (.3)^2 = .61$

Example 15

Find the mean, variance and the standard deviation for the following probability distribution

Values of X,	***x***	**:**	**1**	**2**	**3**	**4**
probability,	***p*(*x*)**	**:**	**0.1**	**0.3**	**0.4**	**0.2**

Solution :

Mean = E(X) = $\Sigma x \, p(x)$

= 1(0.1) + 2(0.3) + 3(0.4) + 4(0.2) = 2.7

Variance = $E(X^2) - [E(X)]^2$

Now $E(X^2) = \Sigma x^2 \, p(x)$

$= 1^2(0.1) + 2^2(0.3) + 3^2(0.4) + 4^2(0.2) = 8.1$

$\therefore$ Variance $= 8.1 - (2.7)^2$

$= 8.1 - 7.29 = .81$

Standard Deviation $= \sqrt{0.81} = 0.9$

Example 16

Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} \frac{1}{2} & \text{for } -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find (i) E(X) (ii) $E(X^2)$ (iii) Var(X)

Solution :

(i) $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$ (by definition)

$$= \frac{1}{2} \int_{-1}^{1} x\, dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_{-1}^{1} = 0$$

(ii) $E(X^2) = \int_{-1}^{1} x^2 f(x)\, dx$

$$= \int_{-1}^{1} x^2 \frac{1}{2} dx$$

$$= \frac{1}{2}\left[\frac{x^3}{3}\right]_{-1}^{1} = \frac{1}{3}$$

(iii) $Var(X) = E(X^2) - [E(X)]^2$

$$= \frac{1}{3} - 0 = \frac{1}{3}$$

EXERCISE 8.2

1) A balanced die is rolled. A person recieves Rs. 10 when the number 1 or 3 or 5 occurs and loses Rs. 5 when 2 or 4 or 6 occurs. How much money can he expect on the average per roll in the long run?

2) Two unbiased dice are thrown. Find the expected value of the sum of the points thrown.

3) A player tossed two coins. If two heads show he wins Rs. 4. If one head shows he wins Rs. 2, but if two tails show he must pay Rs. 3 as penalty. Calculate the expected value of the sum won by him.

4) The following represents the probability distribution of D, the daily demand of a certain product. Evaluate E(D).

D	1	2	3	4	5
$P[D=d]$	0.1	0.1	0.3	0.3	0.2

5) Find E(2X-7) and E(4X + 5) for the following probability distribution.

X	–3	–2	–1	0	1	2	3
$p(x)$	. 05	.1	.3	0	.3	.15	.1

6) Find the mean, variance and standard deviation of the following probability distribution.

Values of X :	–3	–2	–1	0	1	2	3
Probability $p(x)$:	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$

7) Find the mean and variance for the following probability distribution.

$$f(x) = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0 \quad , & x < 0 \end{cases}$$

8.3 DISCRETE DISTRIBUTIONS

We know that the frequency distributions are based on observed data derived from the collected sample information. For example, we may study the marks of the students of a class and formulate a frequency distribution as follows:

Marks	No. of students
0 - 20	10
20-40	12
40-60	25
60-80	15
80-100	18
Total	**80**

The above example clearly shows that the observed frequency distributions are obtained by grouping. Measures like averages, dispersion, correlation, etc. generally provide us a consolidated view of the whole observed data. This may very well be used in formulating certain ideas (inference) about the characteristics of the whole set of data.

Another type of distribution in which variables are distributed according to some definite probability law which can be expressed mathematically are called theoretical probability distribution.

The probability distribution is a total listing of the various values the random variable can take along with the corresponding probabilities of each value. For example; consider the pattern of distribution of machine breakdown in a manufacturing unit. The random variable would be the various values the machine breakdown could assume. The probability corresponding to each value of the breakdown as the relative frequency of occurence of the breakdown. This probability distribution is constructed by the actual breakdown pattern discussed over a period of time.

Theoretical probability distributions are basically of two types

(i) Discrete and (ii) Continuous

In this section, we will discuss theoretical discrete distributions namely, Binomial and Poisson distributions.

8.3.1 Binomial Distribution

It is a distribution associated with repetition of independent trials of an experiment. Each trial has two possible outcomes, generally called success and failure. Such a trial is known as **Bernoulli trial.**

Some examples of Bernoulli trials are :

(i) a toss of a coin (Head or tail)

(ii) the throw of a die (even or odd number)

An experiment consisiting of a repeated number of Bernoulli trials is called a **binomial experiment.** A binomial experiment must possess the following properties:

(i) there must be a fixed number of trials.

(ii) all trials must have identical probabilities of success (p) i.e. if we call one of the two outcomes as "success" and the other as "failure", then the probability p of success remains as constant throughout the experiment.

(iii) the trials must be independent of each other i.e. the result of any trial must not be affected by any of the preceeding trial.

Let X denote the number of successes in 'n' trials of a binomial experiment. Then X follows a binomial distribution with parameters n and p and is denoted by $X \sim B(n, p)$.

A random variable X is said to follow **Binomial distribution** with parameters n and p, if it assumes only non-negative values and its probability mass function is given by

$$P[X=x] = p(x) = {}^nC_x\, p^x\, q^{n-x}\ ;\ x = 0, 1, 2, ..n\ ;\ q = 1 - p$$

Remark

(i) $\sum\limits_{x=0}^{n} p(x) = \sum\limits_{x=0}^{n} {}^nC_x\, p^x\, q^{n-x} = (q + p)^n = 1$

(ii) ${}^nC_r = \dfrac{n(n-1)\ldots(n-\overline{r-1})}{1.2.3\ldots r}$

Mean and Variance

For the binomial distribution

Mean = np

Variance = npq; Standard Deviation = $\sqrt{npq}$

Example 17

What is the probability of getting exactly 3 heads in 8 tosses of a fair coin.

Solution :

Let p denote the probability of getting head in a toss.

Let X be the number of heads in 8 tosses.

Then $p = \frac{1}{2}$, $q = \frac{1}{2}$ and $n = 8$

Probability of getting exactly 3 heads is

$$P(X = 3) = {}^{8}c_{3} \left(\frac{1}{2}\right)^{3} \left(\frac{1}{2}\right)^{5}$$

$$= \frac{8 \times 7 \times 6}{1 \times 2 \times 3} \left(\frac{1}{2}\right)^{8} = \frac{7}{32}$$

Example 18

Write down the Binomial distribution whose mean is 20 and variance being 4.

Solution :

Given mean, $np = 20$; variance, $npq = 4$

Now $q = \frac{npq}{np} = \frac{4}{20} = \frac{1}{5}$ $\therefore$ $p = 1 - q = \frac{4}{5}$

From $np = 20$, we have

$$n = \frac{20}{p} = \frac{20}{\frac{4}{5}} = 25$$

Hence the binomial distribution is

$$p(x) = {}^{n}C_{x} p^{x} q^{n-x} = {}^{25}C_{x} \left(\frac{4}{5}\right)^{x} \left(\frac{1}{5}\right)^{n-x}, \quad x = 0, 1, 2, \ldots, 25$$

Example 19

On an average if one vessel in every ten is wrecked, find the probability that out of five vessels expected to arrive, atleast four will arrive safely.

Solution :

Let the probability that a vessel will arrive safely, $p = \frac{9}{10}$

Then probability that a vessel will be wrecked, $q = 1-p = \frac{1}{10}$

No. of vessels, $n = 5$

$\therefore$ The probability that atleast 4 out of 5 vessels to arrive safely is

$P(X \geq 4) = P(X = 4) + P(X = 5)$

$$= {}^5C_4 \left(\frac{9}{10}\right)^4 \frac{1}{10} + {}^5C_5\left(\frac{9}{10}\right)^5$$

$$= 5(.9)^4(.1) + (.9)^5 \;\; = .91854$$

Example 20

For a binomial distribution with parameters *n* = 5 and *p* = .3 find the probabilities of getting (i) atleast 3 successes (ii) atmost 3 successes.

Solution :

Given $n = 5$, $p = .3$ $\therefore$ $q = .7$

(i) The probability of atleast 3 successes

$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$

$= {}^5C_3 (0.3)^3 (0.7)^2 + {}^5C_4 (0.3)^4 (0.7) + {}^5C_5 (.3)^5(7)^0$

$= .1631$

(ii) The probability of atmost 3 successes

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$

$= (.7)^5 + {}^5C_1 (.7)^4 (.3) + {}^5C_2 (.7)^3 (.3)^2 + {}^5C_3 (.7)^2 (.3)^3$

$= .9692$

8.3.2 Poisson distribution

Poisson distribution is also a discrete probability distribution and is widely used in statistics. Poisson distribuition occurs when there are events which do not occur as outcomes of a definite number of trials of an experiment but which occur at random points of time and space wherein our interest lies only in the number of occurences of the event, not in its non-occurances. This distribution is used to describe the behaviour of rare events such as

(i) number of accidents on road

(ii) number of printing mistakes in a book

(iii) number of suicides reported in a particular city.

Poisson distribution is an approximation of binomial distribution when n (number of trials) is large and *p*, the probability of success is very close to zero with *np* as constant.

A random variable X is said to follow a **Poisson distribution** with parameter $\lambda > 0$ if it assumes only non-negative values and its probability mass function is given by

$$P[X = x] = p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}\,; \quad x = 0, 1, 2, \ldots$$

Remark

It should be noted that

$$\sum_{x=0}^{\infty} P[X = x] = \sum_{x=0}^{\infty} p(x) = 1$$

Mean and Variance

For the of poisson distribution

Mean, $E(X) = \lambda$, Variance, $Var(X) = \lambda$, S.D = $\sqrt{\lambda}$

Note

For poission distribution mean and variance are equal.

Example 21

Find the probability that atmost 5 defective fuses will be found in a box of 200 fuses if experience shows that 2 percent of such fuses are defective. ($e^{-4} = 0.0183$)

Solution :

p = probability that a fuse is defective = $\frac{2}{100}$

$n = 200$

$\therefore\ \lambda = np = \frac{2}{100} \times 200 = 4$

Let X denote the number of defective fuses found in a box.

Then the distribution is given by

$P[X = x] = p(x) = \frac{e^{-4} 4^x}{x!}$

So, probability that atmost 5 defective fuses will be found in a box of 200 fuses

$= P(X \leq 5)$

$= P(X = 0) + P(X = 1) + P(X = 2)$

$\quad + P(X = 3) + P(X = 4) + P(X = 5)$

$= e^{-4} + \frac{e^{-4} 4}{1!} + \frac{e^{-4} 4^2}{2!} + \frac{e^{-4} 4^3}{3!} + \frac{e^{-4} 4^4}{4!} + \frac{e^{-4} 4^5}{5!}$

$= e^{-4} (1 + \frac{4}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!} + \frac{4^4}{4!} + \frac{4^5}{5!})$

$= 0.0183 \times \frac{643}{15} \quad = 0.785$

Example 22

Suppose on an average 1 house in 1000 in a certain district has a fire during a year. If there are 2000 houses in that district, what is the probability that exactly 5 houses will have fire during the year? (e^{-2} = .13534)

Solution :

p = probability that a house catches fire = $\frac{1}{1000}$

Here $n = 2000 \quad \therefore\ \lambda = np = 2000 \times \frac{1}{1000} = 2$

Let X denote the number of houses that has a fire

Then the distribution is given by $P[X = x] = \frac{e^{-2}2^x}{x!}$, $x = 0, 1, 2,...$

Probability that exactly 5 houses will have a fire during the year is

$$P[X = 5] = \frac{e^{-2}2^5}{5!}$$

$$= \frac{.13534 \times 32}{120} = .0361$$

Example 23

The number of accidents in a year attributed to taxi drivers in a city follows poisson distribution with mean 3. Out of 1000 taxi drivers, find the approximate number of drivers with

(i) no accident in a year

(ii) more than 3 accidents in a year

Solution :

Here $\lambda = np = 3$

$N = 1000$

Then the distribution is

$P[X = x] = \frac{e^{-3}3^x}{x!}$ where X denotes the number accidents.

(i) P (no accidents in a year) = P(X = 0)

$$= e^{-3} = 0.05$$

$\therefore$ Number of drivers with no accident = 1000 x 0.05 = 50

(ii) P (that more than 3 accident in a year) = $P(X > 3)$

$$= 1 - P(X \leq 3)$$

$$= 1 - \left[e^{-3} + \frac{e^{-3}3^1}{1!} + \frac{e^{-3}3^2}{2!} + \frac{e^{-3}3^3}{3!}\right]$$

$= 1 - e^{-3}[1 + 3 + 4.5 + 4.5]$

$= 1 - e^{-3}(13) = 1 - .65 = .35$

$\therefore$ Number of drivers with more than 3 accidents

$= 1000 \times 0.35 = 350$

EXERCISE 8.3

1) Ten coins are thrown simultaneously. Find the probability of getting atleast 7 heads.

2) In a binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the parameter '*p*' of the distribution.

3) For a binomial distribution, the mean is 6 and the standard deviation is $\sqrt{2}$. Write down all the terms of the distribution.

4) The average percentage of failure in a certain examination is 40. What is the probability that out of a group of 6 candidates atleast 4 passed in the examination?

5) An unbiased coin is tossed six times. What is the probability of obtaining four or more heads?

6) It is stated that 2% of razor blades supplied by a manufacturer are defective. A random sample of 200 blades is drawn from a lot. Find the probability that 3 or more blades are defective. ($e^{-4} = .01832$)

7) Find the probability that atmost 5 defective bolts will be found in a box of 200 bolts, if it is known that 2% of such bolts are expected to be defective ($e^{-4} = 0.01832$)

8) An insurance company insures 4,000 people against loss of both eyes in car acidents. Based on previous data, the rates were computed on the assumption that on the average 10 persons in 1,00,000 will have car accidents each year that result in this type of injury. What is the probability that more than 3 of the injured will collect on their policy in a given year? ($e^{-0.4} = 0.6703$)

9) It is given that 3% of the electric bulbs manufactured by a company are defective. Find the probability that a sample of 100 bulbs will contain (i) no defective (ii) exactly one defective. (e^{-3} = 0.0498).

10) Suppose the probability that an item produced by particular machine is defective equals 0.2. If 10 items produced from this machine are selected at random, what is the probability that not more than one defective is found? (e^{-2} = .13534)

8.4 CONTINUOUS DISTRIBUTIONS

The binomial and Poisson distributions discussed in the previous section are the most useful theoretical distributions. In order to have mathematical distribution suitable for dealing with quantities whose magnitudes vary continuously like heights and weights of individuals, a continuous distribution is needed. Normal distribution is one of the most widely used continuous distributions.

8.4.1 Normal Distribution

Normal Distribution is considered to be the most important and powerful of all the distributions in statistics. It was first introduced by De Moivre in 1733 in the development of probability. Laplace (1749 - 1827) and Gauss (1827 - 1855) were also associated with the development of Normal distribution.

A random variable X is said to follow a **Normal Distribution** with mean μ and variance σ^2 denoted by X ~ N(μ, σ^2), if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0$$

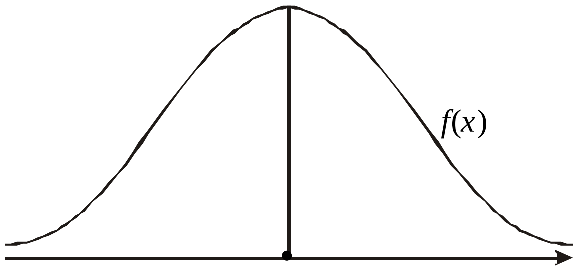
Remark

The parameters μ and σ^2 completely describe the normal distribution. Normal distribution could be also considered as limiting form of binomial distribution under the following conditions:

(i) n, the number of trials is indefinitely large i.e. $n \to \infty$

(ii) neither p nor q is very small.

The graph of the p.d.f of the normal distribution is called the **Normal curve**, and it is given below.



Normal probability curve

8.4.2 Properties of Normal Distribution

The following are some of the important properties of the normal curve and the normal distribution.

(i) The curve is "bell - shaped" and symmetric about $x = \mu$

(ii) Mean, Median and Mode of the distribution coincide.

(iii) There is one maximum point of the normal curve which occurs at the mean (μ). The height of the curve declines as we go in either direction from the mean.

(iv) The two tails of the curve extend infinitely and never touch the horizontal (x) axis.

(v) Since there is only one maximum point, the normal curve is unimodal i.e. it has only one mode.

(vi) Since $f(x)$ being the probability, it can never be negative and hence no portion of the curve lies below the x - axis.

(vii) The points of inflection are given by $x = \mu \pm \sigma$

(viii) Mean Deviation about mean

$$\sqrt{\frac{2}{\pi}}\sigma = \frac{4}{5}\sigma$$

(ix) Its mathematical equation is completely determined if the mean and S.D are known i.e. for a given mean μ and S.D σ, there is only one Normal distribution.

(x) Area Property : For a normal distribution with mean μ and S.D σ, the total area under normal curve is 1, and

(a) $P(\mu - \sigma < X < \mu + \sigma) = 0.6826$

i.e. (mean) $\pm 1\sigma$ covers 68.27%;

(b) $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544$

i.e. (mean) $\pm 2\sigma$ covers 95.45% area

(c) $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$

i.e. (mean) $\pm 3\sigma$ covers 99.73% area

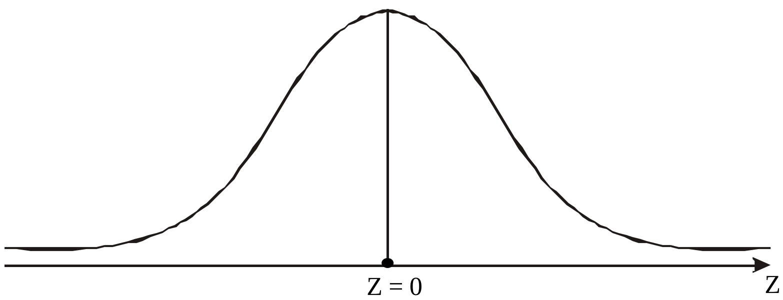
10.4.3 Standard Normal Distribution

A random variable which has a normal distribution with a mean $\mu = 0$ and a standard deviation $\sigma = 1$ is referred to as **Standard Normal Distribution.**

Remark

(i) If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma}$ is a standard normal variate with $E(Z) = 0$ and $var(Z) = 1$ i.e. $Z \sim N(0, 1)$.

(ii) It is to be noted that the standard normal distribution has the same shape as the normal distribution but with the special properties of $\mu = 0$ and $\sigma = 1$.



A random variable Z is said to have a **standard normal distribution** if its probability density function is given by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad , \qquad -\infty < z < \infty$$

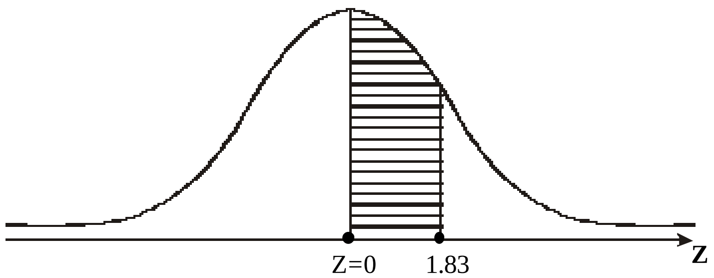
Example 24

What is the probability that Z

(a) lies between 0 and 1.83
(b) is greater than 1.54
(c) is greater than –0.86
(d) lies between 0.43 and 1.12
(e) is less than 0.77

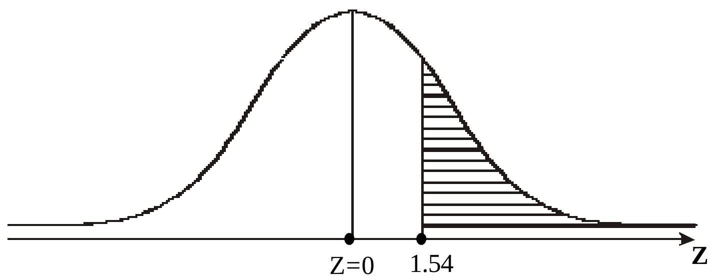
Solution :

(a) **Z lies between 0 and 1.83.**



$P(0 \leq Z \leq 1.83) = 0.4664$ (obtained from the tables directly)

(b) **Z is greater than 1.54 i.e. P(Z $\geq$ 1.54)**

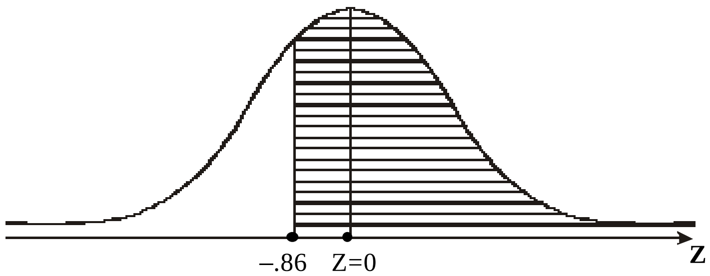


Since the total area to the right of Z = 0 is 0.5 and area between Z = 0 and 1.54 (from tables) is 0.4382

$P(Z \geq 1.54) = 0.5 - P(0 \leq Z \leq 1.54)$

$= 0.5 - .4382 = .0618$

(c) **Z is greater than -0.86 i.e. P(Z $\geq$ -0.86)**



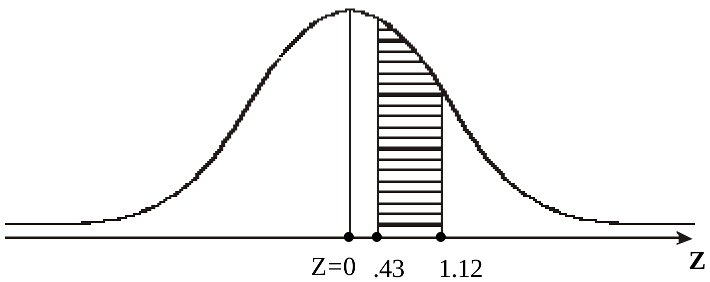
Here the area of interest $P(Z \geq -0.86)$ is represented by the two components.

(i) Area between Z = –0.86 and Z = 0, which is equal to 0.3051 (from tables)

(ii) Z > 0, which is 0.5

$\therefore P(Z \geq -0.86) = 0.3051 + 0.5 = 0.8051$

(d) **Z lies between 0.43 and 1.12**

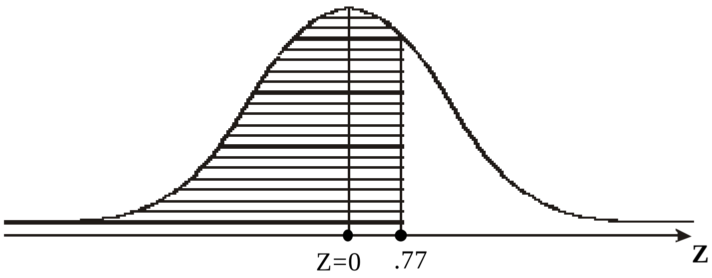


$\therefore P(0.43 \leq Z \leq 1.12) = P(0 \leq Z \leq 1.12) - P(0 \leq Z \leq 0.43)$

$= 0.3686 - 0.1664$ (from tables)

$= 0.2022.$

(e) **Z is less than 0.77**



$P(Z \leq 0.77) = 0.5 + P(0 \leq Z \leq 0.77)$

$= 0.5 + .2794 = .7794$ (from tables)

Example 25

If X is a normal random variable with mean 100 and variance 36

find (i) P(X > 112) (ii) P(X < 106) (iii) P(94 < X < 106)

Solution :

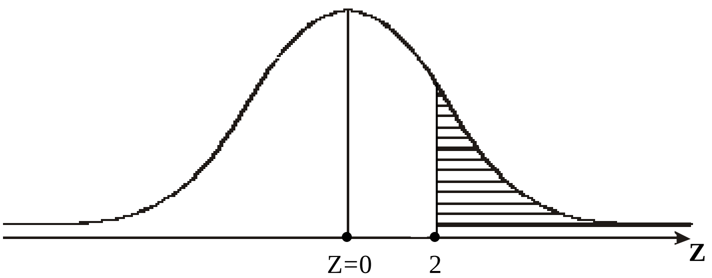
Mean, $\mu = 100$; Variance, $\sigma^2 = 36$; S.D, $\sigma = 6$

Then the standard normal variate Z is given by

$$Z = \frac{X-\mu}{\sigma} = \frac{X-100}{6}$$

(i) When X = 112, then $Z = \frac{112-100}{6} = 2$

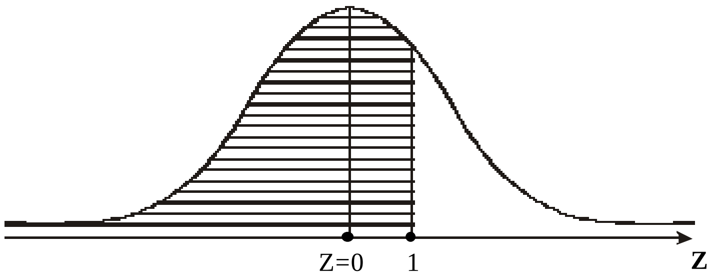
$\therefore P(X > 112) = P(Z > 2)$



$= P(0 \leq Z < \infty) - P(0 \leq Z \leq 2)$

$= 0.5 - 0.4772 = 0.0228$ (from tables)

(ii) For a given value $X = 106$, $Z = \frac{106 - 100}{6} = 1$



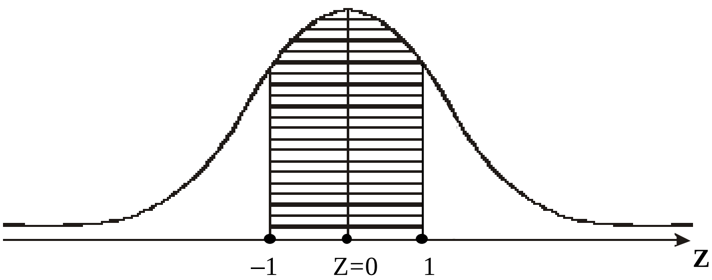
$P(X < 106) = P(Z < 1)$

$= P(-\infty \leq Z \leq 0) + P(0 \leq Z \leq 1)$

$= 0.5 + 0.3413 = 0.8413$ (from tables)

(iii) When $X = 94$, $Z = \frac{94 - 100}{6} = -1$

$X = 106$, $Z = \frac{106 - 100}{6} = +1$



$\therefore$ $P(94 < X < 106) = P(-1 < Z < 1)$

$= P(-1 < Z < 0) + P(0 < Z < 1)$

$= 2\,P(0 < Z < 1)$ (by symmetry)

$= 2\,(0.3413)$

$= 0.6826$

Example 26

In a sample of 1000 candidates the mean of certain test is 45 and S.D 15. Assuming the normality of the distrbution find the following:

(i) How many candidates score between 40 and 60?

(ii) How many candidates score above 50?

(iii) How many candidates score below 30?

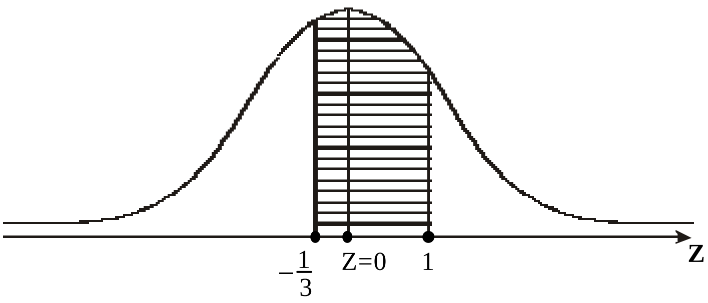
Solution :

Mean = μ = 45 and S.D. = σ = 15

Then $Z = \frac{X-\mu}{\sigma} = \frac{X-45}{15}$

(i) $P(40 < X < 60) = P(\frac{40-45}{15} < Z < \frac{60-45}{15})$

$= P(-\frac{1}{3} < Z < 1)$



$= P(-\frac{1}{3} \leq Z \leq 0) + P(0 \leq Z \leq 1)$

$= P(0 \leq Z \leq 0.33) + P(0 \leq Z \leq 1)$

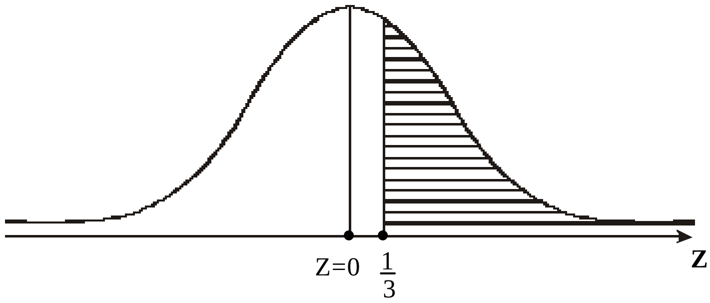
= 0.1293 + 0.3413 (from tables)

$P(40 < X < 60) = 0.4706$

Hence number of candidates scoring between 40 and 60

$= 1000 \times 0.4706 = 470.6 \simeq 471$

(ii) $P(X > 50) = P(Z > \frac{1}{3})$



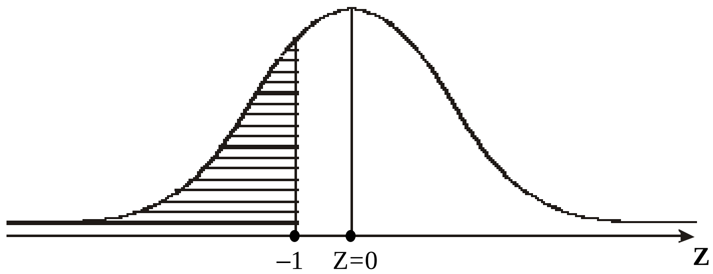
$= 0.5 - P(0 < Z < \frac{1}{3}) = 0.5 - P(0 < Z \; 0.33)$

$= 0.5 - 0.1293 = 0.3707$ (from tables)

Hence number of candidates scoring above 50

$= 1000 \times 0.3707 = 371.$

(iii) $P(X < 30) = P(Z < -1)$



$= 0.5 - P(-1 \leq Z \leq 0)$

$= 0.5 - P(0 \leq Z \leq 1)$ ∈ Symmetry

$= 0.5 - 0.3413 = 0.1587$ (from tables)

∴ Number of candidates scoring less than 30

$= 1000 \times 0.1587 = 159$

Example 27

The I.Q (intelligence quotient) of a group of 1000 school children has mean 96 and the standard deviation 12.

Assuming that the distribution of I.Q among school children is normal, find approximately the number of school children having I.Q.

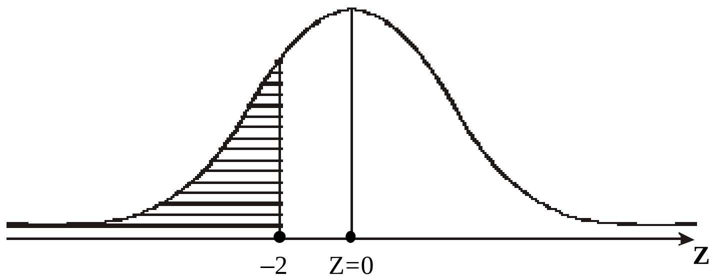
(i) less than 72   (ii) between 80 and 120

Solution :

Given N = 1000, $\mu = 96$ and $\sigma = 12$

Then $Z = \frac{X-\mu}{\sigma} = \frac{X-96}{12}$

(i) $P(X < 72) = P(Z < -2)$



$= P(-\infty < Z \leq 0) - P(-2 \leq Z \leq 0)$

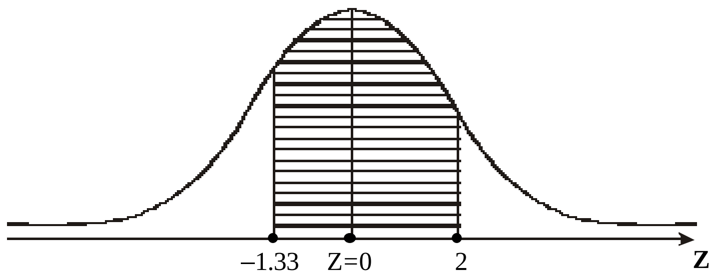
$= P(0 \leq Z < \infty) - P(0 \leq Z \leq 2)$ (By symmetry)

$= 0.5 - 0.4772$ (from tables) $= 0.0228$.

$\therefore$ Number of school children having I.Q less than 72

$= 1000 \times 0.0228 = 22.8 \simeq 23$

(ii) $P(80 < X < 120) = P(-1.33 < Z < 2)$



$= P(-1.33 \leq Z \leq 0) + P(0 \leq Z \leq 2)$

$= P(0 \leq Z \leq 1.33) + P(0 \leq Z \leq 2)$

$= .4082 + .4772$ (from tables)

$= 0.8854$

$\therefore$ Number of school children having I.Q. between 80 and 120

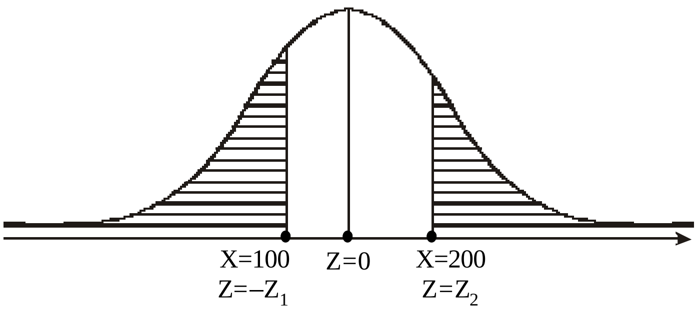
$= 1000 \times .8854 = 885.$

Exercise 28

In a normal distribution 20% of the items are less than 100 and 30% are over 200. Find the mean and S.D of the distribution.

Solution :

Representing the given data diagramtically,



From the diagram

$P(-Z_1 < Z < 0) = 0.3$

i.e. $P(0 < Z < Z_1) = 0.3$

$\therefore$ $Z_1 = 0.84$ (from the normal table)

Hence $-0.84 = \dfrac{100-\mu}{\sigma}$

i.e. $100 - \mu = -0.84\sigma$ ----------(1)

$P(0 < Z < Z_2) = 0.2$

$\therefore$ $Z_2 = 0.525$ (from the normal table)

Hence $\quad 0.525 = \dfrac{200 - \mu}{\sigma}$

i.e. $200 - \mu = 0.525\sigma$ ----------(2)

Solving (1) and (2), $\quad \mu = 161.53$

$\sigma = 73.26$

EXERCISE 8.4

1) Find the area under the standard normal curve which lies
   (i) to the right of $Z = 2.70$
   (ii) to the left of $Z = 1.73$

2) Find the area under the standard normal curve which lies
   (i) between $Z = 1.25$ and $Z = 1.67$
   (ii) between $Z = -0.90$ and $Z = -1.85$

3) The distribution of marks obtained by a group of students may be assumed to be normal with mean 50 marks and standard deviation 15 marks. Estimate the proportion of students with marks below 35.

4) The marks in Economics obtained by the students in Public examination is assumed to be approximately normally distributed with mean 45 and S.D 3. A student taking this subject is chosen at random. What is the probability that his mark is above 70?

5) Assuming the mean height of soldiers to be 68.22 inches with a variance 10.8 inches. How many soldiers in a regiment of 1000 would you expect to be over 6 feet tall ?

6) The mean yield for one-acre plot is 663 kgs with a S.D 32 kgs. Assuming normal distribution, how many one-acre plot in a batch of 1000 plots would you expect to have yield (i) over 700 kgs (ii) below 650 kgs.

7) A large number of measurements is normally distributed with a mean of 65.5" and S.D of 6.2". Find the percentage of measurements that fall between 54.8" and 68.8".

8) The diameter of shafts produced in a factory conforms to normal distribution. 31% of the shafts have a diameter less than 45mm. and 8% have more than 64mm. Find the mean and standard deviation of the diameter of shafts.

9) The results of a particular examination are given below in a summary form.

Result	percentage of candidates
1. passed with distinction	10
2. passed	60
3. failed	30

It is known that a candidate gets plucked if he obtained less tham 40 marks out of 100 while he must obtain atleast 75 marks in order to pass with distinction. Determine the mean and the standard deviation of the distribution assuming this to be normal.

EXERCISE 8.5

Choose the correct answer

1) If a fair coin is tossed three times the probability function $p(x)$ of the number of heads x is

(a)

x	0	1	2	3
$p(x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$

(b)

x	0	1	2	3
$p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

(c)

x	0	1	2	3
$p(x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$

(d) none of these

2) If a discrete random variable has the probability mass function as

x	0	1	2	3
$p(x)$	k	$2k$	$3k$	$5k$

then the value of k is

(a) $\frac{1}{11}$ (b) $\frac{2}{11}$ (c) $\frac{3}{11}$ (d) $\frac{4}{11}$

3) If the probability density function of a variable X is defined as $f(x) = Cx\,(2 - x)$, $0 < x < 2$ then the value of C is

(a) $\frac{4}{3}$ (b) $\frac{6}{4}$ (c) $\frac{3}{4}$ (d) $\frac{3}{5}$

4) The mean and variance of a binomial distribution are

(a) $np, \; npq$ (b) $pq, \; npq$ (c) $np, \sqrt{npq}$ (d) np, nq

5) If X~N (μ, σ), the standard Normal variate is distributed as

(a) N(0, 0) (b) N(1, 0) (c) N(0, 1) (d) N(1, 1)

6) The normal distribution curve is

(a) Bimodal (b) Unimodal

(c) Skewed (d) none of these

7) If X is a poission variate with P(X = 1) = P(X = 2), the mean of the Poisson variate is equal to

(a) 1 (b) 2 (c) –2 (d) 3

8) The standard deviation of a Poissson variate is 2, the mean of the poisson variate is

(a) 2 (b) 4 (c) $\sqrt{2}$ (d) $\frac{1}{\sqrt{2}}$

9) The random variables X and Y are independent if

(a) E(X Y) = 1 (b) E(XY) = 0

(c) E(X Y) = E(X) E(Y) (d) E(X+Y) = E(X) + E(Y)

10) The mean and variance of a binomial distribution are 8 and 4 respectively. Then P(X = 1) is equal to

(a) $\frac{1}{2^{12}}$ (b) $\frac{1}{2^{4}}$ (c) $\frac{1}{2^{6}}$ (d) $\frac{1}{2^{10}}$

11) If X~N (μ, σ^2), the points of inflection of normal distribution curve are

(a) $\pm\,\mu$ (b) $\mu \pm \sigma$ (c) $\sigma \pm \mu$ (d) $\mu \pm 2\sigma$

12) If X~N (μ, σ^2), the maximum probability at the point of inflection of normal distribution is

(a) $\frac{1}{\sqrt{2\pi}}e^{\frac{1}{2}}$ (b) $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}}$ (c) $\frac{1}{\sigma\sqrt{2\pi}}$ (d) $\frac{1}{\sqrt{2\pi}}$

13) If a random variable X has the following probability distribution

X	-1	-2	1	2
p(x)	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

then the expected value of X is

(a) $\frac{3}{2}$ (b) $\frac{1}{6}$ (c) $\frac{1}{2}$ (d) $\frac{1}{3}$

14) If X~N (5, 1), the probability density function for the normal variate X is

(a) $\frac{1}{5\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-1}{5})^2}$ (b) $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-1}{5})^2}$

(c) $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-5)^2}$ (d) $\frac{1}{\sqrt{\pi}}e^{-\frac{1}{2}(x-5)^2}$

15) If X~N (8, 64), the standard normal variate Z will be

(a) $z = \frac{X-64}{8}$ (b) $\frac{X-8}{64}$

(c) $\frac{X-8}{8}$ (d) $\frac{X-8}{\sqrt{8}}$

SAMPLING TECHNIQUES AND STATISTICAL INFERENCE 9

9.1 SAMPLING AND TYPES OF ERRORS

Sampling is being used in our everyday life without knowing about it. For examples, a cook tests a small quantity of rice to see whether it has been well cooked and a grain merchant does not examine each grain of what he intends to purchase, but inspects only a small quantity of grains. Most of our decisions are based on the examination of a few items only.

In a statistical investigation, the interest usually lies in the assessment of general magnitude and the study of variation with respect to one or more characteristics relating to individuals belonging to a group. This group of individuals or units under study is called **population** or **universe.** Thus in statistics, population is an aggregate of objects or units under study. The population may be finite or infinite.

9.1.1 Sampling and sample

Sampling is a method of selecting units for analysis such as households, consumers, companies etc. from the respective population under statistical investigation. The theory of sampling is based on the **principle of statistical regularity.** According to this principle, a moderately large number of items chosen at random from a large group are almost sure on an average to possess the characteristics of the larger group.

A smallest non-divisible part of the population is called a **unit.** A unit should be well defined and should not be ambiguous. For example, if we define unit as a household, then it should be defined that a person should not belong to two households nor should it leave out persons belonging to the population.

A finite subset of a population is called a **sample** and the number of units in a sample is called its **sample size**.

By analysing the data collected from the sample one can draw inference about the population under study.

9.1.2 Parameter and Statistic

The statistical constants of a population like mean (μ), variance (σ^2), proportion (P) are termed as **parameters.** Statistical measures like mean ($\overline{X}$), variance (s^2), proportion (*p*) computed from the sampled observations are known as **statistics.**

Sampling is employed to throw light on the population parameter. A **statistic** is an estimate based on sample data to draw inference about the population parameter.

9.1.3 Need for Sampling

Suppose that the raw materials department in a company receives items in lots and issues them to the production department as and when required. Before accepting these items, the inspection department inspects or tests them to make sure that they meet the required specifications. Thus

(i) it could inspect all items in the lot or

(ii) it could take a sample and inspect the sample for defectives and then estimate the total number of defectives for the population as a whole.

The first approach is called **complete enumeration (census)**. It has two major disadvantages namely, the time consumed and the cost involved in it.

The second approach that uses sampling has two major advantages. (i) It is significantly less expensive. (ii) It takes least possible time with best possible results.

There are situations that involve destruction procedure where sampling is the only answer. A well-designed statistical sampling methodology would give accurate results and at the same time will result in cost reduction and least time. Thus sampling is the best available tool to decision makers.

9.1.4 Elements of Sampling Plan

The main steps involved in the planning and execution of sample survey are :

(i) Objectives

The first task is to lay down in concrete terms the basic objectives of the survey. Failure to define the objective(s) will clearly undermine the purpose of carrying out the survey itself. For example, if a nationalised bank wants to study savings bank account holders perception of the service quality rendered over a period of one year, the objective of the sampling is, here,to analyse the perception of the account holders in the bank.

(ii) Population to be covered

Based on the objectives of the survey, the population should be well defined. The characteristics concerning the population under study should also be clearly defined. For example, to analyse the perception of the savings bank account holders about the service rendered by the bank, all the account holders in the bank constitute the population to be investigated.

(iii) Sampling frame

In order to cover the population decided upon, there should be some list, map or other acceptable material (called the **frame**) which serves as a guide to the population to be covered. The list or map must be examined to be sure that it is reasonably free from defects. The sampling frame will help us in the selection of sample. All the account numbers of the savings bank account holders in the bank are the sampling frame in the analysis of perception of the customers regarding the service rendered by the bank.

(iv) Sampling unit

For the purpose of sample selection, the population should be capable of being divided up into sampling units. The division of the population into sampling units should be unambiguous. Every element of the population should belong to just one sampling unit.

Each account holder of the savings bank account in the bank, form a unit of the sample as all the savings bank account holders in the bank constitute the population.

(v) Sample selection

The size of the sample and the manner of selecting the sample should be defined based on the objectives of the statistical investigation. The estimation of population parameter along with their margin of uncertainity are some of the important aspects to be followed in sample selection.

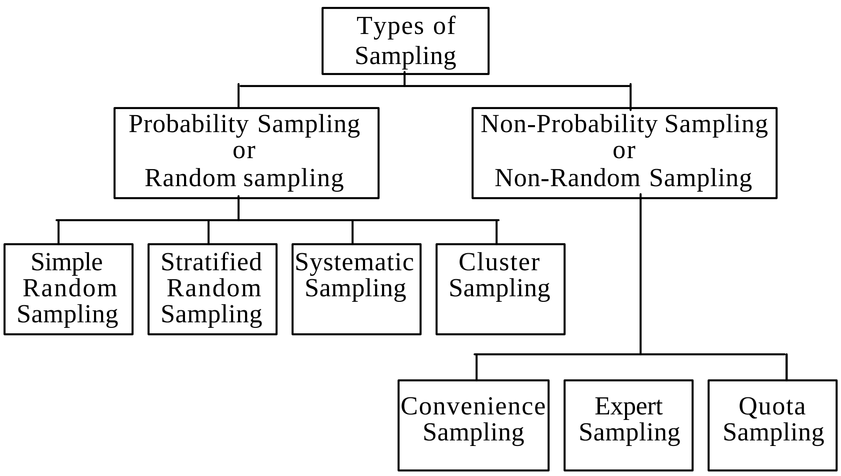
(vi) Collection of data

The method of collecting the information has to be decided, keeping in view the costs involved and the accuracy aimed at. Physical observation, interviewing respondents and collecting data through mail are some of the methods that can be followed in collection of data.

(vii) Analysis of data

The collected data should be properly classified and subjected to an appropriate analysis. The conclusions are drawn based on the results of the analysis.

9.1.5 Types of Sampling



The technique of selecting a sample from a population usually depends on **the nature of the data** and **the type of enquiry**. The procedure of sampling may be broadly classified under the following heads :

(i) Probability sampling or random sampling and
(ii) Non-probability sampling or non-random sampling.

(i) Probability sampling

Probability sampling is a method of sampling that ensures that every unit in the population has a known **non-zero chance** of being included in the sample.

The different methods of random sampling are :

(a) Simple Random Sampling

Simple random sampling is the foundation of probability sampling. It is a special case of probability sampling in which every unit in the population has an **equal chance** of being included in a sample. Simple random sampling also makes the selection of every possible combination of the desired number of units equally likely. Sampling may be done with or without replacement.

It may be noted that when the sampling is with replacement, the units drawn are replaced before the next selection is made.The population size remains constant when the sampling is with replacement.

If one wants to select n units from a population of size N without replacement,then every possible selection of n units must have the same probability. Thus there are ${}^{N}c_{n}$ possible ways to pick up n units from the population of size N. Simple random sampling guarantees that a sample of n units, has the same probability $\frac{1}{{}^{N}c_{n}}$ of being selected.

Example

A bank wants to study the Savings Bank account holders perception of the service quality rendered over a period of one year. The bank has to prepare a complete list of savings bank account holders, called as **sampling frame,** say 500. Now the process involves selecting a sample of 50 out of 500 and interviewing them. This could be achieved in many ways. Two common ways are :

(1) Lottery method : Select 50 slips from a box containing well shuffled 500 slips of account numbers without replacement. This method can be applied when the population is small enough to handle.

(2) Random numbers method : When the population size is very large, the most practical and inexpensive method of selecting a simple random sample is by using the random number tables.

(b) Stratified Random Sampling

Stratified random sampling involves dividing the population into a number of groups called **strata** in such a manner that the units within a stratum are **homogeneous** and the units between the strata are **hetrogeneous**. The next step involves selecting a simple random sample of appropriate size from each stratum. The sample size in each stratum is usually of (a) equal size, (b) proportionate to the number of units in the stratum.

For example, a marketing manager in a consumer product company wants to study the customer's attitude towards a new product in order to improve the sales. Then three typical cities that will influence the sales will be considered as three strata. The customers within a city are similar but between the cities are vastly different. Selection of the customers for the study from each city has to be a random sample to draw meaningful inference on the whole population.

(c) Systematic sampling

Systematic sampling is a convenient way of selecting a sample. It requires less time and cost when compared to simple random sampling.

In this method, the units are selected from the population at a uniform interval. To facilitate this we arrange the items in numerical, alphabetical, geographical or any other order. When a complete list of the population is available, this method is used.

If we want to select a sample of size n from a population of size N under systematic sampling, frist select an item j at random such that $1 \le j \le k$ where $k = \frac{N}{n+1}$ and k is the nearest possible integer. Then $j, j+k, j+2k, \ldots, j+(n-1)k$ th items constitute a systematic random sample.

For example, if we want to select a sample of 9 students out of 105 students numbered as 1, 2, ..., 105 , select a student among 1, 2, ... , 11 at random (say at 3rd position). Here $k = \frac{105}{10} = 10.5$ and $\therefore$ $k = 11$. Hence students at the positions 3, 14, 25, 36, 47, 58, 69, 80, 91 form a random sample of size 9.

(d) Cluster sampling

Cluster sampling is used when the population is divided into **groups** or **clusters** such that each cluster is a representative of the population.

If a study has to be done to find out the number of children that each family in Chennai has, then the city can be divided into several clusters and a few clusters can be chosen at random. Every family in the chosen clusters can be a sample unit.

In using cluster sampling the following points should be noted

(a) For getting precise results clusters should be as small as possible consistent with the cost and limitations of the survey and

(b) The number of units in each cluster must be more or less equal.

(ii) Non-Probability Sampling

The fundamental difference between probability sampling and non-probability sampling is that in non-probability sampling procedure, the selection of the sample units does not ensure a known chance to the units being selected. In other words the units are selected without using the **principle of probability.** Even though the non-probability sampling has advantages such as reduced cost, speed and convenience in implementation, it lacks accuracy in view of the seleciton bias. Non-probability sampling is suitable for pilot studies and exploratory research

The methods of **non-random sampling** are :

(a) Purposive sampling

In this sampling, the sample is selected with definite purpose in view and the choice of the sampling units depends entirely on the discretion and judgement of the investigator.

For example, if an investigator wants to give the picture that the standard of living has increased in the city of Madurai, he may take the individuals in the sample from the posh localities and ignore the localities where low income group and middle class families live.

(b) Quota sampling

This is a restricted type of purposive sampling. This consists in specifying quotas of the samples to be drawn from different groups and then drawing the required samples from these groups by purposive sampling. Quota sampling is widely used in opinion and market research surveys.

(c) Expert opinion sampling or expert sampling

Expert opinion sampling involves gathering a set of people who have the knowledge and expertise in certain key areas that are crucial to decision making. The advantage of this sampling is that it acts as a support mechanism for some of our decisions in situations where virtually no data are available. The major disadvantage is that even the experts can have prejudices, likes and dislikes that might distort the results.

9.1.6 Sampling and non-sampling errors

The errors involved in the collection of data, processing and analysis of data may be broadly classfied as (i) **sampling errors** and (ii) **non-sampling errors.**

(i) Sampling errors

Sampling errors have their origin in sampling and arise due to the fact that only a part of the population has been used to estimate population parameters and draw inference about the population. Increasing in the sample size usually results in decrease in the sampling error.

Sampling errors are primarily due to some of the following reasons :

(a) **Faulty selection of the sample**

Some of the bias is introduced by the use of defective sampling technique for the selection of a sample in which the investigator deliberately selects a representative sample to obtain certain results.

(b) **Substitution**

If difficulty arise in enumerating a particular sampling unit included in the random sample, the investigators usually substitute a convenient member of the population leading to sampling error.

(c) **Faulty demarcation of sampling units**

Bias due to defective demarcation of sampling units is particularly significant in area surveys such as agricultural experiments. Thus faulty demarcation could cause sampling error.

(ii) Non-sampling errors

The non-sampling errors primarily arise at the stages of observation, classification and analysis of data.

Non-sampling errors can occur at every stage of the planning or execution of census or sample surveys. Some of the more important non-sampling errors arise from the following factors :

(a) **Errors due to faulty planning and definitions**
Sampling error arises due to improper data specification, error in location of units, measurement of characteristics and lack of trained investigators.

(b) **Response errors**

These errors occur as a result of the responses furnished by the respondents.

(c) **Non-response bias**

Non-response biases occur due to incomplete information on all the sampling units.

(d) **Errors in coverage**

These errors occur in the coverage of sampling units.

(e) **Compiling errors**

These errors arise due to compilation such as editing and coding of responses.

EXERCISE 9.1

1) Explain sampling distribution and standard error.
2) Distinguish between the terms parameter and statistic
3) Explain briefly the elements of sampling plan.
4) Discuss probability sampling.
5) Discuss non-probability sampling.
6) Distinguish between sampling and non sampling errors.

9.2 SAMPLING DISTRIBUTIONS

Consider all possible samples of size n which can be drawn from a given population. For each sample we can compute a statistic such as mean, standard deviation, etc. which will vary from sample to sample. The aggregate of various values of the statistic under consideration may be grouped into a frequency distribution. This distribution is known as **sampling distribution** of the statistic. Thus the probability distribution of all the possible values that a sample statistic can take, is called the sampling distribution of the statistic. **Sample mean** and **sample proportion** based on a random sample are examples of sample statistic.

Supposing a Market Research Agency wants to estimate the annual household expenditure on consumer durables from among the population of households (say 50000 households) in Tamil Nadu. The agency can choose fifty different samples of 50 households each. For each of the samples, we can calculate the mean annual expenditure on consumer durables as given in the following table :

Sample No.	Total expenditure for 50 households	Mean Rs.
1	100000	2000
2	300000	6000
3	200000	4000
4	150000	3000
.	.	.
.	.	.
.	.	.
49	600000	12000
50	400000	8000

The distribution of all the sample means is known as the **sampling distribution of the mean.** The figures Rs. 2000, 6000 ... 8000 are the sampling distribution of the means.

Similarly, the sampling distribution of the **sample variance** and **sample proportion** can also be obtained.

In a sample of n items if n_1 belongs to Category-1 and n-n_1 belongs to the Category-2, then $\frac{n_1}{n}$ is defined as the sample proportion p belonging to the first category and $\frac{n-n_1}{n}$ or (1-p) is the sample proportion of second category. This concept could be extended to k such categories with proportions (say) p_1, p_2, ... , p_k such that $p_1 + p_2 + \ldots + p_k = 1$

We could also arrive at a sampling distribution of a **proportion.** For example, if in a factory producing electrical switches, 15 different samples of 1000 switches are taken for inspection and number of defectives in each sample could be noted. We could find a probability distribtuion of the **proportion** of defective switches.

9.2.1 Sampling distribution of the Mean from normal population

If $X_1, X_2, \ldots, X_n$ are n independent random samples drawn from a normal population with mean μ and standard deviation σ, then the sampling distribution of $\overline{X}$ (the sample mean) follows a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

It may be noted that

(i) the sample mean $\overline{X} = \frac{\Sigma X_i}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$

Thus $\overline{X}$ is a random variable and will be different every time when a new sample of n observations are taken

(ii) $\overline{X}$ is an unbiased estimator of the population mean μ.

i.e. $E(\overline{X}) = \mu$, denoted by $\mu_{\overline{X}} = \mu$.

(iii) the standard deviation of the sample mean $\overline{X}$ is given by

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

For example, consider a sample of weights of four boys from the normal population of size 10000 with replacement. The mean weight of the four boys is worked out. Again take another new sample of four boys from the same population and find the mean weight. If the process is repeated an infinite number of times, the probability distribution of these infinite number of sample means would be sampling distribution of mean.

9.2.2 Central limit theorem

When the samples are drawn from a normal population with mean μ and standard deviation σ, the sampling distribution of the mean is also normal with mean $\mu_{\overline{X}} = \mu$ and standard deviation $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$. However the sampling distribution of the mean, when the population is not normal is equally important.

The central limit theorem says that from any given population with mean μ and standard deivation σ, if we draw a random sample of *n* observations, the sampling distribution of the mean will approach a normal distribution with a mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ as the sample size increases and becomes large.

In practice a sample size of 30 and above is considered to be large.

Thus the central limit theorem is a hall mark of statistical inference. It permits us to make inference about the population parameter based on random samples drawn from populations that are not neccessarily normally distributed.

9.2.3 Sampling distribution of proportions

Suppose that a population is infinite and that the probability of occurrence of an event, say success, is P. Let Q = 1 – P denote the probability of failure.

Consider all possible samples of size *n* drawn from this population. For each sample, determine the proportion *p* of successes. Applying central limit theorem, if the sample size *n* is large, the distribution of the sample porportion *p* follows a normal distribution with mean $\mu_p = P$ and S.D $\sigma_p = \sqrt{\frac{PQ}{n}}$.

9.2.4 Standard error

The standard deviation of the sampling distribution of a statistic is called the **standard error** of the statistic. The standard deviation of the distribution of the sample means is called the **standard error of the mean**. Likewise, the standard deviation of the distribution of the sample proportions is called the **standard error of the proportion.**

The standard error is popularly known as **sampling error**. Sampling error throws light on the precision and accuracy of the estimate. The standard error is inversely proportional to the sample size i.e. the larger the sample size the samller the standard error.

The standard error measures the dispersion of all possible values of the statistic in repeated samples of a fixed size from a given population. It is used to set up confidence limits for population parameters in tests of significance. Thus the standard errors of sample mean $\overline{X}$ and sample proportion p are used to find confidence limits for the population mean μ and the population proportion P respectively.

Statistic	**Standard error**	**Remarks**
Sample mean $\overline{X}$	$\frac{\sigma}{\sqrt{n}}$	Population size is infinite or sample with replacement.
	$\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$	Population size N finite or sample without replacement
Sample proportion p	$\sqrt{\frac{PQ}{n}}$	Population size is infinite or sample with replacement.
	$\sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$	Population size N finite or sample without replacement

Example 1

A population consists of the five numbers 2, 3, 6, 8, 11. Consider all possible samples of size 2 which can be drawn with replacement from this population. find (i) mean of the population, (i) the standard deviation of the population (iii) the mean of the sample distribution of means and (iv) the standard error of means.

Solution :

(i) The population mean $\mu = \frac{\Sigma x}{N} = \frac{2+3+6+8+11}{5} = 6$

(ii) The variance of the population $\sigma^2 = \frac{1}{N}\Sigma(x - \mu)^2$

$$= \frac{1}{5}\{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2\}$$
$$= 10.8$$

$\therefore$ the standard deviation of the population $\sigma = 3.29$

(iii) There are 25 samples of size two which can be drawn with replacement. They are

(2, 2)	(2, 3)	(2, 6)	(2, 8)	(2, 11)
(3, 2)	(3, 3)	(3, 6)	(3, 8)	(3, 11)
(6, 2)	(6, 3)	(6, 6)	(6, 8)	(6, 11)
(8, 2)	(8, 3)	(8, 6)	(8, 8)	(8, 11)
(11, 2)	(11, 3)	(11, 6)	(11, 8)	(11, 11)

The corresponding sample means are

2.0	2.5	4.0	5.0	6.5
2.5	3.0	4.5	5.5	7.0
4.0	4.5	6.0	7.0	8.5
5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

The mean of sampling distribution of means

$$\mu_{\overline{x}} = \frac{\text{sum of all sample means}}{25} = \frac{150}{25} = 6.0$$

(iv) The variance $\sigma^2_{\overline{x}}$ of the sampling distribution of means is obtained as follows :

$$\sigma^2_{\overline{x}} = \frac{1}{25}\{2-6)^2 + (2.5-6)^2 + \ldots + (6.5-6)^2 + \ldots$$
$$+ (9.5-6)^2 + (11-6)^2\}$$
$$= \frac{135}{25} = 5.4$$

$\therefore$ the standard error of means $\sigma_{\overline{x}} = \sqrt{5.4} = 2.32$

Example 2

Assume that the monthly savings of 1000 employees working in a factory are normally distributed with mean Rs. 2000 and standard deviation Rs. 50 If 25 samples consisting of 4 employees each are obtained, what would be the mean and standard deviation of the resulting sampling distribution of means if sampling were done (i) with replacement, (ii) without replacement.

Solution :

Given N = 1000, $\mu = 2000$, $\sigma = 50$, $n = 4$

(i) *Sampling with replacement*

$$\mu_{\overline{x}} = \mu = 2000$$

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{4}} = 25$$

(ii) *Sampling without replacement*

The mean of the sampling distribution of the means is

$$\mu_{\overline{x}} = \mu = 2000$$

The standard deviation of the sampling distribution of means is

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$= \frac{50}{\sqrt{4}}\sqrt{\frac{100-4}{1000-1}}$$

$$= (25)\sqrt{\frac{996}{999}} = 25\left(\sqrt{.996}\right)$$

$$= (25)\ (0.9984) = 24.96$$

Example 3

A random sample of size 5 is drawn without replacement from a finite population consisting of 41 units. If the population S.D is 6.25, find the S.E of the sample mean.

Solution :

Population size N = 41

Sample size $n = 5$

Standard deviation of the popultion $\sigma = 6.25$

S.E of sample mean $= \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ (N is finite)

$$= \frac{6.25}{\sqrt{5}}\sqrt{\frac{41-5}{41-1}}$$

$$= \frac{6.25 \times 6}{\sqrt{5}\ 2\sqrt{10}} = \frac{3 \times 6.25}{5\sqrt{2}} = 2.65$$

Example 4

The marks obtained by students in an aptitude test are normally distributed with a mean of 60 and a standard deviation of 30. A random sample of 36 students is drawn from this population.

(i) What is the standard error of the sampling mean?

(ii) What is the probability that the mean of a sample of 16 students will be either less than 50 or greater than 80?

[P (0 < Z < 4) = 0.4999]

Solution :

(i) The standard error of the sample mean $\overline{X}$ is given by

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{36}} = 5 \qquad \text{(N is not given)}$$

(ii) The random variable $\overline{X}$ follows normal distribution with mean $\mu_{\overline{X}}$ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

To find P($\overline{X}$<50 or $\overline{X}$>80).

P($\overline{X}$<50 or $\overline{X}$>80) = P($\overline{X}$<50) + P($\overline{X}$>80)

$$= P\left(\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{50-60}{5}\right) + P\left(\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} > \frac{80-60}{5}\right)$$

$= P(Z < -2) + P(Z > 4)$

$= [0.5 - P(0 < Z < 2)] + [0.5 - P(0 < Z < 4)]$

$= (0.5 - 0.4772) + (0.5 - 0.4999)$

$= .02283$, which is the required probability.

Example 5

2% of the screws produced by a machine are defective. What is the probability that in a consignment of 400 such screws, 3% or more will be defective.

Solution :

Here N is not given, but $n = 400$

Population proportion $P = 2\% = 0.02 \quad \therefore Q = 1 - P = 0.98$

The sample size is large

$\therefore$ The sample porportion is normally distributed with mean

$$\mu_p = 0.02 \text{ and S.D} = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.02 \times 0.98}{400}} = 0.007$$

Probability that the sample proportion $p \geq 0.03$

= Area under the normal curve to the right of $Z = 1.43$.

$$\left(Z = \frac{p - P}{S.D} = \frac{0.03 - 0.02}{0.007} = 1.43\right)$$

$\therefore$ required probability = 0.5 – Area between Z = 0 to Z = 1.43

$= 0.5 - 0.4236 = 0.0764$

EXERCISE 9.2

1) A population consists of four numbers 3, 7, 11 and 15. Consider all possible samples of size two which can be drawn with replacement from this population.

Find (i) the population mean

(ii) the population standard deviation.

(iii) the mean of the sampling distribution of mean

(iv) the standard deviation of the sampling distribution of mean.

2) A population consists of four numbers 3, 7, 11 and 15. Consider all possible samples of size two which can be drawn without replacement from this population.

Find (i) the population mean

(ii) the population standard deviation.

(iii) the mean of the sampling distribution of mean

(iv) the standard deviation of the sampling distribution of mean.

3) The weights of 1500 iron rods are normally distributed with mean of 22.4 kgs. and standard deviation of 0.048 kg. If 300 random samples of size 36 are drawn from this population, determine the mean and standard deviation of the sampling distribution of mean when sampling is done (i) with replacement (ii) without replacement.

4) 1% of the outgoing +2 students in a school have joined I.I.T. Madras. What is the probability that in a group of 500 such students 2% or more will be joining I.I.T. Madras.

9.3 ESTIMATION

The technique used for generalising the results of the sample to the population is provided by an important branch of statistics called **statistical inference.** The concept of statistical inference deals with two basic aspects namely (a) **Estimation** and (b) **Testing of hypothesis.**

In statistics, estimation is concerned with making inference about the parameters of the population using information available in the samples. The parameter estimation is very much needed in the decision making process.

The estimation of population parameters such as mean, variance, proportion, etc. from the correspoinding sample statistics is an important function of statistical inference.

9.3.1 Estimator

A sample statistic which is used to estimate a population parameter is known as **estimator.**

A good estimator is one which is as close to the true value of population parameter as possible. A good estimator possesses the following properties:

(i) Unbiasedness

As estimate is said to be unbiased if its expected value is equal to its parameter.

The sample mean $\overline{X} = \frac{1}{n}\Sigma x$ is an unbiased estimator of population mean μ. For a sample of size n, drawn from a population of size N, $s^2 = \frac{1}{n-1} S(x - \bar{x})^2$ is an unbiased estimator of population variance. Hence s^2 is used in **estimation** and in **testing of hypothesis.**

(ii) Consistency

An estimator is said to be consistent if the estimate tends to approach the parameter as the sample size increases.

(iii) Efficiency

If we have two unbiased estimators for the same population prarameter, the first estimator is said to be more efficient than the second estimator if the standard error of the first estimator is smaller than that of the second estimator for the same sample size.

(iv) Sufficiency

If an estimator possesses all information regarding the parameter, then the estimator is said to be a sufficient estimator.

9.3.2 Point Estimate and Interval Estimate

It is possible to find two types of estimates for a population parameter. They are **point estimate** and **interval estimate.**

Point Estimate

An estimate of a population parameter given by a single number is called a point estimator of the parameter. Mean ($\bar{x}$) and the sample variance $[s^2 = \frac{1}{n-1} \Sigma(x-\bar{x})^2]$ are the examples of point estimates.

A point estimate will rarely coincide with the true population parameter value.

Interval Estimate

An estimate of a population parameter given by two numbers between which the parameter is expected to lie is called an interval estimate of the parameter.

Interval estimate indicates the accuracy of an estimate and is therefore preferable to point estimate. As point estimate provides a single value for the population parameter it may not be suitable in some situation.

For example,

if we say that a distance is measured as 5.28mm, we are giving a point estimate. On the other hand, if we say that the distance is 5.28 $\pm$ 0.03 mm i.e. the distance lies between 5.25 and 5.31mm, we are giving an interval estimate.

9.3.3 Confidence Interval for population mean and proportion

The interval within which the unknown value of parameter is expected to lie is called **confidence interval.** The limits so determined are called **confidence limits.**

Confidence intervals indicate the probability that the population parameter lies within a specified range.

Computation of confidence interval

To compute confidence interval we require

(i) the sample statistic,

(ii) the standard error (S.E) of sampling distribution of the statistic

(iii) the degree of accuracy reflected by the Z-value.

If the size of sample is sufficiently large, then the sampling distribution is approximately normal. Therefore, the sample value can be used in estimation of standard error in the place of population value. The Z-distribution is used in case of large samples to estimate the confidence limits.

We give below values of Z corresponding to some confidence levels.

Confidence Levels	99%	98%	96%	95%	80%	50%
Value of Z, Z_c	2.58	2.33	2.05	1.96	1.28	0.674

(i) Confidence interval estimates for means

Let μ and σ be the population mean and standard deviation of the population.

Let $\overline{X}$ and s be the sample mean and standard deviation of the sampling distribution of a statistic.

The confidence limits for m are given below :

Population size	Sample size	confidence limits for μ.
Infinite	n	$\overline{X} \pm (Z_C)\frac{s}{\sqrt{n}}$, z_c is the value of Z corresponding to confidence levels.
Finite, N	n	$\overline{X} \pm (Z_C)\frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

(ii) Confidence intervals for proportions

If p is the proportion of successes in a sample of size n drawn from a population with P as its proportion of successes, then the confidence intervals for P are given below :

Population	Sample size	Confidence limits for P
Infinite	n	$p \pm (Z_C) \sqrt{\frac{pq}{n}}$
Finite, N	n	$p \pm (Z_C) \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$

Example 6

Sensing the downward trend in demand for a leather product, the financial manager was considering shifting his company's resources to a new product area. He selected a sample of 10 firms in the leather industry and discovered their earnings (in %) on investment. Find point estimate of the mean and variance of the population from the data given below.

21.0 25.0 20.0 16.0 12.0 10.0 17.0 18.0 13.0 11.0

Solution :

X	$\overline{X}$	X- $\overline{X}$	(X- $\overline{X}$)^2
21.0	16.3	4.7	22.09
25.0	16.3	8.7	75.69
20.0	16.3	3.7	13.69
16.0	16.3	-0.3	0.09
12.0	16.3	-4.3	18.49
10.0	16.3	-6.3	39.69
17.0	16.3	0.7	0.49
18.0	16.3	1.7	2.89
13.0	16.3	-3.3	10.89
11.0	16.3	-5.3	28.09
163.0			**212.10**

Sample mean, $\overline{X} = \frac{\Sigma X}{n} = \frac{163}{10} = 16.3$

Sample variance, $s^2 = \frac{1}{n-1}\Sigma(X-\overline{X})^2$

$= \frac{212.10}{n-1} = 23.5$ (the sample size is small)

Sample standard deviation = $\sqrt{23.5} = 4.85$

Thus the point estimate of mean and of variance of the population from which the samples are drawn are 16.3 and 23.5 respectively.

Example 7

A sample of 100 students are drawn from a school. The mean weight and variance of the sample are 67.45 kg and 9 kg. respectively. Find (a) 95% and (b) 99% confidence intervals for estimating the mean weight of the students.

Solution :

Sample size, $n = 100$

The sample mean, $\overline{X} = 67.45$

The sample variance $s^2 = 9$

The sample standard deviation $s = 3$

Let μ be the population mean.

(a) The 95% confidence limits for μ are given by

$$\overline{X} \pm (Z_c) \frac{s}{\sqrt{n}}$$

$$\Rightarrow \quad 67.45 \pm (1.96) \frac{3}{\sqrt{100}}$$ (Here $Z_c = 1.96$ for 95% confidence level)

$$\Rightarrow \quad 67.45 \pm 0.588$$

Thus the 95% confidence intervals for estimating μ is given by

(66.86, 68.04)

(b) The 99% confidence limits for estimating μ are given by

$$\overline{X} \pm (Z_c) \frac{s}{\sqrt{n}}$$

$$\Rightarrow \quad 67.45 \pm (2.58) \frac{3}{\sqrt{100}}$$ (Here $z_c = 2.58$ for 99% confidence level)

$$\Rightarrow \quad 67.45 \pm 0.774$$

Thus the 99% confidence interval for estimating μ is given by

(66.67, 68.22)

Example 8

A random sample of size 50 with mean 67.9 is drawn from a normal population. If it is known that the standard error of the sample mean is $\sqrt{0.7}$, find 95% confidence interval for the population mean.

Solution :

$n = 50$, sample mean $\overline{X} = 67.9$

95% confidence limits for population mean μ are :

$\overline{X} \pm (Z_c)\{S.E(\overline{X})\}$

$\Rightarrow \quad 67.9 \pm (1.96)(\sqrt{0.7})$

$\Rightarrow \quad 67.9 \pm 1.64$

Thus the 95% confidence intervals for estimating μ is given by

(66.2, 69.54)

Example 9

A random sample of 500 apples was taken from large consignment and 45 of them were found to be bad. Find the limits at which the bad apples lie at 99% confidence level.

Solution :

We shall find confidence limits for the proportion of bad apples.

Sample size $n = 500$

Proportion of bad apples in the sample = $\frac{45}{500} = 0.09$

p = 0.09

∴ Proportion of good apples in the sample $q = 1-p = 0.91$.

The confidence limits for the population proportion P of bad apples are given by

$$p \pm (Z_c)\left(\sqrt{\frac{pq}{n}}\right)$$

$$\Rightarrow \; 0.09 \pm (2.58) \sqrt{\frac{(.09)(0.91}{500}} \quad \Rightarrow \; 0.09 \pm 0.033$$

The required interval is (0.057, 0.123)

Thus, the bad apples in the consignment lie between 5.7% and 12.3%

Example 10

Out of 1000 TV viewers, 320 watched a particular programme. Find 95% confidence limits for TV viewers who watched this programme.

Solution :

Sample size $n = 1000$

Sample proportion of TV viewers $p = \frac{x}{n} = \frac{320}{1000}$

$= .32$

$\therefore \quad q = 1 - p = .68$

$\text{S.E } (p) = \sqrt{\frac{pq}{n}}$

$= 0.0147$

The 95% confidence limits for population propotion P are given by

$p \pm (1.96) \text{ S.E } (p) = 0.32 \pm 0.028$

$\Rightarrow \quad 0.292$ and 0.348

$\therefore$ TV viewers of this programme lie between 29.2% and 34.8%

Example 11

Out of 1500 school students, a sample of 150 selected at random to test the accuracy of solving a problem in business mathematics and of them 10 did a mistake. Find the limits within which the number of students who did the problem wrongly in whole universe of 1500 students at 99% confidence level.

Solution :

Population size, N = 1500

Sample size, n = 150

Sample proportion, $p = \frac{10}{150} = 0.07$

$\therefore \quad q = 1-p = 0.93$

Standard error of p, $SE(p) = \sqrt{\frac{pq}{n}} = 0.02$

The 99% confidence limits for population proportion P are given by

$$p \pm (Z_c) \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\Rightarrow \quad 0.07 \pm (2.58)(0.02) \sqrt{\frac{1500-150}{1500-1}}$$

$$\Rightarrow \quad 0.07 \pm 0.048$$

$\therefore$ The confidence interval for P is (0.022 , 0.118)

$\therefore$ The number of students who did the problem wrongly in the population of 1500 lies between .022 x 1500 = 33 and .118 x 1500 = 177.

EXERCISE 9.3

1) A sample of five measurements of the diameter of a sphere were recorded by a scientist as 6.33, 6.37, 6.36, 6.32 and 6.37 mm. Determine the point estimate of (a) mean, (b) variance.

2) Measurements of the weights of a random sample of 200 ball bearings made by a certain machine during one week showed mean of 0.824 newtons and a standard deviation of 0.042 newtons. Find (a) 95% and (b) 99% confidence limits for the mean weight of all the ball bearings.

3) A random sample of 50 branches of State Bank of India out of 200 branches in a district showed a mean annual profit of Rs.75 lakhs and a standard deviation of 10 lakhs. Find the 95% confidence limits for the estimate of mean profit of 200 branches.

4) A random sample of marks in mathematics secured by 50 students out of 200 students showed a mean of 75 and a standard deviation of 10. Find the 95% confidence limits for the estimate of their mean marks.

5) Out of 10000 customer's ledger accounts, a sample of 200 accounts was taken to test the accuracy of posting and balancing wherein 35 mistakes were found. Find 95% confidence limits within which the number of defective cases can be expected to lie.

6) A sample poll of 100 voters chosen at random from all voters in a given district indicated that 55% of them were in favour of a particular candidate. Find (a) 95% confidence limits, (b) 99% confidence limits for the proportion of all voters in favour of this candidate.

9.4 HYPOTHESIS TESTING

There are many problems in which, besides estimating the value of a parameter of the population, we must decide whether a statement concerning a parameter is true or false; that is, we must test a hypothesis about a parameter.

To illustrate the general concepts involved in this kind of decision problems, suppose that a consumer protection agency wants to test a manufacturer's claim that the average life time of electric bulbs produced by him is 200 hours. So it instructs a member of its staff to take 50 electric bulbs from the godown of the company and test them for their lifetime continuously with the intention of rejecting the claim if the mean life time of the bulbs is below 180 hours (say); otherwise it will accept the claim.

Thus **hypothesis** is an assumption that we make about an unknown population parameter. We can collect sample data from the population, arrive at the sample statistic and then test if the hypothesis about the population parameter is true.

9.4.1 Null Hypothesis and Alternative Hypothesis

In hypothesis testing, the statement of the hypothesis or assumed value of the population parameter is always stated before we begin taking the sample for analysis.

A statistical statement about the population parameter assumed before taking the sample for possible rejection on the basis of outcome of sample data is known as a **null hypothesis.**

The null hypothesis asserts that there is no diffeence between the sample statistic and population parameter and whatever difference is there is attributable to sampling error.

A hypothesis is said to be **alternative hypothesis** when it is complementary to the null hypothesis.

The null hypothesis and alternative hypothesis are usually denoted by H_0 and H_1 respectively.

For example, if we want to test the null hypothesis that the average height of soldiers is 173 cms, then

$$H_0 : \mu = 173 = \mu_0 \text{ (say)}$$
$$H_1 : \mu \neq 173 \neq \mu_0.$$

9.4.2 Types of Error

For testing the hypothesis, we take a sample from the population, and on the basis of the sample result obtained, we decide whether to accept or reject the hypothesis.

Here, two types of **errors** are possible. A null hypothesis could be rejected when it is true. This is called **Type I error** and the probability of committing type I error is denoted by α.

Alternatively, an error could result by accepting a null hypothesis when it is false. This is known as **Type II error** and the probability of committing type II error is denoted by β.

This is illustrated in the following table :

Actual	Decision based on sampling	Error and its Probability
H_0 is True	Rejecting H_0	Type I error ; $\alpha = P\{H_1 / H_0\}$
H_0 is False	Accepting H_0	Type II error ; $\beta = P\{H_0 / H_1\}$

9.4.3 Critical region and level of significance

A region in the sample space which amounts to rejection of null hypothesis (H_0) is called the **critical region.**

After formulating the null and alternative hypotheses about a population parameter, we take a sample from the population and calculate the value of the relevant statistic, and compare it with the hypothesised population parameter.

After doing this, we have to decide the **criteria** for accepting or rejecting the null hypothesis. These criteria are given as a range of values in the form of an interval, say (a, b), so that if the statistic value falls outside the range, we reject the null hypothesis.

If the statistic value falls within the interval (a, b), then we accept H_0. This criterion has to be decided on the basis of the **level of significance.** A 5% level of signififance means that 5% of the statistical values arrived at from the samples will fall outside this range (a, b) and 95% of the values will be within the range (a, b).

Thus the level of significance is the probability of Type I error α. The levels of significance usually employed in testing of hypothesis are 5% and 1%.

A high significance level chosen for testing a hypothesis would imply that higher is the probability of rejecting a null hypothesis if it is true.

9.4.4 Test of significance

The tests of significance are (a) Test of significance for **large samples** and (b) Test of significance for **small samples.**

For larger sample size (>30), all the distributions like Binomial, Poission etc., are approximated by normal distribution. Thus normal probability curve can be used for testing of hypothesis.

For the test statistic Z (standard normal variate), the critical region at 5% level is given by $| Z | \geq 1.96$ and hence the acceptance region is $| Z | < 1.96$. Where as the critical region for Z at 1% level is $| Z | \geq 2.58$ and the acceptance region is $| Z | < 2.58$.

The testing hypothesis involves five steps :

(i) The formulation of null hypothesis and an alternative hypothesis

(ii) Set up suitable significance level.

(iii) Setting up the statistical test criteria.

(iv) Setting up rejection region for the null hypothesis.

(v) Conclusion.

Example 12

The mean life time of 50 electric bulbs produced by a manufacturing company is estimated to be 825 hours with a standard deviation of 110 hours. If m is the mean life time of all the bulbs produced by the company, test the hypothesis that m = 900 hours at 5% level of significance.

Solution :

Null Hypothesis, $H_0 : \mu = 900$

Alternative Hypothesis, $H_1 : \mu \neq 900$

Test statistic , Z is the standard normal variate.

under H_0, $Z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ where $\overline{X}$ is the sample mean

σ = s.d.of the population

$$= \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}} \qquad \text{(For large sample, } \sigma = s\text{)}$$

$$= \frac{825 - 900}{\frac{110}{\sqrt{50}}} = -4.82.$$

$\therefore \; |Z| = 4.82$

Significant level, $\alpha = 0.05$ or 5%

Critical region is $|Z| \geq 1.96$

Acceptance region is $|Z| < 1.96$

The calculated Z is much greater than 1.96.

Decision : Since the calculated value of $|Z| = 4.82$ falls in the critical region, the value of Z is significant at 5% level.

$\therefore$ the null hypothesis is rejected.

$\therefore$ we conclude that the mean life time of the population of electric bulbs cannot be taken as 900 hours.

Example 13

A company markets car tyres. Their lives are normally distributed with a mean of 50000 kilometers and standard deviation of 2000 kilometers. A test sample of 64 tyres has a mean life of 51250 kms. Can you conclude that the sample mean differs significantly from the population mean? (Test at 5% level)

Solution :

Sample size, $n = 64$

Sample mean, $\overline{X} = 51250$

H_0 : population mean $\mu = 50000$

H_1 : $\mu \neq 50000$

Under H_0, the test statistic $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

$$Z = \frac{51250 - 50000}{\frac{2000}{\sqrt{64}}} = 5$$

Since the calculated Z is much greater than 1.96, it is highly significant.

$\therefore$ $H_0 : \mu = 50000$ is rejected.

$\therefore$ The sample mean differs significantly from the population mean

Example 14

A sample of 400 students is found to have a mean height of 171.38 cms. Can it reasonably be regarded as a sample from a large population with mean height of 171.17 cms and standard deviation of 3.3 cms. (Test at 5% level)

Solution :

Sample size, $n = 400$

Sample mean , $\overline{X} = 171.38$

Population mean, $\mu = 171.17$

Sample standard deviation = s.

Population standard deviation, $\sigma = 3.3$

Set $H_0 : \mu = 171.38$

The test statistic, $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

$= \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ since the sample is large, $s = \sigma$

$$= \frac{171.38 - 171.17}{\frac{3.3}{\sqrt{400}}}$$

$= 1.273$

Since $| Z | = 1.273 < 1.96$, we accept the null hypothesis at 5% level of signifiance.

Thus the sample of 400 has come from the population with mean height of 171.17 cms.

EXERCISE 9.4

1) The mean I.Q of a sample of 1600 children was 99. Is it likely that this was a random sample from a population with mean I.Q 100 and standard deviation 15 ? (Test at 5% level of significance)

2) The income distribution of the population of a village has a mean of Rs.6000 and a variance of Rs.32400. Could a sample of 64 persons with a mean income of Rs.5950 belong to this population?
(Test at both 5% and 1% levels of significance)

3) The table below gives the total income in thousand rupees per year of 36 persons selected randomly from a particular class of people

Income (thousands Rs.)

6.5	10.5	12.7	13.8	13.2	11.4
5.5	8.0	9.6	9.1	9.0	8.5
4.8	7.3	8.4	8.7	7.3	7.4
5.6	6.8	6.9	6.8	6.1	6.5
4.0	6.4	6.4	8.0	6.6	6.2
4.7	7.4	8.0	8.3	7.6	6.7

On the basis of the sample data, can it be concluded that the mean income of a person in this class of people is Rs. 10,000 per year? (Test at 5% level of significance)

4) To test the conjecture of the management that 60 percent employees favour a new bonus scheme, a sample of 150 employees was drawn and their opinion was taken whether they favoured it or not. Only 55 employees out of 150 favoured the new bonus scheme Test the conjecture at 1% level of significance.

EXERCISE 9.5

Choose the correct answer

1) The standard error of the sample mean is

   (a) Type I error (b) Type II error

   (c) Standard deviation of the sampling distribution of the mean

   (d) Variance of the sampling distribution of the mean

2) If a random sample of size 64 is taken from a population whose standard deviation is equal to 32, then the standard error of the mean is

   (a) 0.5 (b) 2 (c) 4 (d) 32

3) The central limit theorem states that the sampling distribution of the mean will approach normal distribution

   (a) As the size of the population increases

   (b) As the sample size increases and becomes larger

   (c) As the number of samples gets larger

   (d) As the sample size decreases

4) The Z-value that is used to establish a 95% confidence interval for the estimation of a population parameter is

   (a) 1.28 (b) 1.65 (c) 1.96 (d) 2.58

5) Probability of rejecting the null hypothesis when it is true is

   (a) Type I error (b) Type II error

   (c) Sampling error (d) Standard error

6) Which of the following statements is true?

   (a) Point estimate gives a range of values

   (b) Sampling is done only to estimate a statistic

   (c) Sampling is done to estimate the population parameter

   (d) Sampling is not possible for an infinite population

7) The number of ways in which one can select 2 customers out of 10 customers is

   (a) 90 (b) 60 (c) 45 (d) 50

APPLIED STATISTICS 10

10.1 LINEAR PROGRAMMING

Linear programming is the general technique of optimum allocation of limited resources such as labour, material, machine, capital etc., to several competing activities such as products, services, jobs, projects, etc., on the basis of given criterion of optimality. The term limited here is used to describe the availability of scarce resources during planning period. The criterion of optimality generally means either performance, return on investment, utility, time, distance etc., The word **linear** stands for the proportional relationship of two or more variables in a model. **Programming** means 'planning' and refers to the process of determining a particular plan of action from amongst several alternatives. It is an extremely useful technique in the decision making process of the management.

10.1.1 Structure of Linear Programming Problem (LPP)

The LP model includes the following three basic elements.

(i) Decision variables that we seek to determine.

(ii) Objective (goal) that we aim to optimize (maximize or minimize)

(iii) Constraints that we need to satisfy.

10.1.2 Formulation of the Linear Programming Problem

The procedure for mathematical formulation of a linear programming consists of the following major steps.

Step 1 : Study the given situation to find the **key decision** to be made

Step 2 : Identify the **variables** involved and designate them by symbols x_j(j = 1, 2 ...)

Step 3 : Express the **feasible alternatives** mathematically in terms of variables, which generally are : $x_j \geq 0$ for all j

Step 4 : Identify the **constraints** in the problem and express them as linear inequalities or equations involving the decision variables.

Step 5 : Identify the **objective function** and express it as a linear function of the decision variables.

10.1.3 Applications of Linear programming

Linear programming is used in many areas. Some of them are

(i) *Transport :* It is used to prepare the distribution plan between source production and destination.

(ii) *Assignment :* Allocation of the tasks to the persons available so as to get the maximum efficiency.

(iii) *Marketing :* To find the shortest route for a salesman who has to visit different locations so as to minimize the total cost.

(iv) *Investment :* Allocation of capital to differerent activities so as to maximize the return and minimize the risk.

(v) *Agriculture :* The allotment of land to different groups so as to maximize the output.

10.1.4 Some useful Definitions

A **feasible solution** is a solution which satisfies all the constraints (including non-negativity) of the problem.

A region which contains all feasible solutions is known as **feasible region**.

A feasible solution which optimizes (maximizes or minimizes) the objective function,is called **optimal solution** to the problem.

Note

Optimal solution need not be unique.

Example 1

A furniture manufacturing company plans to make two products, chairs and tables from its available resources, which consists of 400 board feet of mahogany

timber and 450 man-hours of labour. It knows that to make a chair requires 5 board feet and 10 man-hours and yields a profit of Rs.45, while each table uses 20 board feet and 15 man - hours and has a profit of Rs.80. How many chairs and tables should the company make to get the maximum profit under the above resource constraints? Formulate the above as an LPP.

Solution :

Mathematical Formulation :

The data of the problem is summarised below:

Products	Raw material (per unit)	Labour (per unit)	Profit (per unit)
Chair	5	10	Rs. 45
Table	20	15	Rs. 80
Total availability	400	450	

Step 1 : The key decision to be made is to determine the number of units of chairs and tables to be produced by the company.

Step 2 : Let x_1 designate the number of chairs and x_2 designate the number of tables, which the company decides to produce.

Step 3 : Since it is not possible to produce negative quantities, feasible alternatives are set of values of x_1 and x_2, such that $x_1 \geq 0$ and $x_2 \geq 0$

Step 4 : The constraints are the limited availability of raw material and labour. One unit of chair requires 5 board feet of timber and one unit of table requires 20 board feet of timber. Since x_1 and x_2 are the quantities of chairs and tables, the total requirement of raw material will be $5x_1 + 20x_2$, which should not exceed the available raw material of 400 board feet timber. So, the raw material constraint becomes,

$$5x_1 + 20x_2 \leq 400$$

Similarly, the labour constraint becomes,

$$10x_1 + 15x_2 \le 450$$

Step 5 : The objective is to maximize the total profit that the company gets out of selling their products, namely chairs, tables. This is given by the linear function.

$$z = 45x_1 + 80x_2.$$

The linear programming problem can thus be put in the following mathematical form.

$$\text{maximize } z = 45x_1 + 80x_2$$

$$\text{subject to} \quad 5x_1 + 20x_2 \le 400$$

$$10x_1 + 15x_2 \le 450$$

$$x_1 \ge 0, \ x_2 \ge 0$$

Example 2

A firm manufactures headache pills in two sizes A and B. Size A contains 2 mgs. of aspirin. 5 mgs. of bicarbonate and 1 mg. of codeine. Size B contains 1 mg. of aspirin, 8 mgs. of bicarbonate and 6 mgs. of codeine. It is found by users that it requires atleast 12 mgs. of aspirin, 74 mgs. of bicarbonate and 24 mgs. of codeine for providing immediate relief. It is required to determine the least number of pills a patient should take to get immediate relief. Formulate the problem as a standard LPP.

Solution :

The data can be summarised as follows :

Head ache pills	per pill Aspirin	Bicarbonate	Codeine
Size A	2	5	1
Size B	1	8	6
Minimum requirement	12	74	24

Decision variables :

x_1 = number of pills in size A

x_2 = number of pills in size B

Following the steps as given in (10.1.2) the linear programming problem can be put in the following mathematical format :

$$
\begin{aligned}
\text{Maximize} \quad & z = x_1 + x_2 \\
\text{subject to} \quad & 2x_1 + x_2 \geq 12 \\
& 5x_1 + 8x_2 \geq 74 \\
& x_1 + 6x_2 \geq 24 \\
& x_1 \geq 0, \quad x_2 \geq 0
\end{aligned}
$$

10.1.5 Graphical method

Linear programming problem involving two decision variables can be solved by graphical method. The major steps involved in this method are as follows.

Step 1 : State the problem mathematically.

Step 2 : Plot a graph representing all the constraints of the problem and identify the feasible region (solution space). The feasible region is the intersection of all the regions represented by the constraints of the problem and is restricted to the first quadrant only.

Step 3 : Compute the co-ordinates of all the corner points of the feasible region.

Step 4 : Find out the value of the objective function at each corner point determined in step3.

Step 5 : Select the corner point that optimizes (maximzes or minimizes) the value of the objective function. It gives the optimum feasible solution.

Example 3

A company manufactures two products P_1 and P_2. The company has two types of machines A and B for processing

the above products. Product P_1 takes 2 hours on machine A and 4 hours on machine B, whereas product P_2 takes 5 hours on machine A and 2 hours on machine B. The profit realized on sale of one unit of product P_1 is Rs.3 and that of product P_2 is Rs. 4. If machine A and B can operate 24 and16 hours per day respectively, determine the weekly output for each product in order to maximize the profit, through graphical method.

Solution :

The data of the problem is summarized below.

Product	Hours on Machine A	Hours on Machine B	profit (per unit)
P_1	2	4	3
P_2	5	2	4
Max. hours / week	120	80	

Let x_1 be the number of units of P_1 and x_2 be the number of units of P_2 produced. Then the mathematical formulation of the problem is

$$\text{Maximize } z = 3x_1 + 4x_2$$

$$\text{subject to} \quad 2x_1 + 5x_2 \leq 120$$

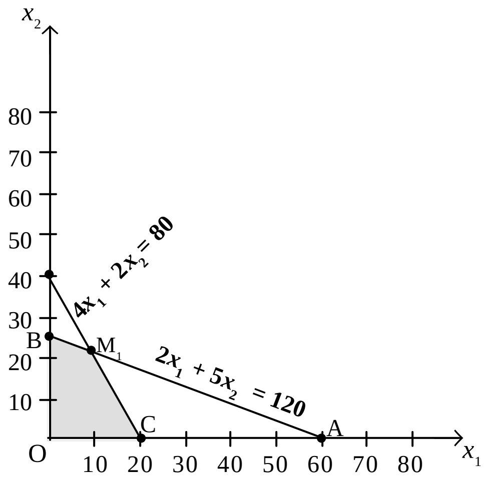
$$4x_1 + 2x_2 \leq 80$$

$$x_1, \; x_2 \geq 0$$

Solution by graphical method

Consider the equation $2x_1 + 5x_2 = 120$, and $4x_1 + 2x_2 = 80$. Clearly (0, 24) and (60, 0) are two points on the line $2x_1 + 5x_2 = 120$. By joining these two points we get the straight line $2x_1 + 5x_2 = 120$. Similarly, by joining the points (20, 0) and (0, 40) we get the sraight line $4x_1 + 2x_2 = 80$. (Fig. 10.1)

Now all the constraints have been represented graphically.



(Fig. 10.1)

The area bounded by all the constrints called feasible region or solution space is as shown in the Fig. 10.1, by the shaded area OCM_1B

The optimum value of objective function occurs at one of the extreme (corner) points of the feasible region. The coordinates of the extreme points are

0 = (0, 0), C = (20, 0), M_1 = (10, 20), B = (0, 24)

We now compute the *z*-values correspoinding to extreme points.

Extreme point	coordinates (x_1, x_2)	$z = 3x_1 + 4x_2$
O	(0, 0)	0
C	(20, 0)	60
M_1	(10, 20)	110
B	(0, 24)	96

The optimum solution is that extreme point for which the objective function has the largest (maximum) value. Thus the optimum solution occurs at the point M_1 i.e. $x_1 = 10$ and $x_2 = 20$.

Hence to maximize profit of Rs.110, the company should produce 10 units of P_1 and 20 units of P_2 per week.

Note

In case of maximization problem, the corner point at which the objective function has a maximum value represent the optimal solution. In case of minimization problem, the corner point at which the objective function has a minimum value represents the optimnal solution.

Example 4

Solve graphically :

Minimize $z = 20x_1 + 40x_2$

Subject to $36x_1 + 6x_2 \geq 108$

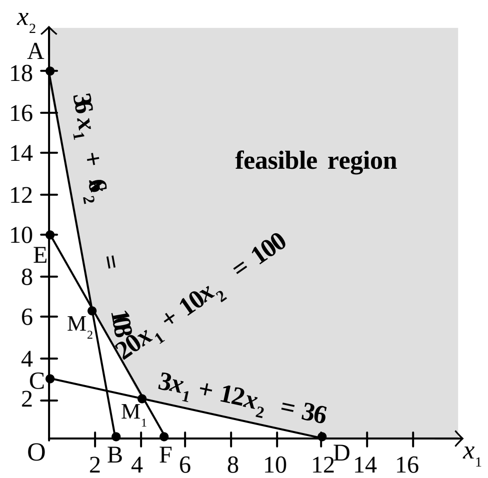
$3x_1 + 12x_2 \geq 36$

$20x_1 + 10x_2 \geq 100$

$x_1, \quad x_2 \geq 0$

Solution :

A(0, 18) and B(3, 0) ; C(0, 3) and D(12, 0) ; E(0, 10) and F(5, 0) are the points on the lines $36x_1 + 6x_2 = 108$, $3x_1 + 12x_2 = 36$ and $20x_1 + 10x_2 = 100$ respectively. Draw thte above lines as Fig. 10.2.



(Fig. 10.2)

Now all the constraints of the given problem have been graphed. The area beyond three lines represents the **feasible region** or **solution space**, as shown in the above figure. Any point from this region would satisfy the constraints.

The coordinates of the extreme points of the feasible region are:

A = (0, 18), M_2 = (2, 6), M_1 = (4, 2), D = (12, 0)

Now we compute the z-values corresponding to extreme points.

Extreme point	coordinates (x_1, x_2)	$z = 20x_1 + 40x_2$
A	(0, 18)	720
M_1	(4, 2)	160
M_2	(2, 6)	280
D	(12, 0)	240

The optimum solution is that extreme point for which the objective function has minimum value. Thus optimum solution occurs at the point M_1 i.e. $x_1 = 4$ and $x_2 = 2$ with the objective function value of $z = 160$ $\therefore$ Minimum $z = 160$ at $x_1 = 4$, $x_2 = 2$

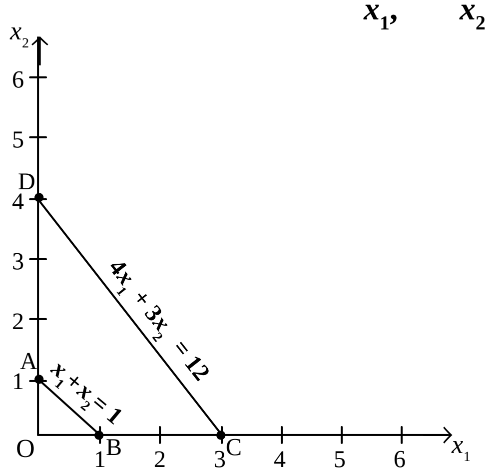
Example 5

Maximize $z = x_1 + x_2$ subject to $x_1 + x_2 \leq 1$

$4x_1 + 3x_2 \geq 12$

$x_1, \quad x_2 \geq 0$

Solution :



(Fig. 10.3)

From the graph, we see that the given problem has no solutiion as the feasible region does not exist.

EXERCISE 10.1

1) A company produces two types of products say type A and B. Profits on the two types of product are Rs.30/- and Rs.40/- per kg. respectively. The data on resources required and availability of resources are given below.

	Requirements		Capacity available per month
	Product A	Product B	
Raw materials (kgs)	60	120	12000
Machining hours / piece	8	5	600
Assembling (man hours)	3	4	500

Formulate this problem as a linear programming problem to maximize the profit.

2) A firm manufactures two products A & B on which the profits earned per unit are Rs.3 and Rs.4 respectively. Each product is processed on two machines M_1 and M_2. Product A requires one minute of processing time on M_1 and two minutes on M_2, while B requires one minute on M_1 and one minute on M_2. Machine M_1 is available for not more than 7 hrs 30 minutes while M_2 is available for 10 hrs during any working day. Formulate this problem as a linear programming problem to maximize the profit.

3) Solve the following, using graphical method

Maximize $z = 45x_1 + 80x_2$

subject to the constraints

$5x_1 + 20x_2 \leq 400$

$10x_1 + 15x_2 \leq 450$

$x_1, x_2 \geq 0$

4) Solve the following, using graphical method

Maximize $z = 3x_1 + 4x_2$

subject to the constraints

$2x_1 + x_2 \leq 40$

$2x_1 + 5x_2 \leq 180$

$x_1, \ x_2 \geq 0$

5) Solve the following, using graphical method

Minimize $z = 3x_1 + 2x_2$

subject to the constraints

$5x_1 + x_2 \geq 10$

$2x_1 + 2x_2 \geq 12$

$x_1 + 4x_2 \geq 12$

$x_1, x_2 \geq 0$

10.2 CORRELATION AND REGRESSION

10.2.1 Meaning of Correlation

The term correlation refers to the degree of relationship between two or more variables. If a change in one variable effects a change in the other variable, the variables are said to be correlated. There are basically three types of correlation, namely positive correlation, negative correlation and zero correlatiion.

Positive correlation

If the values of two variables deviate (change) in the same direction i.e. if the increase (or decrease) in one variable results in a corresponding increase (or decrease) in the other, the correlation between them is said to be positive.

Example

(i) the heights and weights of individuals

(ii) the income and expenditure

(iii) experience and salary.

Negative Correlation

If the values of the two variables constantly deviate (change) in the opposite directions i.e. if the increase (or decrease) in one results in corresponding decrease (or increase) in the other, the correlation between them is said to be negative.

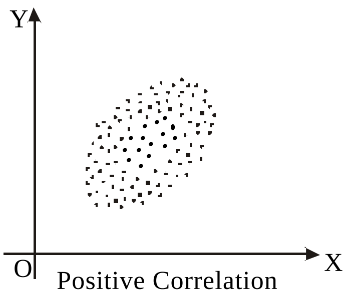
Example

(i) price and demand ,

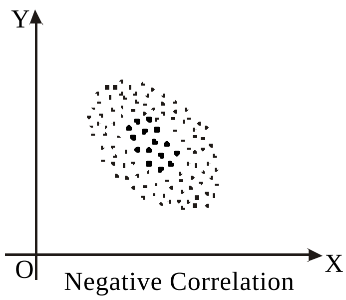
(ii) repayment period and EMI

10.2.2 Scatter Diagram

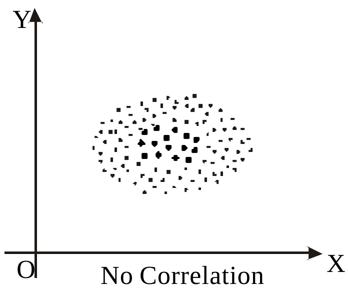
Let (x_1, y_1), (x_2, y_2) ... (x_n, y_n) be the n pairs of observation of the variables x and y. If we plot the values of x along x-axis and the corresponding values of y along y-axis, the diagram so obtained is called a **scatter diagram.** It gives us an idea of relationship between x and y. The types of scatter diagram under simple linear correlation are given below.



Positive Correlation
(Fig. 10.4)



Negative Correlation
(Fig. 10.5)



No Correlation
(Fig. 10.6)

(i) If the plotted points show an upward trend, the correlation will be positive (Fig. 10.4).

(ii) If the plotted points show a downward trend, the correlation will be negative (Fig. 10.5).

(iii) If the plotted point show no trend the variables are said to be uncorrelated (Fig. 10.6).

10.2.3 Co-efficient of Correlation

Karl pearson (1867-1936) a British Biometrician, developed the coefficient of correlation to express the degree of linear relationship between two variables. Correlation co-efficient between two random variables X and Y denoted by $r(\mathrm{X}, \mathrm{Y})$, is given by

$$r(\mathrm{X}, \mathrm{Y}) = \frac{\mathrm{Cov}(\mathrm{X}, \mathrm{Y})}{\mathrm{SD}(\mathrm{X})\ \mathrm{SD}(\mathrm{Y})}$$

where

$$\mathrm{Cov}\ (\mathrm{X}, \mathrm{Y}) = \frac{1}{n}\sum_i (\mathrm{X}_i - \overline{\mathrm{X}})(\mathrm{Y}_i - \overline{\mathrm{Y}}) \quad \text{(covariance between X and Y)}$$

$$\mathrm{SD}\ (\mathrm{X}) = \sigma_x = \sqrt{\frac{1}{n}\sum_i (\mathrm{X}_i - \overline{\mathrm{X}})^2} \quad \text{(standard deviation of X)}$$

$$\mathrm{SD}\ (\mathrm{Y}) = \sigma_y = \sqrt{\frac{1}{n}\sum_i (\mathrm{Y}_i - \overline{\mathrm{Y}})^2} \quad \text{(standard deviationof Y)}$$

Hence the formula to compute Karl Pearson correlation co-efficient is

$$r(\mathrm{X}, \mathrm{Y}) = \frac{\frac{1}{n}\sum_i (\mathrm{X}_i - \overline{\mathrm{X}})(\mathrm{Y}_i - \overline{\mathrm{Y}})}{\sqrt{\frac{1}{n}\sum_i (\mathrm{X}_i - \overline{\mathrm{X}})^2}\sqrt{\frac{1}{n}\sum_i (\mathrm{Y}_i - \overline{\mathrm{Y}})^2}}$$

$$= \frac{\sum_i (\mathrm{X}_i - \overline{\mathrm{X}})(\mathrm{Y}_i - \overline{\mathrm{Y}})}{\sqrt{\sum_i (\mathrm{X}_i - \overline{\mathrm{X}})^2}\sqrt{\sum_i (\mathrm{Y}_i - \overline{\mathrm{Y}})^2}} = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$

Note

The following formula may also be used to compute correlation co-efficient between the two variables X and Y.

(i) $r(\text{X, Y}) = \dfrac{\text{N}\Sigma\text{XY - }\Sigma\text{X}\,\Sigma\text{Y}}{\sqrt{\text{N}\Sigma\text{X}^2 - (\Sigma\text{X})^2}\sqrt{N\Sigma\text{Y}^2 - (\Sigma\text{Y})^2}}$

(ii) $r(\text{X, Y}) = \dfrac{N\Sigma dxdy - \Sigma dx\,\Sigma dy}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2}\sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$

where $dx = x - \text{A}$; $dy = y - \text{B}$ are the deviations from arbitrary values A and B.

10.2.4 Limits for Correlation co-efficient

Correlation co-efficient lies between -1 and +1.
i.e. $-1 \leq r\,(x, y) \leq 1$.

(i) If $r(\text{X , Y}) = +1$ the variables X and Y are said to be perfectly possitively correlated.

(ii) If $r(\text{X , Y}) = -1$ the variables X and Y are said to be perfectly negatively correlated.

(iii) If $r(\text{X , Y}) = 0$ the variables X and Y are said to be uncorrelated.

Example 6

Calculate the correlation co-efficient for the following heights (in inches) of fathers(X) and their sons(Y).

X	**:**	**65**	**66**	**67**	**67**	**68**	**69**	**70**	**72**
Y	**:**	**67**	**68**	**65**	**68**	**72**	**72**	**69**	**71**

Solution :

$$\overline{\text{X}} = \frac{\Sigma\text{X}}{\text{n}} = \frac{544}{8} = 68$$

$$\overline{\text{Y}} = \frac{\Sigma\text{Y}}{\text{n}} = \frac{552}{8} = 69$$

X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	y^2	xy
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	65	-1	-4	1	16	4
67	68	-1	-1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
544	**552**	**0**	**0**	**36**	**44**	**24**

Karl Pearson Correlation Co-efficient,

$$r(x, y) = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} = \frac{24}{\sqrt{36}\sqrt{44}} = .603$$

Since $r(x, y) = .603$, the variables X and Y are positively correlated. i.e. heights of fathers and their respective sons are said to be positively correlated.

Example 7

Calculate the correlation co-efficient from the data below:

X :	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Y :	**9**	**8**	**10**	**12**	**11**	**13**	**14**	**16**	**15**

Solution :

X	Y	X^2	Y^2	XY
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
8	16	64	256	128
9	15	81	225	135
45	**108**	**285**	**1356**	**597**

$$r\,(X,Y) = \frac{N\Sigma XY - \Sigma X\,\Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{9(597) - (45)\,(108)}{\sqrt{9(285) - (45)^2}\sqrt{9(1356) - (108)^2}} = .95$$

∴ X and Y are highly positively correlated.

Example 8

Calculate the correlation co-efficient for the ages of husbands (X) and their wives (Y)

X :	**23**	**27**	**28**	**29**	**30**	**31**	**33**	**35**	**36**	**39**
Y :	**18**	**22**	**23**	**24**	**25**	**26**	**28**	**29**	**30**	**32**

Solution :

Let A = 30 and B = 26 then dx = X– A dy = Y–B

X	Y	d_x	d_y	d^2_x	d^2_y	d_xd_y
23	18	-7	-8	49	64	56
27	22	-3	-4	9	16	12
28	23	-2	-3	4	9	6
29	24	-1	-2	1	4	2
30	25	0	-1	0	1	0
31	26	1	0	1	0	0
33	28	3	2	9	4	6
35	29	5	3	25	9	15
36	30	6	4	36	16	24
39	32	9	6	81	36	54
		11	**-3**	**215**	**159**	**175**

$$r\,(x,y) = \frac{N\Sigma dxdy - \Sigma dx\,\Sigma dy}{\sqrt{N\Sigma d_x^2 - (\Sigma dx)^2}\sqrt{N\Sigma d_y^2 - (\Sigma dy)^2}}$$

$$= \frac{10(175)-(11)(-3)}{\sqrt{10(215)-(11)^2}\sqrt{10(159)-(-3)^2}}$$

$$= \frac{1783}{1790.8} = 0.99$$

$\therefore$ X and Y are highly positively correlated. i.e. the ages of husbands and their wives have a high degree of correlation.

Example 9

Calculate the correlation co-efficient from the following data

N = 25, SX = 125, SY = 100

SX^2 = 650 SY^2 = 436, SXY = 520

Solution :

We know,

$$r = \frac{N\Sigma XY - \text{ÓXÓY}}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

$$= \frac{25(520) - (125)(100)}{\sqrt{25(650)-(125)^2}\sqrt{25(436)-(100)^2}}$$

$$r = -0.667$$

10.2.5 Regression

Sir Francis Galton (1822 - 1911), a British biometrician, defined **regression** in the context of heriditary characteristics. The literal meaning of the word "regression" is **"Stepping back towards the average".**

Regression is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data.

There are two types of variables considered in regression analysis, namely dependent variable and independent variable(s).

10.2.6 Dependent Variable

The variable whose value is to be predicted for a given independent variable(s) is called dependent variable, denoted by Y. For example, if advertising (X) and sales (Y) are correlated, we could estimate the expected sales (Y) for given advertising expenditure (X). So in this case Y is a dependent variable.

10.2.7 Independent Variable

The variable which is used for prediction is called an independent variable. For example, it is possible to estimate the required amount of expenditure (X) for attaining a given amount of sales (Y), when X and Y are correlated. So in this case Y is independent variable. There can be more than one independent variable in regression.

The line of regression is the line which gives the best estimate to the value of one variable for any specific value of the other variable.

Thus the line of Regression is the **line of best fit** and is obtained by the **principle of least squares**. (Refer pages 61 & 62 of Chapter 7). The equation corresponds to the line of regression is also referred to as regression equation.

10.2.8 Two Regression Lines

For the pair of values of (X, Y), where X is an independent variable and Y is the dependent variable the line of regression of Y on X is given by

$$Y - \overline{Y} = b_{yx}(X - \overline{X})$$

where b_{yx} is the regression co-efficient of Y on X and given by $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$, where r is the correlation co-efficient between X and Y and σ_x and σ_y are the standard deviations of X and Y respectively.

$$\therefore \quad b_{yx} = \frac{\Sigma xy}{\Sigma x^2} \quad \text{where } x = X - \overline{X} \text{ and } y = Y - \overline{Y}$$

Similarly when Y is treated as an independent variable and X as dependent variable, the line of regression of X on Y is given by

$$(X - \overline{X}) = b_{xy} (Y - \overline{Y})$$

where $b_{xy} == r \frac{\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma y^2}$ here $x = X - \overline{X}$; $y = Y - \overline{Y}$

Note

The two regression equations are not reversible or interchangeable because of the simple reason that the basis and assumption for deriving these equations are quite different.

Example 10

Calculate the regression equation of X on Y from the following data.

X	**:**	**10**	**12**	**13**	**12**	**16**	**15**
Y	**:**	**40**	**38**	**43**	**45**	**37**	**43**

Solution :

X	Y	$x = X - \overline{X}$	$y = Y - \overline{Y}$	x^2	y^2	xy
10	40	-3	-1	9	1	3
12	38	-1	-3	1	9	3
13	43	0	2	0	4	0
12	45	-1	4	1	16	-4
16	37	3	-4	9	16	-12
15	43	2	2	4	4	4
78	**246**	**0**	**0**	**24**	**50**	**6**

$$\overline{X} = \frac{\Sigma X}{n} = \frac{78}{6} = 13 \qquad \overline{Y} = \frac{\Sigma Y}{n} = \frac{246}{6} = 41$$

$$b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{-6}{50} = -0.12$$

$\therefore$ Regression equation of X on Y is $(X - \overline{X}) = b_{xy} (Y - \overline{Y})$

$$X - 13 = -0.12\ (Y - 41) \Rightarrow \quad X = 17.92 - 0.12Y$$

Example 11

Marks obtained by 10 students in Economics and Statistics are given below.

Marks in Economics : 25 28 35 32 31 36 29 38 34 32

Marks in Statistics : 43 46 49 41 36 32 31 30 33 39

Find (i) the regression equation of Y on X

(ii) estimate the marks in statistics when the marks in Economics is 30.

Solution :

Let the marks in Economics be denoted by X and statistics by Y.

X	Y	$x = X - \overline{X}$	$y = Y - \overline{Y}$	x^2	y^2	xy
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
320	**380**	**0**	**0**	**140**	**398**	**-93**

$$\overline{X} = \frac{\Sigma X}{n} = \frac{320}{10} = 32 \qquad \overline{Y} = \frac{\Sigma Y}{n} = \frac{380}{10} = 38$$

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{-93}{140} = -0.664$$

(i) Regression equation of Y on X is

$$Y - \overline{Y} = b_{yx}(X - \overline{X})$$

$$Y - 38 = -0.664\,(X - 32)$$

$$\Rightarrow \quad Y = 59.25 - 0.664X$$

(ii) To estimate the marks in statistics (Y) for a given marks in the Economics (X), put X = 30, in the above equation we get,

$$Y = 59.25 - 0.664(30)$$
$$= 59.25 - 19.92 = 39.33 \text{ or } 39$$

Example 12

Obtain the two regression equations for the following data.

X	**:**	**4**	**5**	**6**	**8**	**11**
Y	**:**	**12**	**10**	**8**	**7**	**5**

Solution :

The above values are small in magnitude. So the following formula may be used to compute the regression co-efficient.

$$b_{xy} = \frac{N\Sigma XY - \text{Ó}X\text{Ó}Y}{N\Sigma Y^2 - (\Sigma Y)^2}$$

$$b_{yx} = \frac{N\Sigma XY - \text{Ó}X\text{Ó}Y}{N\Sigma X^2 - (\Sigma X)^2}$$

X	Y	X^2	Y^2	XY
4	12	16	144	48
5	10	25	100	50
6	8	36	64	48
8	7	64	49	56
11	5	121	25	55
34	**42**	**262**	**382**	**-257**

$$\overline{X} = \frac{\Sigma X}{n} = \frac{34}{5} = 6.8 \qquad \overline{Y} = \frac{\Sigma Y}{n} = \frac{42}{5} = 8.4$$

$$b_{XY} = \frac{5(257) - (34)(42)}{5(382) - (42)^2} = -0.98$$

$$b_{YX} = \frac{5(257) - (34)(42)}{5(262) - (34)^2} = -0.93$$

Regression Equation of X on Y is

$$(X - \overline{X}) = b_{XY}(Y - \overline{Y})$$

$\Rightarrow \qquad X - 6.8 = -0.98(Y - 8.4)$

$X = 15.03 - 0.98Y$

Regression equation of Y on X is

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

$Y - 8.4 = -0.93\,(X - 6.8)$

$\Rightarrow \qquad Y = 14.72 - 0.93X$

EXERCISE 10.2

1) Calculate the correlation co-efficient from the following data.

X	:	12	9	8	10	11	13	7
Y	:	14	8	6	9	11	12	3

2) Find the co-efficient of correlation for the data given below.

X	:	10	12	18	24	23	27
Y	:	13	18	12	25	30	10

3) From the data given below, find the correlation co-efficient.

X	:	46	54	56	56	58	60	62
Y	:	36	40	44	54	42	58	54

4) For the data on price (in rupees) and demand (in tonnes) for a commodity, calculate the co-efficient of correlation.

Price (X) :	22	24	26	28	30	32	34	36	38	40
Demand(Y) :	60	58	58	50	48	48	48	42	36	32

5) From the following data, compute the correlation co-efficient.
$N = 11, \quad \Sigma X = 117, \quad \Sigma Y = 260, \quad \Sigma X^2 = 1313$
$\Sigma Y^2 = 6580, \quad \Sigma XY = 2827$

6) Obtain the two regression lines from the following

X	:	6	2	10	4	8
Y	:	9	11	5	8	7

7) With the help of the regression equation for the data given below calculate the value of X when Y = 20.

X	:	10	12	13	17	18
Y	:	5	6	7	9	13

8) Price indices of cotton (X) and wool (Y) are given below for the 12 months of a year. Obtain the equations of lines of regression between the indices.

X :	78	77	85	88	87	82	81	77	76	83	97	93
Y :	84	82	82	85	89	90	88	92	83	89	98	99

9) Find the two regression equations for the data given below.

X :	40	38	35	42	30
Y :	30	35	40	36	29

10.3 TIME SERIES ANALYSIS

Statistical data which relate to successive intervals or points of time, are referred to as "time series".

The following are few examples of time series.

(i) Quarterly production, Half-yearly production, and yearly production for particular commodity.

(ii) Amount of rainfall over 10 years period.

(iii) Price of a commodity at different points of time.

There is a strong notion that the term " time series" usually refer only to Economical data. But it equally applies to data arising in other natural and social sciences. Here the time sequence plays a vital role and it requires special techniques for its analysis. In analysis of time series, we analyse the past in order to understand the future better.

10.3.1 Uses of analysis of Time Series

(i) It helps to study the past conditions, assess the present achievements and to plan for the future.

(ii) It gives reliable forcasts.

(iii) It provides the facility for comparison.

Thus wherever time related data is given in Economics, Business, Research and Planning, the analysis of time series provides the opportunity to study them in proper perspective.

10.3.2 Components of Time Series

A graphical representation of a Time Series data, generally shows the changes (variations) over time. These changes are known as principal components of Time Series . They are

(i) Secular trend (ii) Seasonal variation

(iii) Cyclical variation (iv) Irregular variation.

Secular trend (or Trend)

It means the smooth, regular, long-term movement of a series if observed long enough. It is an upward or downward trend. It may increase or decrease over period of time. For example, time series relating to population, price, production, literacy,etc. may show increasing trend and time series relating to birth rate, death rate, poverty may show decreasing trend.

Seasonal Variation

It is a short-term variation. It means a periodic movement in a time series where the period is not longer than one year. A periodic movement in a time series is one which recurs or repeats at regular intervals of time or periods. Following are the examples of seasonal variation.

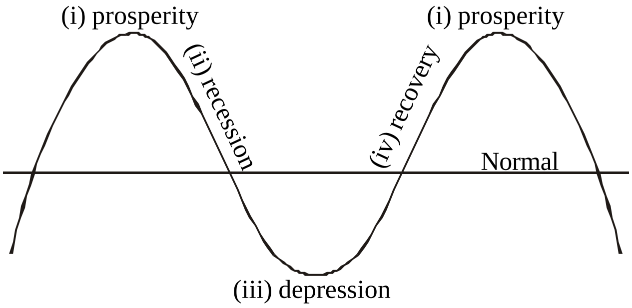
(i) passenger traffic during 24 hours of a day

(ii) sales in a departmental stores during the seven days of a week.

The factors which cause this type of variation are due to climatic changes of the different season, customs and habits of the people. For example more amount of ice creams will be sold in summer and more number of umbrellas will be sold during rainy seasons.

Cyclical Variation

It is also a short-term variation. It means the oscillatory movement in a time series, the period of oscillation being more than a year. One complete period is called a cycle. Business cycle is the suitable example for cyclical variation. There are many time series relating to Economics and Business, which have certain wave-like movements called business cycle. The four phases in business cycle, (i) prosperity (ii) recession (iii) depression (iv) recovery, recur one after another regularly.



(Fig. 10.7)

Irregular Variation

This type of variation does not follow any regularity. These variations are either totally unaccountable or caused by unforeseen events such as wars, floods, fire, strikes etc. Irregular variation is also called as **Erratic Variation.**

10.3.3 Models

In a given time series, some or all the four components, namely secular trend, seasonal variation, cyclical variation and irregular variation may be present. It is important to separate the different components of times series because either our interest may be on a particular component or we may want to study the series after eliminating the effect of a particular component. Though there exist many models, here we consider only two models.

Multiplicative Model

According to this model, it is assumed that there is a multiplicative relationship among the four components. i.e.,

$$y_t = T_t \times S_t \times C_t \times I_t,$$

Where y_t is the value of the variable at time *t*, or observed data at time *t*, T_t is the Secular trend or trend, S_t is the Seasonal variation, C_t is the Cyclical variation and I is the Irregular variation or Erratic variation.

Additive Model

According to this model, it is assumed that y_t be the sum of the four components.

$$y_t = T_t + S_t + C_t + I_t,$$

10.3.4 Measurement of secular trend

The following are the four methods to estimate the secular trend

(i) Graphic method or free - hand method
(ii) Method of Semi - Averages
(iii) Method of Moving Averages
(iv) Method of least squares.

(i) Graphic Method / Free - hand Method

This is the simplest method of studying the trend procedure. Let us take time on the x - axis, and observed data on the y-axis. Mark a point on a graph sheet, corresponding to each pair of time and observed value. After marking all such possible points, draw a straight line which will best fit to the data according to personal judgement.

It is to be noted that the line should be so drawn that it passes between the plotted points in such a manner that the fluctuations in one direction are approximately equal to those in other directions.

When a trend line is fitted by the free hand method an attention should be paid to conform the following conditions.

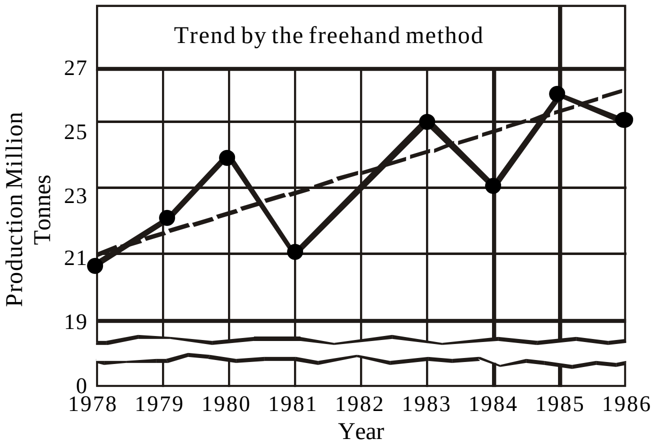
(i) The number of points above the line is equal to the number of points below the line, as far as possible.
(ii) The sum of the vertical deviations from the trend of the annual observation above the trend should equal the sum of the vertical deviations from the trend of the observations below the trend.
(iii) The sum of the squares of the vertical deviations of the observations from the trend should be as small as possible.

Example 13

Fit a trend line to the following data by the free hand mehtod.

year **production of steel**	**1978** **20**	**1979** **22**	**1980** **24**	**1981** **21**	**1982** **23**
year **production of steel**	**1983** **25**	**1984** **23**	**1985** **26**	**1986** **25**	

Solution :



(Fig. 10.8)

Note

(i) The trend line drawn by the free hand method can be extended to predict future values. However, since the free hand curve fitting is too subjective, this method should not be generally used for predictions

(ii) In the above diagram **false base line** (zig-zag) has been used. Generally we use false base line following objectives.

(a) Variations in the data are clearly shown

(b) A large part of the graph is not wasted or space is saved.

(c) The graph provides a better visual communications.

(ii) Method of Semi Averages

This method involves very simple calculations and it is easy to adopt. When this method is used the given data is divided into two equal parts. For example, if we are given data from 1980 to 1999, i.e., over a period of 20 years, the two equal points will be first 10 years, i.e. from 1980 to 1989, and from 1990 to 1999. In case of odd number of years like 7, 11, 13 etc., two equal parts can be made by omitting the middle year. For example, if the data are given for 7 years from 1980 to 1986, the two equal parts would be from 1980 to 1982 and from 1984 to 1986. The middle year 1983 will be omitted.

After dividing the data in two parts, find the arithmetic mean of each part. Thus we get semi averages from which we calculate the annual increase or decrease in the trend.

Example 14

Find trend values to the following data by the method of semi-averages.

Year	**1980**	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**
Sales	**102**	**105**	**114**	**110**	**108**	**116**	**112**

Solution :

No. of years = 7 (odd no.) By omitting the middle year (1983) we have

Year	Sales	Semi total	Semi-average
1980	102		
1981	105	321	107
1982	114		
1983	110		
1984	108		
1985	116	336	112
1986	112		

Difference between middle periods = 1985 – 1981 = 4

Difference between semi averages = 112 - 107 = 5

Annual increase in trend $= \frac{5}{4} = 1.25$

Year	1980	1981	1982	1983	1984	1985	1986
Trend	105.75	107	108.25	109.50	110.75	112	113.25

Example 15

The sales in tonnes of a commodity varied from 1994 to 2001 as given below:

Year	**1994**	**1995**	**1996**	**1997**
Sales	**270**	**240**	**230**	**230**
Year	**1998**	**1999**	**2000**	**2001**
Sales	**220**	**200**	**210**	**200**

Find the trend values by the method of semi-average. Estimate the sales in 2005.

Solution :

Year	Sales	Semi total	Semi average
1994	270		
1995	240		
1996	230	⟶ 970 ⟶	242.5
1997	230		
1998	220		
1999	200		
2000	210	⟶ 830 ⟶	207.5
2001	200		

Difference between middle periods = 1999.5 - 1995.5 = 4

Decrease in semi averages = 242.5 - 207.5 = 35

Annual decrease in trend $= \frac{35}{4} = 8.75$

Half yearly decrease in trend = 4.375

Year	1994	1995	1996	1997
Sales trend	255.625	246.875	238.125	229.375
Year	1998	1999	2000	2001
Sales trend	220.625	211.875	203.125	194.375

Trend value for the year 2005 = 194.375 – (8.75 x 4) = 159.375

(iii) Method of Moving Averages

This method is simple and flexible algebraic method of measuring trend. The method of Moving Average is a simple device for eliminating fluctuations and obtaining trend values with a fair degree of accuracy. The technique of Moving Average is based on the arithmetic mean but with a distinction. In arithmetic mean we sum all the items and divide the sum by number of items, whereas in Moving Average method there are various averages in one series depending upon the number of years taken in a Moving Average. While applying this method, it is necessary to define a period for Moving Average such as 3 yearly moving average (odd number of years), 4-yearly moving average (even number of years) etc.

Moving Average - Odd number of years (say 3 years)

To find out the trend values by the method of 3-yearly moving averages, the following steps are taken into consideration.

1) Add up the values of the first three years and place the yearly sum against the median year i.e. the 2nd year.
2) Leave the first year item and add up the values of the next three years. i.e. from the 2nd year to 4 th year and place the sum (known as moving total) against the 3rd year.
3) Leave the first two items and add the values of the next three years. i.e. from 3rd year to the 5th year and place the sum (moving total) against the 4th year.
4) This process must be continued till the value of the last item is taken for calculating the moving average.
5) Each 3-yearly moving total must be divided by 3 to get themoving averages. This is our required trend values.

Note

The above 5 steps can be applied to get 5-years, 7-years, 9-years etc., Moving Averages.

Moving Average - Even number of years (say 4 years)

1) Add up the values of the first four years and place the sum against the middle of 2nd and 3rd years.
2) Leave the first year value and add from the 2nd year onwards to the 5th year and write the sum (moving total) against the middle of the 3rd and the 4th items.
3) Leave the first two year values and add the values of the next four years i.e. from the 3rd year to the 6th year. Place the sum (moving total) against the middle of the 4th and the 5th items.
4) This process must be continued till the value of the last item is taken into account.
5) Add the first two 4-years moving total and write the sum against 3rd year.
6) Leave the first 4-year moving total and add the next two 4-year moving total. This sum must be placed against 4th year.
7) This process must be continued till all the four-yearly moving totals are summed up and centred.
8) Divide the 4-years moving total centred by 8 and write the quotient in a new column. These are our required trend values.

Note

The above steps can be applied to get 6-years, 8-years, 10-years etc., Moving Averages.

Example 16

Calculate the 3-yearly Moving Averages of the production figures (in mat. tonnes) given below

Year	1973	1974	1975	1976	1977	1978	1979	1980
Production	**15**	**21**	**30**	**36**	**42**	**46**	**50**	**56**
Year	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**	**1987**	
Production	**63**	**70**	**74**	**82**	**90**	**05**	**102**	

Solution :

Calculation of 3-yearly Moving Averages

Year	Productiion y	3-yearly Moving total	3-yearly Moving average
1973	15	---	---
1974	21	66	22.00
1975	30	87	29.00
1976	36	108	36.00
1977	42	124	41.33
1978	46	138	46.00
1979	50	152	50.67
1980	56	169	56.33
1981	63	189	63.00
1982	70	207	69.00
1983	74	226	75.33
1984	82	246	82.00
1985	90	267	89.00
1986	95	287	95.67
1987	102	---	---

Example 17

Estimate the trend values using the data given below by taking 4-yearly Moving Average.

Year	1974	1975	1976	1977	1978	1979	1980
Value	**12**	**25**	**39**	**54**	**70**	**37**	**105**
Year	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**	**1987**
Value	**100**	**82**	**65**	**49**	**34**	**20**	**7**

Solution :

Year	value	4 year moving total	4 year moving total centered	Two 4-year moving total (Trend values)
1974	12	---	---	---
1975	25		---	---
		→ 130		
1976	39		318	39.75
		→ 188		
1977	54		388	48.50
		→ 200		
1978	70		466	58.25
		→ 266		
1979	37		578	72.25
		→ 312		
1980	105		636	79.50
		→ 324		
1981	100		676	84.50
		→ 352		
1982	82		648	81.00
		→ 296		
1983	65		526	65.75
		→ 230		
1984	49		398	49.75
		→ 168		
1985	34		278	34.75
		→ 110		
1986	20		---	---
1987	7		---	---

(iv) Method of Least Squares

The method of least squares is most widely used in practice. The method of least squares may be used to fit a straight line trend.

The straight line trend is generally expressed by an equation

$$y_t = a + bx$$

Where y_t is used to represent the trend values, 'a' is the intercept, 'b' represents slope of the line which is also known as the ratio of growth during a unit of time. The variable x represents the time periods.

In order to determine the values of the unknown constants '*a*' and '*b*' the following equations, known as normal equations, are used.

$$\Sigma y = na + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2,$$

where n represents number of observations (years, months or any other period) for which the data are given.

For derivation of normal equations, refer pages 61 & 62 of Chapter 7. To solve the above normal equations and get trend values the following are the computational steps.

Case (i) When the number of years is odd

1) Denote the years as the X variable and its corresponding values as Y.

2) Assume the middle year as the period of origin and take deviations accordingly. Thus $\Sigma X = 0$.

3) Find ΣX^2, ΣY^2 and ΣXY.

4) Substitute ΣX, ΣX^2, ΣY and ΣXY in the above normal equation and the solve it.

   Hence $a = \dfrac{\Sigma Y}{n}$ $b = \dfrac{\Sigma XY}{\Sigma X^2}$

5) Put the values of '*a*' and '*b*' in the equation and solving for each value of X, we get the trend values.

Case (ii) When the number of years is even

In this case, assume the variable X as

$$X = \frac{x - \text{A.M. of two Middle years}}{.5}$$

and all other steps are similar to case (i)

Note

If the time lag between consecutive years is not one assume the variable X as

$$X = \frac{x - \text{A.M. of two Middle years}}{d}$$

$$\text{where } d = \frac{\text{Difference between two consecutive years}}{2}$$

Example 18

Below are the given the figures of production (in thousand tonnes) of a sugar factory.

Year	1980	1981	1982	1983	1984	1985	1986
Production	**80**	**90**	**92**	**83**	**94**	**99**	**92**

Fit a straight line trend to these figures by the method of least squares and estimate the production in the year of 1990.

Solution :

Given the numbers of years $n = 7$ (odd)

Year x	Production ('000 tonnes) $y = Y$	$X = x - 1983$	X^2	XY
1980	80	-3	9	-240
1981	90	-2	4	-180
1982	92	-1	1	-92
1983	83	0	0	0
1984	94	1	1	94
1985	99	2	4	198
1986	92	3	9	276
	630	**0**	**28**	**56**

The equation of the straight line trend is $Y_t = a + bX$

Substituting the values of ΣX, ΣX^2, ΣXY and n in normal equation, we get

$630 = 7a + b(0) \quad \Rightarrow \; a = 90$

$56 = a(0) + b(28) \quad \Rightarrow \; b = 2$

Hence the equation of the straight line trend is

$$Y_t = 90 + 2X$$

Trend values

For X = -3, $Y_t = 90 + 2(-3) = 84$

For X = -2, $Y_t = 90 + 2(-2) = 86$

For X = -2, $Y_t = 90 + 2(-1) = 88$

For X = 0, $Y_t = 90 + 2(0) = 90$

For X = 1, $Y_t = 90 + 2(1) = 92$

For X = 2, $Y_t = 90 + 2(2) = 94$

For X = 3, $Y_t = 90 + 2(3) = 96$

To estimate the production in 1990, substitute X = 7 in the trend equation.

$\therefore \; Y_{1990} = 90 + 2(7) = 104 \times 1000$ tonnes

Example 19

Fit a straight line trend by the method of least squares to the following data. Also predict the earnings for the year 1988.

Year	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**
Earnings	**38**	**40**	**65**	**72**	**69**	**60**	**87**	**95**

Solution :

Number of years $n = 8$ (even)

year x	Earnings (in lakhs) y = Y	$X = \frac{x-1982.5}{.5}$	X^2	XY
1979	38	-7	49	-266
1980	40	-5	25	-200
1981	65	-3	9	-195
1982	72	-1	1	-72
1983	69	1	1	69
1984	60	3	9	180
1985	87	5	25	435
1986	95	7	49	665
	526	**0**	**168**	**616**

The equation of the straight line trend is $Y_t = a + bX$

Substituting the values of ΣX, ΣX^2, ΣXY, n in Normal equation, we get,

$526 = 8a + b(0)$ -----------(1)
$616 = a(0) + b(168)$ -----------(2)

Solving (1) and (2) we get $a = 65.75$ and $b = 3.667$

Hence the equation of the straight line trend is

$$Y_7 = 65.75 + 3.667X$$

Year	1979	1980	1981	1982
Trend values	40.08	47.415	54.749	62.083

Year	1983	1984	1985	1986
Trend values	69.417	76.751	84.085	91.419

To estimate the earnings in 1988, substitute X = 11 in the trend equation and we get,

$Y_t = 65.75 + 3.667(11) = 106.087$

The estimate earnings for the year 1988 are Rs.106.087 lakhs.

Measurement of seasonal variation

Seasonal variation can be measured by the method of simple average.

Method of simple average

This method is the simple method of obtaining a seasonal Index. In this method the following steps are essential to calculate the Index.

(i) Arrange the data by years, month or quarters as the case may be.

(ii) Compute the totals of each month or each quarter.

(iii) Divide each total by the number of years for which the data are given. This gives seasonal averages (monthly or quarterly)

(iv) Compute average of seasonal averages. This is called grand average.

(v) Seasonal Index for every season (monthly or quarterly) is calculated as follows

$$\text{Seasonal Index (S.I)} = \frac{\text{Seasonal Average}}{\text{Grand Average}} \times 100$$

Note

(i) If the data is given monthwise,

$$\text{Seasonal Index} = \frac{\text{Monthly Average}}{\text{Grand Average}} \times 100$$

(ii) If quarterly data is given,

$$\text{Seasonal Index} = \frac{\text{Quarterly Average}}{\text{Grand Average}} \times 100$$

Example 20

From the data given below calculate Seasonal Indices.

	year				
Quarter	**1984**	**1985**	**1986**	**1987**	**1988**
I	**40**	**42**	**41**	**45**	**44**
II	**35**	**37**	**35**	**36**	**38**
III	**38**	**39**	**38**	**36**	**38**
IV	**40**	**38**	**40**	**41**	**42**

Solution :

	Quarters			
year	**I**	**II**	**III**	**IV**
1984	40	35	38	40
1985	42	37	39	38
1986	41	35	38	40
1987	45	36	36	41
1988	44	38	38	42
Total	**212**	**181**	**189**	**201**
Average	**42.4**	**36.2**	**37.8**	**40.2**

Grand Average = $\frac{42.4+36.2+37.8+40.2}{4}$ = 39.15

Seasonal Index (S. I) = $\frac{\text{Quarterly Average}}{\text{Grand Average}}$ x 100

Hence, S.I for I Quarter = $\frac{42.4}{39.15}$ x 100 = 108.30

S.I for II Quarter = $\frac{36.2}{39.15}$ x 100 = 92.54

S.I for III Quarter = $\frac{37.8}{39.15}$ x 100 = 96.55

S.I for IV Quarter = $\frac{40.2}{39.15}$ x 100 = 102.68

Note

Measurement of cyclical variation, and measurement of irregular variation is beyond the scope of this book.

EXERCISE 10.3

1) Draw a trend line by graphic method (free hand)

year	1995	1996	1997	1998	1999	2000	2001
Production	20	22	25	26	25	27	30

2) Draw a trend line by graphic method

year	1997	1998	1999	2000	2001
Production	20	24	25	38	60

3) Obtain the trend values by the method of Semi-Average

year	1987	1988	1989	1990	1991	1992	1993
Production (in tonnes)	90	110	130	150	100	150	200

[redacted]

year	1993	1994	1995	1996	1997	1998	1999	2000
Netprofit (Re lakhs)	38	39	41	43	40	39	35	25

5) Using three year moving averages determine the trend values for the following data.

year	1983	1984	1985	1986	1987	1988	1989
Production (in tonnes)	21	22	23	25	24	22	25

year	1990	1991	1992
Production (in tonnes)	26	27	26

6) Below are given figures of production (in thousand tonnes) of a sugar factory. Obtain the trend values be 3-year moving average.

year	1980	1981	1982	1983	1984	1985	1986
Production	80	90	92	83	94	99	92

7) Using four yearly moving averages calculate the trend values.

year	1981	1982	1983	1984	1985	1986	1987
Production	464	515	518	467	502	540	557

year	1988	1989	1990
Production	571	586	612

8) Calculate the trend values by four year moving averages method.

Year	1978	1979	1980	1981	1982	1983	1984
Production	614	615	652	678	681	655	717

Year	1985	1986	1987	1988
Production	719	708	779	757

9) Given below are the figures of production of a sugar factory.

year	1989	1990	1991	1992	1993	1994	1995
Production (in tonnes)	77	88	94	85	91	98	90

Fit a straight line trend to these figures by the method of least squares and estimate trend values. Also estimate the production for the year 2000.

10) Fit the straight line trend, find the trend values and estimate the net profit in 2002.

Year	1992	1993	1994	1995	1996	1997	1998
Net profit	65	68	59	55	50	52	54

year	1999	2000
Net profit	50	42

11) The following data relate to the profit earned by public limited company from 1984 to 1989.

Year	1984	1985	1986	1987	1988	1989
Profit (Rs in 000)	10	12	15	16	18	19

Fit a straight line trend by the method of least squares to the data and tabulate the trend values.

12) Fit a straight line trend and estimate the trend values.

Year	1992	1993	1994	1995	1996	1997	1998	1999
Net profit	47	53	50	46	41	39	40	36

13) Calculate the seasonal indices by the method of simple average for the following data

year	I quarter	II quarter	III quarter	IV quarter
1985	68	62	61	63
1986	65	58	66	61
1987	68	63	63	67

14) Calculate the seasonal indices for the following data by the method of simple Average.

	Quarters			
year	I	II	III	IV
1994	78	66	84	80
1995	76	74	82	78
1996	72	68	80	70
1997	74	70	84	74
1998	76	74	86	82

15) Calculate the seasonal Indices for the following data using average method.

	Quarters			
year	I	II	III	IV
1982	72	68	80	70
1983	76	70	82	74
1984	74	66	84	80
1985	76	74	84	78
1986	78	74	86	82

10.4 INDEX NUMBERS

"An **Index Number** is a single ratio (usually in percentages) which measures the (combined average) change of several variables between two different times, places and situations" - Alva. M. Tuttle.

Index numbers are the devices for measuring differences in the magnitude of a group of related variables, over two different situations or defined as a measure of the average change in a group of related variables over two different situations. For example, the price of commodities at two different places or two different time periods at the same location. We need Index Numbers to compare the cost of living at different times or in different locations.

10.4.1 Classification of Index Numbers

Index Numbers may be classified in terms of what they measure. They are

(i) Price Index Numbers

(ii) Quantity Index Numbers

(iii) Value Index Numbers

(iv) Special purpose Index Numbers

We shall discuss (i) and (ii).

10.4.2 Uses of Index Numbers

(i) Index numbers are used to evolve business policies.

(ii) Index numbers determine the inflation or deflation in economy.

(iii) Index numbers are used to compare intelligence of students in different locations or for different year.

(iv) Index numbers serve as economic barometers.

10.4.3 Method of construction of Index Numbers

Index Numbers are broadly divided into two groups

(i) Unweighted Index

(ii) Weighted Index

We confine our attention to weighted index numbedrs.

10.4.4 Weighted Index Numbers

The method of construction of weighted indices are

(c) Weighted aggregative method

(d) Weighted averages of relatives method

Weighted Aggregative Index Numbers

Let p_1 and p_0 be the prices of the current year and the base year respctively. Let q_1 and q_0 be the quantities of the current year and the base year respectively. The formulae for assigning weights to the items are :

(i) Laspeyre's Price Index

$P_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$ x 100 where w = $p_0 q_0$ is the weight assigned to the items and P_{01} is the price index.

(ii) Paasche's price index

$$P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

Here the weights assigned to the items are the current year quantities i.e. W = $p_0 q_1$

(iii) Fisher's price Index

$$P_{01}^{F} = \sqrt{P_{01}^{L} \times P_{01}^{P}} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

Note

Fisher's price index is the geometric mean of Laspeyre's and Paasche's price index numbers.

Example 21

Compute (i) Laspeyre's (ii) Paasche's and (iii) Fisher's Index Numbers for the 2000 from the following :

Commodity	**Price**		**Quantity**	
	1990	**2000**	**1990**	**2000**
A	**2**	**4**	**8**	**6**
B	**5**	**6**	**10**	**5**
C	**4**	**5**	**14**	**10**
D	**2**	**2**	**19**	**13**

Solution :

Commodity	Price		Quantity					
	Base year p_0	Current year p_1	Base year q_0	Current year q_1	p_0q_0	p_1q_0	p_0q_1	p_1q_1
A	2	4	8	6	16	32	12	24
B	5	6	10	5	50	60	25	30
C	4	5	14	10	56	70	40	50
D	2	2	19	13	38	38	26	26
					160	**200**	**103**	**130**

(i) Laspeyre's Index : $P_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$= \frac{200}{160} \times 100 = 125$

(ii) Paasche's Index : $P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

$= \frac{130}{103} \times 100 = 126.21$

(iii) Fisher's Index : $P_{01}^{F} = \sqrt{P_{01}^{L} \times P_{01}^{P}}$

$= 125.6$

Example 22

From the following data, calculate price index number by (a) Laspeyre's method (b) Paasche's method (iii) Fisher's method.

Commodity	Base year		Current year	
	Price	Quantity	Price	Quantity
A	**2**	**40**	**6**	**50**
B	**4**	**50**	**8**	**40**
C	**6**	**20**	**9**	**30**
D	**8**	**10**	**6**	**20**
E	**10**	**10**	**5**	**20**

Solution :

Commodity	Base year Price	Qty	Current year Price	Qty				
	p_0	q_0	p_1	q_1	p_0q_0	p_1q_0	p_0q_1	p_1q_1
A	2	40	6	50	80	240	100	300
B	4	50	8	40	200	400	160	320
C	6	20	9	30	120	180	180	270
D	8	10	6	20	80	60	160	120
E	10	10	5	20	100	50	200	100
					580	**930**	**800**	**1110**

(i) Laspeyre's Price Index : $P_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$$= \frac{930}{580} \times 100 = 160.34$$

(ii) Paasche's Price Index : $P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

$$= \frac{1100}{800} \times 100 = 137.50$$

(iii) Fisher's Index : $P_{01}^{F} = \sqrt{P_{01}^{L} \times P_{01}^{P}} = 148.48$

10.4.5 Test of adequacy for Index Number

Index Numbers are constructed to study the relative changes in prices, quantities, etc. of one time in comparison with another. Several formulae have been suggested for constructing index numbers and one should select the most appropriate one in a given situation. Following are the tests suggested for choosing an appropriate index.

1) Time reversal test

2) Factor reversal test

Time reversal test

It is a test to determine whether a given method will work both ways in time, forward and backward. When the data for any two years are treated by the same method, but with the bases reversed, the two index numbers secured should be reciprocals of each other so that their product is unity. Symbolically the following relation should be satisfied.

$$\therefore \; \mathbf{P_{01} \times P_{10} = 1}$$

(ignoring the factor 100 in each index) where P_{01} is the index for current period 1 on base period 0 and P_{10} is the index for the current period 0 on base period 1.

Factor reversal test

This test holds that the product of a price index and the quantity index should be equal to the corresponding value index. The test is that the change in price multiplied by the change in quantity should be equal to the total change in value.

$$\therefore \; \mathbf{P_{01} \times Q_{01}} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \text{ (ignoring the factor 100 in each index)}$$

P_{01} gives the relative change in price and Q_{01} gives the relative change in quantity. The total value of a given commodity in a given year is the product of the quantity and the price per unit.

$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$ is the ratio of the total value in the current period to the total value in the base period and this ratio is called the **true value ratio.**

Fisher's index is known as Ideal Index Number since it is the only index number that satisfies both reversal tests.

Example 23

Calculate Fisher's Ideal Index from the following data and verify that it satisfies both Time Reversal and Factor Reversal test

Commodity	Price		Quantity	
	1985	1986	1985	1986
A	**8**	**20**	**50**	**60**
B	**2**	**6**	**15**	**10**
C	**1**	**2**	**20**	**25**
D	**2**	**5**	**10**	**8**
E	**1**	**5**	**40**	**30**

Solution :

Commodity	1985		1986					
	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	8	50	20	60	1000	400	1200	480
B	2	15	6	10	90	30	60	20
C	1	20	2	25	40	20	50	25
D	2	10	5	8	50	20	40	16
E	1	40	5	30	200	40	150	30
					1380	**510**	**1500**	**571**

Fisher's Ideal Index $= \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$

$= \sqrt{\frac{1380}{510} \times \frac{1500}{571}} \times 100$

$= 2.6661 \times 100 = 266.61$

Time reversal test

Test is satisfied when $P_{01} \times P_{10} = 1$

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} = \sqrt{\frac{1380}{510} \times \frac{1500}{571}}$$

$$P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_1}} = \sqrt{\frac{571}{1500} \times \frac{510}{1380}}$$

$$P_{01} \times P_{10} = \sqrt{\frac{1380}{510} \times \frac{1500}{571} \times \frac{571}{1500} \times \frac{510}{1380}}$$

$$= \sqrt{1} = 1$$

Hence Fisher's Ideal Index satisfies Time reversal test.

Factor reversal test

Test is satisfied when $P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \sqrt{\frac{571}{510} \times \frac{1500}{1380}}$$

$$\therefore \quad P_{01} \times Q_{01} = \sqrt{\frac{1380}{510} \times \frac{1500}{571} \times \frac{571}{510} \times \frac{1500}{1380}}$$

$$= \frac{1500}{510} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Hence Fisher's Ideal Index satisfies Factor reversal test.

Example 24

Compute Index Number using Fisher's formula and show that it satisfies time reversal test and factor reversal test.

Commodity	Base year		Current year	
	Price	Quantity	Price	Quantity
A	10	12	12	15
B	7	15	5	20
C	5	24	9	20
D	16	5	14	5

Solution :

Commodity	Base year		Current year					
	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	10	12	12	15	144	120	180	150
B	7	15	5	20	75	105	100	140
C	5	24	9	20	216	120	180	100
D	16	5	14	5	70	80	70	80
					505	**425**	**530**	**470**

Fisher's Ideal Index = $\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$

$= \sqrt{\frac{505}{425} \times \frac{530}{470}} \times 100 = 115.75$

Time reversal test

Test is satisfied when $P_{01} \times P_{10} = 1$

$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} = \sqrt{\frac{505}{425} \times \frac{530}{470}}$

$P_{10} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = \sqrt{\frac{470}{530} \times \frac{425}{505}}$

$P_{01} \times P_{10} = \sqrt{\frac{505}{425} \times \frac{530}{470} \times \frac{470}{530} \times \frac{425}{505}}$

$= \sqrt{1} = 1$

Hence Fisher's Ideal Index satisfies Time Reversal Test.

Factor reversal test

Test is satisfied when $P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \sqrt{\frac{470}{530} \times \frac{425}{505}}$$

$$P_{01} \times Q_{01} = \sqrt{\frac{505}{425} \times \frac{530}{470} \times \frac{470}{425} \times \frac{530}{505}}$$

$$= \frac{530}{425} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Hence Fisher's Ideal Index satisfies Factor reversal test.

10.4.6 Cost of Living Index (CLI)

Cost of living index numbers are generally designed to represent the average change over time in the prices paid by the ultimate consumer for a specified quantity of goods and services. Cost of living index number is also known as **Consumer price index number**

It is well known that a given change in the level of prices (retail) affects the cost of living of different classes of people in different manners. The general index number fails to reveal this. Therefore it is essential to construct a cost of living index number which helps us in determining the effect of rise and fall in prices on different classes of consumers living in different areas. It is to be noted that the demand for a higher wage is based on the cost of living index. The wages and salaries in most countries are adjusted in accordance with the cost of living index.

10.4.7 Methods of constructing cost of living index

Cost of living index may be constructed by the following methods.

(i) Aggregate expenditure method or weighted aggregative method

(ii) Family budget method

Aggregate expenditure method

In this method, the quantities of commodities consumed by the particular group in the base year are used as the weights. On the basis of these weights, aggregate (total) expenditure in current year and base year are calculated and the percentage of change is worked out .

$$\therefore \quad \text{Cost of Living Index (C.L.I)} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

This method is the most popular method for constructing cost of living index and the method is same as Laspeyre's price index.

Family budget method

In this method, the value weights obtained by multiplying prices by quantities consumed (i.e. $p_0 q_0$) are taken as weights. To get the cost of living index, find the sum of respective products of price relatives and value weights and then divide this sum by the sum of the value weights.

$$\therefore \quad \text{Cost of living Index} = \frac{\Sigma PV}{\Sigma V} \text{ where}$$

$P = \frac{p_1}{p_0} \times 100$ is the price relative and

$V = p_0 q_0$ is the value weight for each item.

This method is same as the Weighted average of price relative method.

10.4.8 Uses of cost of living index number

(i) The cost of living index number is mainly used in wage negotiations and wage contracts.

(ii) It is used to calculate the dearness allowance for the employees.

Example 25

Calculate the cost of living index by aggregate expenditure method

Commodity	Quantity 2000	Price (Rs)	
		2000	2003
A	**100**	**8**	**12.00**
B	**25**	**6**	**7.50**
C	**10**	**5**	**5.25**
D	**20**	**48**	**52.00**
E	**65**	**15**	**16.50**
F	**30**	**19**	**27.00**

Solution :

Commodity	Quantity 2000	Price 2000	Price 2003		
	q_0	p_0	p_1	p_1q_0	p_0q_0
A	100	8	12.00	1200.00	800
B	25	6	7.50	187.50	150
C	10	5	5.25	52.50	50
D	20	48	52.00	1040.00	960
E	65	15	16.50	1072.50	975
F	30	19	27.00	810.00	570
				4362.50	**3505**

$$\text{C. L. I} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{4362.50}{3505} \times 100 = 124.46$$

Example 26

Construct the cost of living Index Number for 2003 on the basis of 2000 from the following data using family Budget method.

Items	**Price**		**Weights**
	2000	**2003**	
Food	**200**	**280**	**30**
Rent	**100**	**200**	**20**
Clothing	**150**	**120**	**20**
Fuel & lighting	**50**	**100**	**10**
Miscellaneous	**100**	**200**	**20**

Solution :

Calculation of CLI by family budget method

Items	p_0	p_1	weights V	$P=\frac{p_1}{p_0}x100$	PV
Food	200	280	30	140	4200
Rent	100	200	20	200	4000
Clothing	150	120	20	80	1600
Fuel & Lighting	50	100	10	200	2000
Misc.	100	200	20	200	4000
			100		**15800**

Cost of living index = $\frac{\Sigma PV}{\Sigma V} = \frac{15800}{100} = 158$

Hence, there is 58% increase in cost of living in 1986 compared to 1980.

EXERCISE 10.4

1) Compute (i) Laspeyre's (ii) Paasche's and (iii) Fisher's index Numbers

Commodity	Price		Quantity	
	Base year	Current year	Base year	Current year
A	6	10	50	50
B	2	2	100	120
C	4	6	60	60
D	10	12	30	25

2) Construct the price index number from the following data by applying
(i) Laspeyre's (ii) Paasche's (iii) Fisher's method

Commodity	1999		1998	
	Price	Quantity	Price	Quantity
A	4	6	2	8
B	6	5	5	10
C	5	10	4	14
D	2	13	2	19

3) Compute (a) Laspyre's (b) Paasche's (c) Fisher's method of index numbers for 1990 from the following :

Commodity	Price		Quantity	
	1980	1990	1980	1990
A	2	4	8	6
B	5	6	10	5
C	4	5	14	10
D	2	2	19	13

4) From the following data calculate the price index number by
(a) Laspeyre's method (b) paasche's method
(c) Fisher's method

Commodity	Base year		Current year	
	Price	Quantity	Price	Quantity
A	5	25	6	30
B	10	5	15	4
C	3	40	2	50
D	6	30	8	35

5) Using the following data, construct Fisher's Ideal index and show that it satisfies Factor Reversal test and Time Reversal test.

Commodity	Price		Quantity	
	Base year	Current year	Base year	Current year
A	6	10	50	56
B	2	2	100	120
C	4	6	60	60
D	10	12	30	24
E	8	12	40	36

6) Calculate Fisher's Ideal Index from the following data and show how it satisfies time reversal test and factor reversal test.

Commodity	Base year (1997)		Current year (1998)	
	Price	Quantity	Price	Quantity
A	10	10	12	8
B	8	12	8	13
C	12	12	15	8
D	20	15	25	10
E	5	8	8	8
F	2	10	4	6

7) Construct cost of living index for 2000 taking 1999 as the base year from the following data using Aggregate Expenditure method.

Commodity	Quantity (kg.) 1999	Price 1999	Price 2000
A	6	5.75	6.00
B	1	5.00	8.00
C	6	6.00	9.00
D	4	8.00	10.00
E	2	2.00	1.80
F	1	20.00	15.00

8) Calculate the cost of living Index Number using Family Budget method

Commodity	A	B	C	D	E	F	G	H
Quantity in Base year (unit)	20	50	50	20	40	50	60	40
Price in Base year (Rs.)	10	30	40	200	25	100	20	150
Price in current year (Rs)	12	35	50	300	50	150	25	180

9) Calculate the cost of living index number using Family Budget method for the following data taking the base year as 1995

Commodity	Weight	Price (per unit)	
		1995	1996
A	40	16.00	20.00
B	25	40.00	60.00
C	5	0.50	0.50
D	20	5.12	6.25
E	10	2.00	1.50

10) From the data given below, construct a cost of living index number by family budget method for 1986 with 1976 as the base year.

Commodity	P	Q	R	S	T	U
Quantity in 1976 Base year	50	25	10	20	30	40
Price per unit in 1976 (Rs.)	10	5	8	7	9	6
Price per unit in 1986 (Rs.)	6	4	3	8	10	12

10.5 STATISTICAL QUALITY CONTROL (SQC)

Every product manufactured is required for a specific purpose. It means that if the product meets the specifications required for its rightful use, it is of good quality and if not, then the quality of the product is considered to be poor.

It is a well known fact that all repetitive process no matter how carefully arranged are not exactly identical and contain some variability. Even in the manufacture of commodities by highly specialised machines it is not unusual to come across differences between various units of production. For example, in the manufacture of corks, bottles etc. eventhough highly efficient machines are used some difference may be noticed in various units. If the differences are not much, it can be ignored and the product can be passed off as within specifications. But if it is beyond certain limits, the article has to be rejected and the cause of such variation has to be investigated.

10.5.1 Causes for variation

The variation occurs due to two types of causes namely (i) Chance causes (ii) Assignable causes

(i) Chance causes

If the variation occurs due to some inherent pattern of variation and no causes can be assigned to it, it is called **chance** or **random variation.** Chance Variation is tolerable, permissible inevitable and does not materially affect the quality of a product.

(ii) Assignable causes

The causes due to faulty process and procedure are known as assignable causes. the variation due to assignable causes is of non-random nature. Chance causes cannot detected. However assignable causes can be detected and corrected.

10.5.2 Role and advantages of SQC

The role of statistical quality control is to collect and analyse relevant data for the purpose of detecting whether the process is under control or not.

The value of quality control lies in the fact that assignable causes in a process can be quickly detected. Infact the variations are often discovered before the product becomes defective.

SQC is a well accepted and widespread process on the basis of which it is possible to understand the principles and techniques by which decisions are made based on variation.

Statistical quality control is only diagonstic. It tells us whether the standard is being maintained or not. The remedial action rests with the technicians who employ techniques for the maintenance of uniform quality in a continuous flow of manufactured products.

The purpose for which SQC are used are two fold namely, (a) **Process control,** (b) **Product control.**

In process control an attempt is made to find out if a particular process is within control or not. Process control helps in studying the future performance.

10.5.3 Process and Product control

The main objective in any production process is to control and maintain quality of the manufactured product so that it conforms to specified quality standards. In otherwords, we want to ensure that the proportion of defective items in the manufactured product is not too large. This is called **process control** and is achieved through the technique of control charts.

On the otherhand, by **product control** we mean controlling the quality of the product by critical examination at strategic points and this is achieved through **product control plans** pioneered by Dodge and Romig. Product control aims at guaranteeing a certain quality level to the consumer regardless of what quality level is being maintained by the producer. In otherwords, it attempts to ensure that the product marketed by the sale department does not contain a large number of defective items.

10.5.4 Control Charts

The statistical tool applied in process control is the **control chart.** Control charts are the devices to describe the patterns of variation. The control charts were developed by the physicist, Walter A. Stewart of Bell Telephone Company in 1924. He suggested that the control chart may serve, first to define the goal or standard for the process that the management might strive to attain. Secondly, it may be used as an instrument to attain that goal and thirdly, it may serve as a means of judging whether the goal is being achieved. Thus, control chart is an instrument to be used in specification, production and inspection and is the core of statistical quality control.

A control chart is essentially a graphic device for presenting data so as to directly reveal the frequency and extent of variations from established standards or goals. Control charts are simple to construct and easy to interpret and they tell the manager at a glance whether or not the process is in control, i.e. within the tolerance limits.

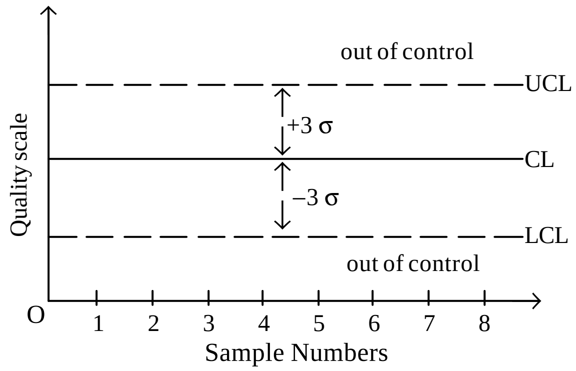
In general a control chart consists of three horizontal lines

(i) A central line to indicate the desired standard or level of the process (CL)

(ii) Upper control limit (UCL) and

(iii) Lower control limit (LCL)

Outline of a control chart



(Fig. 10.9)

From time to time a sample is taken and the data are plotted on the graph. If all the sample points fall within the upper and lower control limits, it is asumed that the process is "in control" and only chance causes are present. When a sample point falls outside the control limits, it is assumed that variations are due to assignable causes.

Types of Control Charts

Broadly speaking, control charts can be divided under two heads.

(i) Control charts of Variables

(ii) Control charts of Attributes

Control charts of variables concern with measurable data on quality characteristics which are usually continous in nature. Such type of data utilises $\overline{X}$ and R chart.

Control charts of attirbutes, namely *c*, *np* and *p* charts concern with the data on quality characteristics, which are not amenable to measurement or attributes (prodcut defective or non defective)

In this chapter, we consider only the control charts of variables, namely $\overline{\mathbf{X}}$ **chart** and **R chart.**

R-Chart (Range chart)

The R chart is used to show the variability or dispersion of the quality produced by a given process. R chart is the companion chart to the $\overline{X}$ chart and both are usually required for adequate analysis of the production process under study. The R chart is generally presented along with the $\overline{X}$ chart. The general procedure for constructing the R chart is similar to that for the $\overline{X}$ chart. The required values for constructing the R chart are :

(i) The range of each sample, R.

(ii) The Mean of the sample ranges, $\overline{R}$

(iii) The control limits are set at

$$U.C.L = D_4\overline{R}$$

$$L.C.L = D_3\overline{R}$$

The values of D_4 and D_3 can be obtained from tables.

$\overline{\mathbf{X}}$ Chart

The $\overline{X}$ chart is used to show the quality averages of the samples drawn from a given process. The following values must first be computed before an $\overline{X}$ chart is constructed.

1) Obtain the mean of each sample $\overline{X}_i$: i = 1, 2 ... *n*

2) Obtain the mean of the sample means

i.e $\overline{\overline{X}} = \dfrac{\overline{X}_1 + \overline{X}_2 + ... + \overline{X}_n}{n}$

where n is the total number of observations

3) The control limits are set at

U.C.L. = $\overline{\overline{X}} + A_2\overline{R}$

LCL = $\overline{\overline{X}} - A_2\overline{R}$, where $\overline{R} = \dfrac{\sum_{i=1}^{n} R_i}{n}$, where R_i are the sample ranges.

The values of A_2 for different n can be obtained from the tables.

Example 28

The following data relate to the life (in hours) of 10 samples of 6 electric bulbs each drawn at an interval of one hour from a production process. Draw the control chart for $\overline{X}$ and R and comment.

Sample No.	**Life time (in hours)**					
1	**620**	**687**	**666**	**689**	**738**	**686**
2	**501**	**585**	**524**	**585**	**653**	**668**
3	**673**	**701**	**686**	**567**	**619**	**660**
4	**646**	**626**	**572**	**628**	**631**	**743**
5	**494**	**984**	**659**	**643**	**660**	**640**
6	**634**	**755**	**625**	**582**	**683**	**555**
7	**619**	**710**	**664**	**693**	**770**	**534**
8	**630**	**723**	**614**	**535**	**550**	**570**
9	**482**	**791**	**533**	**612**	**497**	**499**
10	**706**	**524**	**626**	**503**	**661**	**754**

(Given for n = 6, A_2 = 0.483, D_3 = 0, D_4 = 2.004)

Solution :

Sample No.	Total	Sample Mean $\bar{X}$	Sample Range R
1	4086	681	118
2	3516	586	167
3	3906	651	134
4	3846	641	171
5	4080	680	490
6	3834	639	200
7	3990	665	236
8	3622	604	188
9	3414	569	309
10	3774	629	251
Total		**6345**	**2264**

Central line $\bar{\bar{X}}$ = mean of the sample means = 634.5

$\bar{R}$ = mean of the sample ranges = 226.4

U.C.L. $= \bar{\bar{X}} + A_2\bar{R}$

$= 634.5 + 0.483 \times 226.4$

$= 634.5 + 109.35 = 743.85$

L.C.L. $= \bar{\bar{X}} - A_2\bar{R}$

$= 634.5 - 0.483 \times 226.4$

$= 634.5 - 109.35 = 525.15$

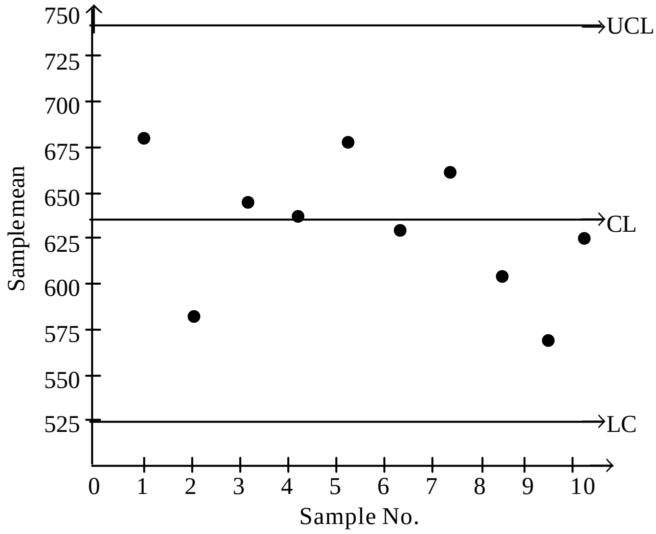
Central line $\bar{R} = 226.4$

U.C.L. $= D_4\bar{R} = 2.004 \times 226.4$

$= 453.7056$

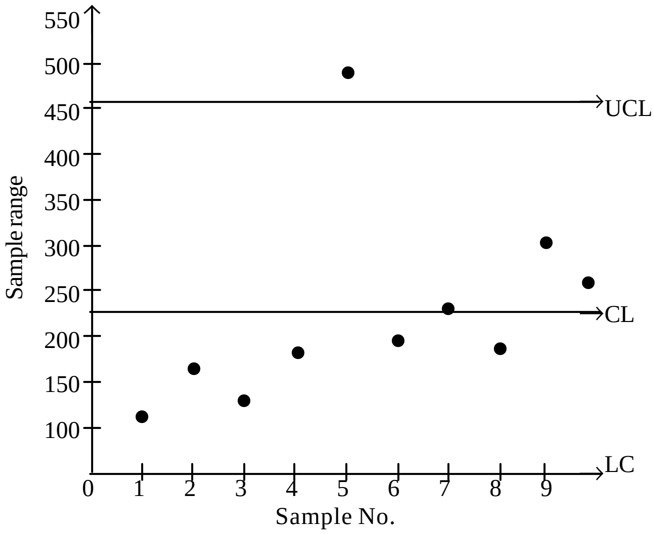
L.C.L. $= D_3\bar{R} = 0 \times 226.4 = 0$

$\overline{X}$ Chart



(Fig. 10.10)

R Chart



(Fig. 10.11)

Conclusion :

Since one of the points of the sample range is outside the UCL of R chart, the process is not in control.

Example 29

The following data shows the value of sample mean $\overline{X}$ and the range R for ten samples of size 5 each. Calculate the values for central line and control limits for mean chart and range chart and determine whether the process is in control

Sample no.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mean $\overline{X}$	**11.2**	**11.8**	**10.8**	**11.6**	**11.0**	**9.6**	**10.4**	**9.6**	**10.6**	**10.0**
Range (R)	**7**	**4**	**8**	**5**	**7**	**4**	**8**	**4**	**7**	**9**

(Given for $n = 5$, $A_2 = 0.577$, $D_3 = 0$ $D_4 = 2.115$)

Solution :

Control limits for $\overline{X}$ chart

$$\overline{\overline{X}} = \frac{1}{n}\Sigma\overline{X}$$

$$= \frac{1}{10}(11.2 + 11.8 + 10.8 + ...+ 10.0) = 10.66$$

$$\overline{R} = \frac{1}{n}\Sigma R = \frac{1}{10}(63) = 6.3$$

$$\text{U.C.L} = \overline{\overline{X}} + A_2\overline{R}$$

$$= 10.66 + (0.577 \times 6.3) = 14.295$$

$$\text{L.C.L.} = \overline{\overline{X}} - A_2\overline{R}$$

$$= 10.66 - (0.577 \times 6.3) = 7.025$$

CL = Central line = $\overline{\overline{X}}$ = 10.66

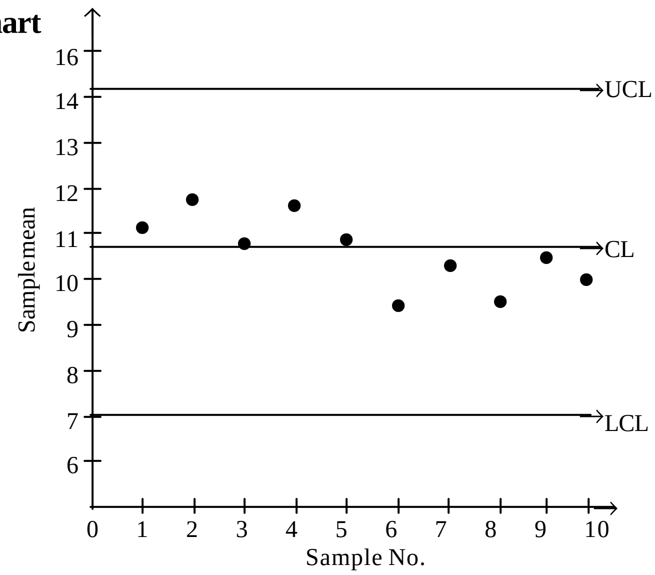
Range chart

U.C.L. = $D_4\overline{R}$ = 2.115 x 6.3 = 13.324

L.C.L. = $D_3\overline{R}$ = 0

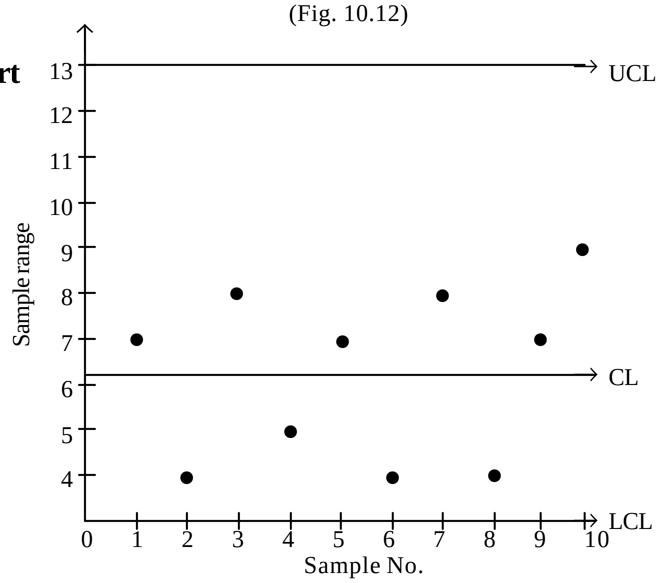
C.L. = $\overline{R}$ = 6.3

$\overline{X}$ Chart



(Fig. 10.12)

R Chart



(Fig. 10.13)

Conclusion :

Since all the points of sample mean and range are within the control limits, the process is in control.

EXERCISE 10.5

1) The following are the $\overline{X}$ and R values for 20 samples of 5 readings. Draw $\overline{X}$ chart and R chart and write your conclusion.

Samples	1	2	3	4	5	6	7	8	9	10
$\overline{X}$	34	31.6	30.8	33	35	33.2	33	32.6	33.8	37.8
R	4	4	2	3	5	2	5	13	19	6

Samples	11	12	13	14	15	16	17	18	19	20
$\overline{X}$	35.8	38.4	34	35	38.8	31.6	33	28.2	31.8	35.6
R	4	4	14	4	7	5	5	3	9	6

(Given for $n = 5$, $A_2 = 0.58$, $D_3 = 0$ $D_4 = 2.12$)

2) From the following, draw $\overline{X}$ and R chart and write your conclusion.

Sample no.	1	2	3	4	5	6	7	8	9	10
	140	138	139	143	142	136	142	143	141	142
	143	143	133	141	142	144	147	137	142	137
	137	143	147	137	145	143	137	145	147	145
	134	145	148	138	135	136	142	137	140	140
	135	146	139	140	136	137	138	138	140	132

Sample no.	11	12	13	14	15	16	17	18	19	20
	137	137	142	137	144	140	137	137	142	136
	147	146	142	145	142	132	137	142	142	142
	142	142	139	144	143	144	142	142	143	140
	137	142	141	137	135	145	143	145	140	139
	135	140	142	140	144	141	141	143	135	137

(Given for $n = 5$, $A_2 = 0.58$, $D_3 = 0$, $D_4 = 2.12$)

3) From the following data construct $\overline{X}$ and R chart and write your conclusion

Sample no.	1	2	3	4	5	6	7	8	9
	46	41	43	37	37	37	44	35	37
	40	42	40	40	40	38	39	39	44
	48	49	46	47	46	49	43	48	48

Sample no.	10	11	12	13	14	15	16	17	18
	45	48	36	40	42	38	47	42	47
	43	44	42	39	40	40	44	45	42
	49	48	48	48	48	48	49	37	49

(Given for $n = 3$, $A_2 = 1.02$, $D_3 = 0$, $D_4 = 2.58$)

EXERCISE 10.6

Choose the correct answer

1) A time series is a set of data recorded

(a) periodically (b) at equal time intervals

(c) at successive points of time (d) all the above

2) A time series consists of

(a) two components (b) three components

(c) four components (d) none of these

3) The component of a time series attached to long term variation is termed as

(a) cyclic variations (b) secular trend

(c) irregular variation (d) all the above

4) The component of a time series which is attached to short term fluctuations is

(a) seasonal variation (b) cyclic variation

(c) irregular variation (d) all the above

5) Cyclic variations in a time series are casued by

(a) lock out in a factory (b) war in a country

(c) floods in the states (d) none of these

6) The terms prosperity, recession depression and recovery are in particular attached to

(a) Secular trend (b) seasonal fluctuation

(c) cyclic movements (d) irregular variation

7) An additive model of time series with the components T, S, C and I is

(a) $Y = T + S + C - I$ (b) $Y = T + S \times C + I$

(c) $Y = T + S + C + I$ (d) $Y = T + S + C \times I$

8) A decline in the sales of ice cream during November to March is associated with

(a) Seasonal variation (b) cyclical variation

(c) random variation (d) secular trend

9) Index number is a

(a) measure of relative changes

(b) a special type of an average

(c) a percentage relative

(d) all the above.

10) Index numbers are expressed

(a) in percentages (b) in ratios

(c) in terms of absolute value (d) all the above

11) Most commonly used index numbers are

(a) Diffusion index number (b) price index number

(c) value index number (d) none of these

12) Most frequently used index number formulae are

(a) weighted formulae (b) Unweighted formulae

(c) fixed weighted formulae (d) none of these

13) Laspeyre's index formula uses the weights of the

(a) base year quantities (b) current year prices

(c) average of the weights of number of years

(d) none of these

14) The weights used in Paasche's formula belong to

(a) the base period (b) the current period

(c) to any arbitrary chosen period (d) none of these

15) Variation in the items produced in a factory may be due to

(a) chance causes (b) assignable causes

(c) both (a) and (b) (d) neither (a) or (b)

16) Chance variation in the manufactured product is

(a) controlable (b) not controlable

(c) both (a) and (b) (d) none of these

17) The causes leading to vast variation in the specification of a product are usually due to

(a) random process (b) assignable causes

(c) non-traceable causes (d) all the above

18) Variation due to assignable causes in the product occur due to

(a) faulty process (b) carelessness of operators

(c) poor quality of raw material (d) all the above

19) Control charts in statistical quality consist of

(a) three control lines

(b) upper and lower control limits

(c) the level of process

(d) all the above

20) The range of correlation co-efficient is

(a) 0 to ∞ (b) $-\infty$ 10 ∞

(c) -1 to 1 (d) none of these

21) If X and Y are two variates, there can be atmost

(a) one regression line (b) two regression lines

(c) three regression lines (d) none of these

22) In a regression line of Y on X, the variable X is known as

(a) independent variable (b) dependent variable

(c) both (a) and (b) (d) none of these

23) Scatter diagram of the variate values (X, Y) give the idea about

(a) functional relationship (b) regression model

(c) distribution of errors (d) none of these

24) The lines of regression intersect at the point

(a) (X, Y) (b) $(\overline{X}, \overline{Y})$

(c) (0, 0) (d) none of these

25) The term regression was introduced by

(a) R.A.Fisher (b) Sir Francis Galton

(c) Karl pearson (d) none of these.