

Modular Arithmetic

— Manjula M, Assistant Professor
BGSCET

Divisibility

a divides b [$a \neq 0$], if there exists an integer ' k ' such that $b = ka$. Here a & b are two integers.

a divides b is written as $a|b$.

Ex:- $4|-16$ [$\because -16 = -4 \times 4$].

$3 \nmid 19$

Fundamental theorem of Arithmetic

Every positive integer greater than 1 can be written uniquely as a prime or as the product of two or more primes [where the prime factors are written in order of non-decreasing size]

Ex:- $64 = 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^6$

$641 = 641$

$100 = 2^2 \times 5^2$

Note: Consequence of Fundamental Theorem

- 1) If a prime number p divides ab , then either $p|a$ or $p|b$.
- 2) If $m|a$ & $n|a$ and if m & n have no divisors greater than 1 in common, then $mn|a$.

Division Algorithm

Let a & b be positive integers where $a > b$ then $\exists$ unique integers q & r such that $a = bq + r$, $0 \leq r < b$. Here a is called dividend, b is called the divisor, q is called the quotient & r is called the remainder [$r = a \bmod b$]

Ex: 1) $a = 7$, $b = 3$, then

$$7 = 3 \times 2 + 1 \Rightarrow q = 2 \text{ \& } r = 1$$

2) $a = -5$, $b = 2$ then

$$-5 = 2 \times (-3) + 1 \Rightarrow q = -3 \text{ \& } r = 1$$

$$\begin{array}{r} -3 \\ 2 \overline{) -5} \\ \underline{+6} \\ 1 \end{array}$$

Modulo operator

a modulo b or $a \bmod b$ is denoted by $a \% b$. It is remainder obtained when a is divided by b .

If $a \bmod b = 0$, then $b | a$ where a and b are the integers.

Ex: 1) $3 \bmod 2 = 1$ & $2 \bmod 3 = 2$

2) $1547 \bmod 560 = 427$

3) $-3 \bmod 12 = 9$

4) $-7 \bmod 12 = 5$

5) $-13 \bmod 12 = 11$

$$\begin{array}{r} -1 \\ 12 \overline{) -3} \\ \underline{+12} \\ 9 \\ 12 \overline{) -1} \\ \underline{+12} \\ -11 \end{array} \quad \begin{array}{r} -2 \\ 12 \overline{) -13} \\ \underline{+24} \\ 11 \end{array}$$

(3)

Greater Common divisor [G.C.D] or Highest Common Factor [H.C.F]

Given 2 integers a and b both being non-zero, the $GCD(a, b)$ is the largest integer ' d ' dividing both a & b .

Ex:- Common divisors of 24 & 36 are
1, 2, 3, 4, 6, 12

$$\therefore GCD(24, 36) \text{ or } (24, 36) = 12$$

Note:

If $GCD(a, b) = 1$ then a & b are relatively prime or co-prime

Ex:- $GCD(8, 15) = 1$.

The Euclidean Algorithm

This algorithm is used to find GCD of 2 numbers, a & b with $a > b$.

Step 1: Let a, b be 2 integers, $a > b$

Step 2: $a \bmod b = R$

Step 3: Let $a = b$ & $b = R$

Step 4: Repeat steps 2 & 3 until $a \bmod b$ is equal to 0.

Step 5: The last non zero remainder is the $GCD(a, b)$.

1) Find the GCD of 32 & 54 and express it in the form $32x + 54y$.

Soln:-

$$\begin{array}{r} 32 \overline{) 54} (1 \\ \underline{32} \\ 22 = R \end{array}$$

$$22 = 54 - 1(32) \rightarrow (1)$$

$$\begin{array}{r} 22 \overline{) 32} (1 \\ \underline{22} \\ 10 \end{array}$$

$$10 = 32 - 1(22) \rightarrow (2)$$

$$\begin{array}{r} 10 \overline{) 22} (2 \\ \underline{20} \\ 2 \end{array}$$

$$2 = 22 - 2(10) \rightarrow (3)$$

$$\begin{array}{r} 2 \overline{) 10} (5 \\ \underline{10} \\ 0 \end{array}$$

$\therefore$ The last non zero remainder is 2

$$\therefore \text{GCD}(32, 54) = 2$$

From (3), $2 = 22 - 2(10)$

$$2 = [54 - 1(32)] - 2[32 - 1(22)]$$

$$= 54 - 3(32) + 2(22)$$

$$= 54 - 3(32) + 2[54 - 1(32)]$$

$$= 3(54) - 5(32)$$

$$2 = 32(-5) + 54(3)$$

i.e. $2 = 32x + 54y$, where $x = -5$, $y = 3$.

9) Find GCD of 25520 & 19314 & express it in the form $25520x + 19314y$.

Soln $19314 \overline{) 25520} (1$
 $\underline{19314}$
 6206

$$6206 = 25520 - 1(19314)$$

$$6206 \overline{) 19314} (3$$

$$\underline{18618}$$

$$696$$

$$696 = 19314 - 3(6206)$$

$$696 \overline{) 6206} (8$$

$$\underline{5568}$$

$$638$$

$$638 = 6206 - 8(696)$$

$$638 \overline{) 696} (1$$

$$\underline{638}$$

$$58$$

$$58 = 696 - 1(638)$$

$$58 \overline{) 638} (11$$

$$\underline{638}$$

$$0$$

$$58 = 696 - 1(638)$$

$$58 = 19314 - 3(6206) - [6206 - 8(696)]$$

$$= 19314 - 4(6206) + 8(696)$$

$$= 19314 - 4[25520 - 1(19314)]$$

$$+ 8[19314 - 3(6206)]$$

$$= 13(19314) - 4(25520) - 24(6206)$$

$$= 13(19314) - 4(25520) - 24[25520 - 1(19314)]$$

$$= -28(25520) + 37(19314)$$

Assignment

1) Find GCD(123, 277) Ans $\rightarrow (1)$

2) Find GCD(252, 198) $\rightarrow (18) = 252(4) + 198(-5)$

3) Find GCD(360, 294) $\rightarrow (6)$

Congruences

Let 'm' be a +ve integer greater than 1.
If a & b are any integers then a is said to be congruent to b modulo m [$a \equiv b \pmod{m}$]
iff $m \mid a-b$. [or a & b leave same remainder when divided by m]

Ex:- $25 \equiv 3 \pmod{11} \Rightarrow 11 \mid 22$
 $79 \not\equiv 8 \pmod{9} [\because 9 \nmid 71]$
 $-69 \equiv -5 \pmod{16}$

Problems

Solve $7x \equiv 9 \pmod{15}$

$15 \mid 7x-9$

$7x-9 = 15K$, where $K \in \mathbb{Z}$

$$x = \frac{15K+9}{7}$$

By inspection, $K=5$ gives the integer value of x.

i.e. $x = \frac{15(5)+9}{7}$

$$\boxed{x=12}$$

$\therefore x \equiv 12 \pmod{15}$

Extended Euclidean Algorithm

If $d = \gcd(a, b)$, where $a > b$, then $\exists$ integers x & y such that $d = ax + by$.

2) Solve $5x \equiv 4 \pmod{13}$

Soln:

$$13 \mid 5x - 4$$

$$5x - 4 = 13k, \text{ where } k \in \mathbb{Z}$$

$$x = \frac{13k + 4}{5}$$

By inspection, $k=2$ gives integer value of x

$$x = \frac{13(2) + 4}{5}$$

$$\boxed{x=6} \quad \text{i.e. } x \equiv 6 \pmod{13}$$

$\therefore$ The general soln is $x = 6 + 13r, r \in \mathbb{Z}$.

3) If $2^8 \equiv a \pmod{13}$, find 'a'.

Soln:

$$2^8 = (2^4)^2 = 256$$

$$256 \equiv 9 \pmod{13}$$

$$\therefore \boxed{a=9}$$

$$\begin{array}{r} 13 \overline{) 256} \quad (19) \\ \underline{13} \\ 126 \\ \underline{117} \\ 9 \end{array}$$

4) Find the least +ve value of x such that

~~*** Model QP~~
(P) $71 \equiv x \pmod{8}$

Soln

$$71 \equiv 7 \pmod{8}$$

$$\therefore \boxed{x=7}$$

$$\begin{array}{r} 8 \overline{) 71} \quad (8) \\ \underline{64} \\ 7 \end{array}$$

*** Model QP

(ii) $78 + x \equiv 3 \pmod{5}$

Soln: $5 \mid 78 + x - 3$

$$78 + x = 5K$$

Let $x = 5$, $75 + 5 = 80$ [$\because 80$ is multiple of 5]

$\therefore$ The least value of x is 5.

*** Model QP

(iii) $89 \equiv (x+3) \pmod{4}$

$$4 \mid 89 - x - 3$$

Soln: $89 - x - 3 \equiv 4K$

$$86 - x = 4K$$

Let $x = 2$, $86 - 2 = 84$

$\therefore$ The least value of x is 2.

(iv) $96 \equiv \left(\frac{x}{7}\right) \pmod{5} \Rightarrow 5 \mid 96 - \frac{x}{7}$

$$96 - \frac{x}{7} = 5K$$

Let $x = 7$, $96 - \frac{7}{7} = 95$

$\therefore$ The least value of x is 7.

(v) $5x \equiv 4 \pmod{6}$

$$5x - 4 = 6K$$

Let $x = 2$, $10 - 4 = 6$

$\therefore$ The least +ve value of x is 2

5) If $2x \equiv 3 \pmod{7}$, find 'x' such that $9 \leq x \leq 30$

Soln:- $7 \mid 2x - 3 \Rightarrow 2x - 3 = 7k$

By inspection, $x = 5$

$\therefore$ General soln is $x \equiv 5 \pmod{7}$

Soln set = $\{ \dots, -9, -2, 5, 12, 19, 26, 33, \dots \}$

$\therefore$ Required values of x are 12, 19, 26.

6) Solve $6x \equiv 15 \pmod{21}$

Soln:- $21 \mid 6x - 15 \Rightarrow 6x - 15 = 21k$

$$x = \frac{21k + 15}{6}$$

By inspection, $k=1$ gives $x = 6$

$\therefore$ General soln is $x \equiv 6 \pmod{21}$

7) Solve $12x \equiv 6 \pmod{21}$

Soln:- $21 \mid 12x - 6 \Rightarrow 12x - 6 = 21k$

$$x = \frac{21k + 6}{12}$$

By inspection, $k=2$ gives $x = 4$

$\therefore$ General soln is $x \equiv 4 \pmod{21}$

8) Solve $18x \equiv 30 \pmod{42}$

Soln:- $42 \mid 18x - 30 \Rightarrow 18x - 30 = 42k \Rightarrow x = \frac{42k + 30}{18}$

By inspection, $k=1 \Rightarrow x = 5$

Properties of Congruences

- 1) $a \equiv a \pmod{m}$
- 2) If $a \equiv b \pmod{m}$, then $b \equiv a \pmod{m}$
- 3) If $a \equiv b \pmod{m}$ & $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$.
- 4) If $a \equiv b \pmod{m}$ & $c \equiv d \pmod{m}$, then $a + c \equiv b + d \pmod{m}$
- 5) If $a \equiv b \pmod{m}$ & $c \equiv d \pmod{m}$, then $ac \equiv bd \pmod{m}$
- 6) If $a \equiv b \pmod{m}$ then $a^k \equiv b^k \pmod{m}$ for every positive integer k
- 7) If $ac \equiv bc \pmod{m}$ & $(c, m) = d$, then $a \equiv b \pmod{\frac{m}{d}}$.

Steps to find x in linear
Congruence, $ax \equiv b \pmod{m}$ by
Inverse method

- 1) Find $(a, m) = d$
- 2) Find $\frac{b}{d} \rightarrow$ If this is an integer
soln exists.
- 3) Multiply both sides by multiplicative
inverse of 'a'.
i.e. $(a^{-1})x \equiv a^{-1}b \pmod{m}$, which
gives value of 'x'.
- 4) General solution is given by
$$x_k = x_0 + k\left(\frac{m}{d}\right), \text{ where}$$
$$k = \{0, 1, 2, \dots, (d-1)\}$$

*** Model QP

▷ Solve $11x \equiv 4 \pmod{25}$

Soln $25 \mid 11x - 4$ $x = \frac{25k + 4}{11}$

$11x - 4 = 25k$

By inspection
 $k = 6$

$x = 14$

OR

Inverse method [To find x s.t. $11x \equiv 1 \pmod{25}$]

$11 \mid 25(2$

$\frac{22}{3}$

$\frac{9}{2}$

$3 \mid 11(3$

$\frac{2}{1}$

$\frac{1}{2}$

$2 \mid 3(1$

$\frac{2}{1}$

$d = \frac{2}{1} = \text{GCD}(11, 25) \Rightarrow 1 = [25 - 2(11)] - 1[11 - 3(3)]$

$1 \mid 2(2$

$\frac{2}{0}$

$\frac{0}{0}$

$(11, 25) = 1$

$\frac{b}{d} = \frac{4}{1} = 4 \in \mathbb{Z} \Rightarrow$
soln exists.

$3 = 25 - 2(11) \rightarrow \textcircled{1}$

$2 = 11 - 3(3) \rightarrow \textcircled{2}$

$1 = 3 - 1(2) \rightarrow \textcircled{3}$

$1 = 25 - 2(11) - 1[11 - 3(3)]$

$1 = 25 - 3(11) + 3(3)$

$1 = 25 - 3(11) + 3[25 - 2(11)]$

$1 = 4(25) - 9(11)$

$1 \pmod{25} \equiv 4(25) \pmod{25}$

$-9(11) \pmod{25}$

$1 \pmod{25} \equiv 0 - 9(11) \pmod{25}$

$1 \equiv -9(11) \pmod{25}$

$\Rightarrow 11(-9) \equiv 1 \pmod{25}$

$[(11) 11^{-1} \equiv 1 \pmod{25}]$

$\Rightarrow 11^{-1} = -9$

We have,

$$11x \equiv 4 \pmod{25}$$

$$(11^{-1}) 11x \equiv (11^{-1}) 4 \pmod{25}$$

$$(11^{-1}, 11)x \equiv (11^{-1}) 4 \pmod{25}$$

$$x \equiv (9)(4) \pmod{25} \Rightarrow x \equiv -36 \pmod{25}$$

$$x \equiv 14 \pmod{25} \Rightarrow \boxed{x = 14}$$

2) Solve $14x \equiv 12 \pmod{18}$

Soln:- $7x \equiv 6 \pmod{9} \Rightarrow 9 \mid 7x - 6$

$$7x - 6 = 9k \Rightarrow \text{By inspection, } \boxed{x = 6}$$

OR

Find 'x'

$$\text{a.t. } 7x \equiv 1 \pmod{9}$$

$$\begin{array}{r} 7 \overline{) 9(1} \\ \underline{7} \\ 2 \end{array}$$

$$\begin{array}{r} 2 \overline{) 7(3} \\ \underline{6} \\ 1 \end{array}$$

$$\begin{array}{r} 1 \overline{) 2(2} \\ \underline{2} \\ 0 \end{array}$$

$$(7, 9) = 1 \quad \& \quad \frac{b}{d} = \frac{6}{1} = 6 \Rightarrow \text{soln exists}$$

$$2 = 9 - 1(7)$$

$$1 = 7 - 3(2)$$

$$\Rightarrow 1 = 7 - 3[9 - 1(7)]$$

$$1 = 4(7) + 3(9)$$

$$1 \pmod{9} \equiv 4(7) \pmod{9} + 0$$

$$1 \equiv 4(7) \pmod{9}$$

$$7(4) \equiv 1 \pmod{9} \Rightarrow 7^{-1} = 4$$

Given, $7x \equiv 6 \pmod{9}$

$$7^{-1} \cdot 7x \equiv 7^{-1} \cdot 6 \pmod{9}$$

$$x \equiv 4 \cdot 6 \pmod{9}$$

$$x \equiv 24 \pmod{9}$$

$$x \equiv 6 \pmod{9}$$

$$\Rightarrow \boxed{x = 6}$$

Fermat's theorem

[i.e. $(a, p) = 1$]

Let 'p' be a prime & $p \nmid a$, then
 $a^{p-1} \equiv 1 \pmod{p}$

Ex: $a=2$ & $5=p$, $5 \nmid 2$ or $(5, 2) = 1$

$$2^4 \equiv 1 \pmod{5}$$

Euler's Totient Function

It is defined as no of positive integers less than 'n' & relatively prime to 'n'. Ex: $\phi(6) = 2$ [1, 2, 3, 4, 5]

* $\phi(1) = 1$

* If 'p' is prime, then $\phi(p) = p - 1$

* For any positive integer 'n' if $n = pq$ where 'p' & q are primes with $p \neq q$, then $\phi(n) = \phi(pq) = \phi(p) \cdot \phi(q) = (p-1)(q-1)$

* For any positive integer n, if $n = p_1^{a_1} \cdot p_2^{a_2} \cdots p_m^{a_m}$, then

$$\phi(n) = n \left[1 - \frac{1}{p_1} \right] \cdot \left[1 - \frac{1}{p_2} \right] \cdots \left[1 - \frac{1}{p_m} \right]$$

Ex: 1) $\phi(7) = 7 - 1 = 6$ [1, 2, 3, 4, 5, 6]

2) $\phi(15) = \phi(3 \times 5) = 2 \times 4 = 8$

3) $\phi(24) = \phi(2^3 \times 3) = 2^3 \left(1 - \frac{1}{2} \right) \times \frac{2}{3} = 8$
 $= 2^3 \left[1 - \frac{1}{2} \right] \left[1 - \frac{1}{3} \right]$

Euler's theorem:

For every a and n that are relatively prime, $a^{\phi(n)} \equiv 1 \pmod{n}$

Ex: $a = 7, n = 10, (7, 10) = 1$

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

$$7^4 \equiv 1 \pmod{10}$$

$$\left[\begin{aligned} \phi(10) &= \phi(2) \times \phi(5) \\ &= 1 \times 4 = 4 \end{aligned} \right]$$

1) Find the least positive remainder when 2^{301} is divided by 5.

Soln:- $p=5, a=2, (2,5)=1$

By Fermat's theorem,

$$a^{p-1} \equiv 1 \pmod{p}$$

$$2^4 \equiv 1 \pmod{5}$$

$$(2^4)^{75} \equiv 1^{75} \pmod{5}$$

$$2^{300} \equiv 1 \pmod{5} \text{ \& } 2 \equiv 2 \pmod{5}$$

$$\Rightarrow 2^{301} \equiv 2 \pmod{5} \Rightarrow \text{Remainder} = 2$$

$$\begin{array}{r} 4 \overline{) 301} \text{ (75)} \\ \underline{300} \\ 1 \end{array}$$

2) Find the unit digit in the no 7^{289}

Soln:- $a=7, n=10, (7,10)=1$

By Euler's thm, $a^{\phi(n)} \equiv 1 \pmod{n}$

$$a^{\phi(10)} \equiv 1 \pmod{10} \quad [\phi(10) = \phi(5) \times \phi(2) = 4 \times 1 = 4]$$

$$7^4 \equiv 1 \pmod{10}$$

$$(7^4)^{72} \equiv 1^{72} \pmod{10}$$

$$7^{288} \equiv 1 \pmod{10} \text{ \& } 7 \equiv 7 \pmod{10}$$

$$\begin{array}{r} 4 \overline{) 289} \text{ (72)} \\ \underline{288} \\ 1 \end{array}$$

$$\Rightarrow 7^{289} \equiv 7 \pmod{10}$$

$\therefore$ Unit digit in 7^{289} is 7.

*** Model QP

3) Find the last digit of 7^{2013} .

Soln:- $a=7, n=10, (7,10)=1$

By Euler's thm, $a^{\phi(n)} \equiv 1 \pmod{n}$

$$7^4 \equiv 1 \pmod{10}$$

$$(7^4)^{503} \equiv 1^{503} \pmod{10}$$

$$7^{2012} \equiv 1 \pmod{10} \text{ \& } 7 \equiv 7 \pmod{10}$$

$$7^{2013} \equiv 7 \pmod{10} \Rightarrow \text{Last digit} = \underline{7}$$

$$\begin{array}{r} 4) 2013 \text{ (503)} \\ \underline{2012} \\ 1 \end{array}$$

*** Model QP

4) Find the unit digit of 7^{118} .

Soln:- $a=7, n=10, (7,10)=1$

By Euler's thm, $a^{\phi(n)} \equiv 1 \pmod{n}$

$$7^4 \equiv 1 \pmod{10} \Rightarrow (7^4)^{29} \equiv 1^{29} \pmod{10}$$

$$7^{116} \equiv 1 \pmod{10} \text{ \& } 7^2 \equiv 7^2 \pmod{10} \quad [7^2 \equiv 9 \pmod{10}]$$

$$7^{118} \equiv 49 \pmod{10}$$

$$7^{118} \equiv 9 \pmod{10} \Rightarrow \text{Unit digit} = \underline{9}$$

*** Model QP

5) Find the last digit of 13^{37} .

Soln:- $a=13, n=10, (10,13)=1$

By Euler's thm, $13^{\phi(10)} \equiv 1 \pmod{10}$

$$13^4 \equiv 1 \pmod{10} \Rightarrow (13^4)^9 \equiv 1^9 \pmod{10}$$

$$13^{36} \equiv 1 \pmod{10} \text{ \& } 13^1 \equiv 13 \pmod{10}$$

$$13^{37} \equiv 13 \pmod{10}, 13^{37} \equiv 3 \pmod{10} \Rightarrow \text{Last digit} = \underline{3}$$

6) What is the remainder in the division of 2^{50} by 7?

Soln:- $p=7$ $a=2$ $(2, 7)=1$

By Fermat's thm,

$$a^{p-1} \equiv 1 \pmod{p}$$

$$2^6 \equiv 1 \pmod{7}$$

$$(2^6)^8 \equiv 1^8 \pmod{7} \quad \& \quad 2^2 \equiv 4 \pmod{7}$$

$$\Rightarrow 2^{50} \equiv 4 \pmod{7} \Rightarrow \boxed{r=4}$$

*** Model QP

7) Find the remainder when 2^{23} is divided by 47.

Soln:- $a=2$ $p=47$ $(2, 47)=1$

By Fermat's thm,

$$a^{p-1} \equiv 1 \pmod{p}$$

$$2^{46} \equiv 1 \pmod{47}$$

46 can't be reduced to 23 & hence Fermat's thm cannot be used

$$2^6 \equiv 17 \pmod{47}$$

$$(2^6)^3 \equiv 17^3 \pmod{47}$$

$$2^{18} \equiv 25 \pmod{47} \quad \& \quad 2^5 \equiv 2^5 \pmod{47}$$

$$2^{23} \equiv 800 \pmod{47}$$

$$2^{23} \equiv 1 \pmod{47} \Rightarrow \boxed{r=1}$$

*** Model QP

8) Find the remainder when the number 2^{1000} is divided by 13.

Soln: $a=2$, $p=13$, $(2,13)=1$

By Fermat's thm, $a^{p-1} \equiv 1 \pmod{p}$

$$2^{12} \equiv 1 \pmod{13}$$

$$(2^{12})^{83} \equiv 1^{83} \pmod{13}$$

$$2^{996} \equiv 1 \pmod{13} \quad \& \quad 2^4 \equiv 3 \pmod{13}$$

$$2^{1000} \equiv 3 \pmod{13}$$

$$\Rightarrow \boxed{r=3}$$

$$\begin{array}{r} 12 \overline{) 1000} \\ \underline{996} \\ 4 \end{array}$$

*** 9) Find the remainder when $135 \times 74 \times 48$ is divided by 7.

Soln: $135 \equiv 2 \pmod{7}$

$$74 \equiv 4 \pmod{7}$$

$$48 \equiv 6 \pmod{7}$$

$$135 \times 74 \times 48 \equiv 48 \pmod{7}$$

$$135 \times 74 \times 48 \equiv 6 \pmod{7}$$

$$\therefore \boxed{r=6}$$

10) Find the remainder obtained when $64 \times 65 \times 66$ is divided by 67.

Soln:-

$$64 \equiv -3 \pmod{67}$$

$$65 \equiv -2 \pmod{67}$$

$$66 \equiv -1 \pmod{67}$$

$$\begin{array}{r} 67 \overline{) 6441} \\ \underline{-67} \\ -3 \end{array}$$

$$64 \times 65 \times 66 \equiv -6 \pmod{67}$$

$$64 \times 65 \times 66 \equiv 61 \pmod{67}$$

$$\begin{array}{r} 67 \overline{) -61} \\ \underline{67} \\ 61 \end{array}$$

$\therefore \boxed{r = 61}$

*** Model QP

11) Find the remainder when $349 \times 74 \times 36$ is divided by 3.

Soln:-

$$349 \equiv 1 \pmod{3}$$

$$74 \equiv 2 \pmod{3}$$

$$36 \equiv 0 \pmod{3}$$

$$349 \times 74 \times 36 \equiv 0 \pmod{3}$$

$\therefore \boxed{r = 0}$

*** Model QP

12) Find the remainder when $175 \times 113 \times 53$ is divided by 11.

Soln:-

$$175 \equiv 10 \pmod{11}$$

$$113 \equiv 3 \pmod{11}$$

$$53 \equiv 9 \pmod{11}$$

$$175 \times 113 \times 53 \equiv 270 \pmod{11}$$

$$175 \times 113 \times 53 \equiv 6 \pmod{11}$$

$$\therefore \boxed{x = 6}$$

18) Solve $3^{202} \pmod{13}$ by Euler's theorem

Soln:- $(3, 13) = 1$

Hence by Euler's theorem,

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

$$3^{\phi(13)} \equiv 1 \pmod{13}$$

$$3^{12} \equiv 1 \pmod{13}$$

$$(3^{12})^{16} \equiv 1 \pmod{13}$$

$$3^{192} \equiv 1 \pmod{13} \checkmark$$

$$3^{10} \equiv 3^1 \pmod{13} \Rightarrow 3^{10} \equiv 3 \pmod{13} \checkmark$$

$$3^{202} \equiv 3 \pmod{13}$$

$$\therefore 3^{202} \pmod{13} = 3$$

19) Solve $4^{99} \pmod{35}$

Soln:- $(4, 35) = 1$

$$\begin{aligned} \phi(35) &= \phi(7) \times \phi(5) \\ &= 6 \times 4 = 24 \end{aligned}$$

$$4^{\phi(35)} \equiv 1 \pmod{35}$$

$$4^{24} \equiv 1 \pmod{35}$$

$$(4^{24})^4 \equiv 1 \pmod{35}$$

$$4^{96} \equiv 1 \pmod{35} \text{ \& } 4^3 \equiv 29 \pmod{35}$$

$$\therefore 4^{99} \equiv 29 \pmod{35}$$

$$\Rightarrow 4^{99} \pmod{35} = 29$$

15) Use Euler's theorem to find the unit digit in 3^{100} .

Soln: $(3, 10) = 1$ $\phi(10) = 4 \times 1 = 4$

$$3^{\phi(10)} \equiv 1 \pmod{10}$$

$$3^4 \equiv 1 \pmod{10}$$

$$(3^4)^{25} \equiv 1 \pmod{10}$$

Hence 1 is the unit digit.

16) ~~*** Model QP~~
S.T $8^{30} - 1$ is divisible by 31.

Soln: $a = 8, P = 31$

$$\text{GCD}(8, 31) = 1$$

By Fermat's little theorem

$$a^{P-1} \equiv 1 \pmod{P}$$

$$a^{P-1} - 1 \equiv 0 \pmod{P}$$

$$8^{31-1} - 1 \equiv 0 \pmod{31}$$

$$8^{30} - 1 \equiv 0 \pmod{31}$$

17) Find the remainder when 72^{1001} is divided by 31

Soln: $(72, 31) = 1, p = 31$

By Fermat's thm,

$$72^{31-1} \equiv 1 \pmod{31}$$

$$72^{30} \equiv 1 \pmod{31}$$

$$[(72)^{30}]^{33} \equiv 1^{33} \pmod{31}$$

$$(72)^{990} \equiv 1 \pmod{31} \rightarrow (1)$$

$$(72)^2 \equiv 7 \pmod{31}$$

$$[(72)^2]^5 \equiv 7^5 \pmod{31}$$

$$(72)^{10} \equiv 5 \pmod{31} \rightarrow (2)$$

$$72 \equiv 72 \pmod{31} \rightarrow (3)$$

$$(72)^{1001} \equiv (72^{4 \times 5 \times 1}) \pmod{31}$$

$$(72)^{1001} \equiv 360 \pmod{31}$$

$$(72)^{1001} \equiv 19 \pmod{31}$$

$\therefore$ remainder = 19.

18) S.T $2^{340} \equiv 1 \pmod{31}$ by Fermat's little theorem [OR $2^{340} - 1$ is divisible by 31]

Soln: $a = 2, p = 31$
 $(2, 31) = 1$

By Fermat's theorem
 $a^{p-1} \equiv 1 \pmod{p}$

$$2^{30} \equiv 1 \pmod{31}$$

$$(2^{30})^{11} \equiv 1^{11} \pmod{31}$$

$$2^{330} \equiv 1 \pmod{31} \quad \& \quad 2^{10} \equiv 1 \pmod{31}$$

$$\therefore 2^{340} \equiv 1 \pmod{31} \Rightarrow 2^{340} - 1 \equiv 0 \pmod{31}$$

19) Find $7^{121} \pmod{13}$

Soln: $13 \rightarrow$ prime, $(13, 7) = 1$
 $7^{12} \equiv 1 \pmod{13}$ [Using Fermat's theorem]

$$(7^{12})^{10} \equiv 1^{10} \pmod{13}$$

$$\checkmark 7^{120} \equiv 1 \pmod{13}$$

$$\checkmark 7 \equiv 7 \pmod{13}$$

$$7^{121} \equiv 7 \pmod{13}$$

$$\therefore 7^{121} \pmod{13} = 7$$

20) Find the remainder when 41^{75} is divided by 3.

Soln:- $a = 41$, $p = 3$, $(a, p) = 1$

By Fermat's thm,

$$a^{p-1} \equiv 1 \pmod{p}$$

$$41^2 \equiv 1 \pmod{3}$$

$$(41^2)^{37} \equiv 1^{37} \pmod{3}$$

$$41^{74} \equiv 1 \pmod{3} \text{ \& } 41 \equiv 2 \pmod{3}$$

$$41^{75} \equiv 2 \pmod{3}$$

$$\therefore \boxed{r = 2}$$

21) Find the remainder when 5^{11} is divisible by 7.

Soln:- $a = 5$, $p = 7$, $(5, 7) = 1$

By Fermat's thm,

$$a^{p-1} \equiv 1 \pmod{p}$$

$$5^6 \equiv 1 \pmod{7}$$

$$(5^6) \equiv 1 \pmod{7} \text{ \& } 5^5 \equiv 3 \pmod{7}$$

$$5^{11} \equiv 3 \pmod{7}$$

$$\therefore \boxed{r = 3}$$

29) S.T $5^{38} \equiv 4 \pmod{11}$

Soln? $(5, 11) = 1$, $a=5$, $p=11$

By Fermat's thm,

$$a^{p-1} \equiv 1 \pmod{p}$$

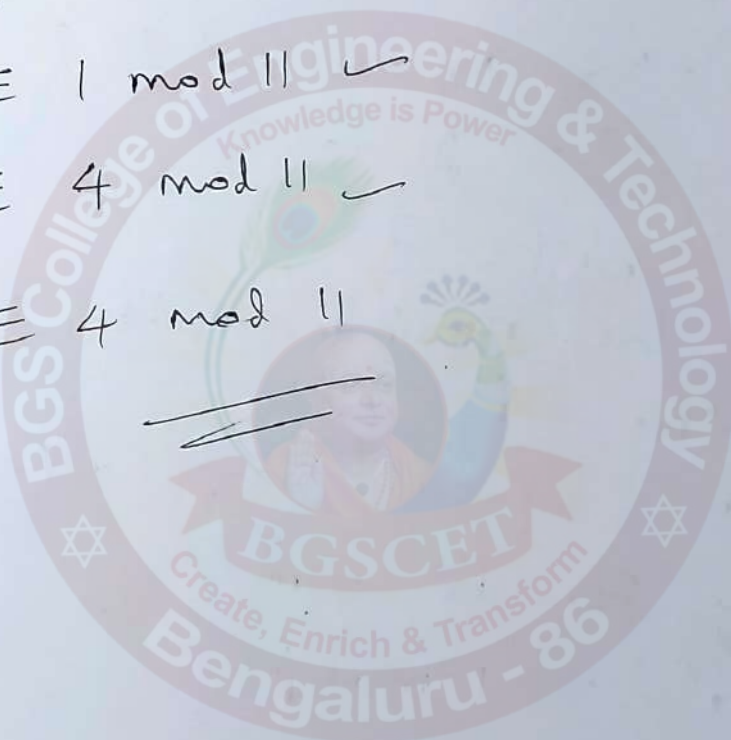
$$5^{10} \equiv 1 \pmod{11}$$

$$(5^{10})^3 \equiv 1^3 \pmod{11}$$

$$5^{30} \equiv 1 \pmod{11} \checkmark$$

$$5^8 \equiv 4 \pmod{11} \checkmark$$

$$5^{38} \equiv 4 \pmod{11}$$



Wilson's Theorem:

If p is a prime no then
 $(p-1)! \equiv -1 \pmod{p}$
 or

Difference = 1 $\Rightarrow (p-1)! \equiv (p-1) \pmod{p}$
 or

Difference = 2 $\Rightarrow (p-2)! \equiv 1 \pmod{p}$
 or

Difference = 3 $\Rightarrow (p-3)! \equiv \frac{(p-1)}{2} \pmod{p}$

Problems:-

1) Find the remainder when $15!$ is divided by 17.

Soln:-

$$p = 17 \quad 15! = (17-2)!$$

$$(p-2)! \equiv 1 \pmod{p}$$

$$15! \equiv 1 \pmod{17}$$

$$\therefore \text{Remainder} = 1.$$

2) S.T $4(29)! + 5!$ is divisible by 31.

Soln:- $p = 31, \quad 29! = (31-2)!$

$$(p-2)! \equiv 1 \pmod{p}$$

$$29! \equiv 1 \pmod{31}$$

$$4(29)! \equiv 4 \pmod{31}$$

$$4(29)! + 5! \equiv 4 + 5! \pmod{31}$$

$$4(29)! + 5! \equiv 124 \pmod{31}$$

$$4(29)! + 5! \equiv 0 \pmod{31}$$

$\therefore 4(29)! + 5!$ is divisible by 31.

3) Find the remainder when $2(26)!$ is divided by 29.

Soln:- $p = 29, \quad 26! = (29-3)!$

$$(p-3)! \equiv \frac{p-1}{2} \pmod{p}$$

$$26! \equiv 14 \pmod{29}$$

$$2(26)! \equiv 28 \pmod{29}$$

$\therefore$ Remainder is 28.

4) Find the remainder when $14!$ is divided by 17.

Soln:- $p = 17, \quad 14! = (17-3)!$

$$(p-3)! = \frac{p-1}{2} \pmod{p}$$

$$14! = 8 \pmod{17}$$

$$R = 8$$

5) If $p = 11$, s.t. $10! + 1$ is divisible by 11.

Soln:- $p = 11, \quad 10! = (11-1)!$

$$(p-1)! \equiv -1 \pmod{p}$$

$$10! \equiv -1 \pmod{11}$$

$$10! + 1 \equiv 0 \pmod{11}$$

$\therefore 10! + 1$ is divisible by 11.

6) Find the least positive remainder when $2(28!)$ is divided by 31.

Soln:- $p = 31, \quad 28! = (31-3)!$

$$(p-3)! = \frac{p-1}{2} \pmod{p}$$

$$28! \equiv 15 \pmod{31}$$

$$2(28)! \equiv 30 \pmod{31}$$

$$R = 30$$

Linear Diophantine equation

An eqn of the form $ax+by=c$ where $a \neq 0$, $b \neq 0$ & c is an integer is called a linear Diophantine eqn in 2 variables x & y , provided its solution is always integers.

Ex:- $2x + 3y = 12$ is an LDE.

Note:-

- 1) The LDE $ax+by=c$ has a sol'n iff $d|c$ where $d = (a, b)$.
- 2) A pair (x_0, y_0) is a sol'n of LDE $ax+by=c$ iff $ax_0+by_0=c$.
The general sol'n is given by
 $x = x_0 + \frac{b}{d}t$ & $y = y_0 - \frac{a}{d}t$, where t is an integer & $d = (a, b)$.

Problems:-

- 1) Which of the following Diophantine eqns can be solved?

a) $6x + 51y = 22$

Soln:- $(6, 51) = 3 = d$
 $d \nmid c = 3 \nmid 22$

$\therefore$ It doesn't have sol'n

$$b) 33x + 14y = 115$$

Soln - $(33, 14) = 1$

$$d | c = 1 | 115$$

$\therefore$ It has a soln.

$$c) 14x + 35y = 93$$

$$(14, 35) = 7$$

$$7 \nmid 93$$

It cannot be solved

2) Find the general soln of the eqn
 $70x + 112y = 168$, $a=70, b=112, d=14$

Soln -

$$(70, 112) = 14$$

$$14 \mid 168$$

$$168 = 70x + 112y \rightarrow (1)$$

$$\begin{array}{r} 70 \mid 112 \quad (1) \\ \underline{70} \\ 42 \end{array}$$

$$42 = 112 - 1(70)$$

$$28 = 70 - 1(42)$$

$$14 = 42 - 1(28)$$

$$\begin{array}{r} 42 \mid 70 \quad (1) \\ \underline{42} \\ 28 \end{array}$$

$$14 = [112 - 1(70)] - 1[70 - 1(42)]$$

$$14 = 112 - 2(70) + 1[112 - 1(70)]$$

$$\begin{array}{r} 28 \mid 42 \quad (1) \\ \underline{28} \\ 14 \end{array}$$

$$14 = 70(-3) + 112(2) \rightarrow (2)$$

$$(2) \times 12 \Rightarrow$$

$$\begin{array}{r} 14 \mid 28 \quad (2) \\ \underline{28} \\ 0 \end{array}$$

$$168 = 70(-36) + 112(24)$$

$$\therefore x_0 = -36 \text{ \& } y_0 = 24$$

$$\text{Gr. Soln} \Rightarrow x = x_0 + \frac{b}{d}t, \quad y = y_0 - \frac{a}{d}t$$

$$x = -36 + 8t \text{ \& } y = 24 - 5t$$

3) Find the general solution of the eqn
 $39x - 56y = 11$.

Soln: $(39, 56) = 1$

$$\begin{array}{r} 39 \overline{) 56} \quad (1 \\ \underline{39} \\ 17 \end{array}$$

$$\begin{array}{r} 17 \overline{) 39} \quad (2 \\ \underline{34} \\ 5 \end{array}$$

$$\begin{array}{r} 5 \overline{) 17} \quad (3 \\ \underline{15} \\ 2 \end{array}$$

$$\begin{array}{r} 2 \overline{) 5} \quad (2 \\ \underline{4} \\ 1 \end{array}$$

$$\begin{array}{r} 1 \overline{) 2} \quad (2 \\ \underline{2} \\ 0 \end{array}$$

$$11 = 39x - 56y \rightarrow (1)$$

$$17 = 56 - 1(39)$$

$$5 = 39 - 2(17)$$

$$2 = 17 - 3(5)$$

$$1 = 5 - 2(2)$$

$$1 = [39 - 2(17)] - 2[17 - 3(5)]$$

$$1 = 39 - 4[56 - 1(39)] + 6[39 - 2(17)]$$

$$1 = 17(39) - 4(56) - 12(17)$$

$$= 11(39) - 4(56) - 12[56 - 1(39)]$$

$$1 = 39(23) - 16(56) \rightarrow (2)$$

$$(2) \times 11 \Rightarrow$$

$$11 = 39(253) - 56(176)$$

$$\boxed{x_0 = 253} \quad , \quad \boxed{y_0 = 176}$$

$\therefore$ General soln is given by

$$x = x_0 + \frac{b}{d}t$$

$$y = y_0 - \frac{a}{d}t$$

$$\boxed{x = 253 - 56t}$$

$$\boxed{y = 176 - 39t}$$

Where 't' is an integer

4) Find all the ^{positive} solution of LDE
 $172x + 20y = 1000$

Soln: $(172, 20) = 4, 4 | 1000 \checkmark \rightarrow$ soln exists.
 $1000 = 172x + 20y \rightarrow (1)$

$$\begin{array}{r} 20 \overline{) 172} \\ \underline{160} \\ 12 \end{array}$$

$$12 = 172 - 8(20)$$

$$8 = 20 - 1(12)$$

$$4 = 12 - 1(8)$$

$$\begin{array}{r} 12 \overline{) 20} \\ \underline{12} \\ 8 \end{array}$$

$$4 = 12 - 1[20 - 1(12)]$$

$$\begin{array}{r} 8 \overline{) 12} \\ \underline{8} \\ 4 \end{array}$$

$$= 2(12) - 1(20)$$

$$= 2[172 - 8(20)] - 1(20)$$

$$4 = 172(2) + 20(-17) \rightarrow (2)$$

$$\begin{array}{r} 4 \overline{) 8} \\ \underline{8} \\ 0 \end{array}$$

$$(4) \times 250$$

$$1000 = 172(500) + 20(-4250)$$

$$\therefore x_0 = 500 \text{ \& } y_0 = -4250$$

G.Soln $\Rightarrow x = x_0 + \frac{b}{d}t, y = y_0 - \frac{a}{d}t$

$$x = 500 + \frac{20}{4}t$$

$$\text{ \& } y = -4250 - \frac{172}{4}t$$

$$\boxed{x = 500 + 5t} \text{ \& } \boxed{y = -4250 - 43t}$$

$$500 + 5t > 0$$

$$t > -100$$

$$-4250 - 43t > 0$$

$$t < -98.8$$

$$-100 < t < -98.8$$

$$\therefore t = -99 \text{ [} \because t \text{ is an integer]} \text{]}$$

$$x = 500 + 5(-99) \text{ \& } y = -4250 - 43(-99)$$

$$\therefore \boxed{x = 5}$$

$$\boxed{y = 7}$$

System of Linear Congruence

The system of Linear congruence

$$ax + by \equiv r \pmod{n}$$

$$cx + dy \equiv s \pmod{n}$$

has a unique solution modulo 'n' whenever $(ad - bc \downarrow n) = 1$.

*** Model QP

1) Solve, $2x + 6y \equiv 1 \pmod{7} \rightarrow (1)$

$$4x + 3y \equiv 2 \pmod{7} \rightarrow (2)$$

Soln:- $|ad - bc| = 18$ & $(18, 7) = 1$

$\therefore$ The given system has unique soln

$$(1) \times (2) \Rightarrow 4x + 12y \equiv 2 \pmod{7}$$

$$4x + 3y \equiv 2 \pmod{7}$$

$$\begin{array}{r} (-) \quad (-) \quad (-) \\ \hline 9y \equiv 0 \pmod{7} \end{array}$$

$\boxed{y = 0}$ $[7 | 9y \Rightarrow 9y \text{ should be multiple of } 7 \Rightarrow y = 0]$

$$(1) \Rightarrow 2x \equiv 1 \pmod{7}$$

$$7 \nmid 2x - 1$$

$$\Rightarrow 2x - 1 = 7K$$

$$2x = 7K + 1$$

$$x = \frac{7K + 1}{2}$$

$$K = 1 \Rightarrow \boxed{x = 4}$$

2) ^{solve} $7x + 3y \equiv 10 \pmod{16} \rightarrow (1)$

$2x + 5y \equiv 9 \pmod{16} \rightarrow (2)$

Soln: $|ad - bc| = 29 \quad \& \quad (29, 16) = 1$

$\therefore$ The given system has unique soln

$(1) \times 5 \Rightarrow 35x + 15y \equiv 50 \pmod{16}$

$(2) \times 3 \Rightarrow 6x + 15y \equiv 27 \pmod{16}$

$\begin{array}{r} (-) \quad (-) \quad (-) \\ \hline 29x \equiv 23 \pmod{16} \end{array}$

$29x \equiv 7 \pmod{16}$

$16 \mid 29x - 7$

$29x - 7 = 16K$

$\boxed{x=3}$, by inspection.

$2x + 5y \equiv 9 \pmod{16}$

$6 + 5y \equiv 9 \pmod{16}$

$5y \equiv 3 \pmod{16}$

$16 \mid 5y - 3$

$5y - 3 = 16K$

By inspection $\boxed{y=7}$

$$3) \text{ Solve } 3x + 4y \equiv 5 \pmod{13} \rightarrow (1)$$

$$2x + 5y \equiv 7 \pmod{13} \rightarrow (2)$$

Soln:- $|ad - bc| = 7 \quad \& \cdot (7, 13) = 1$

$\therefore$ The given system has unique soln.

(1) $\times 2$ & (2) $\times 3$ gives,

$$6x + 8y \equiv 10 \pmod{13}$$

$$\begin{array}{r} 6x + 15y \equiv 21 \pmod{13} \\ (-) \quad (-) \quad (-) \end{array}$$

$$-7y \equiv -11 \pmod{13}$$

$$7y \equiv 11 \pmod{13}$$

$$13 \mid 7y - 11$$

$$\Rightarrow 7y - 11 = 13K$$

$$y = \frac{13K + 11}{7}$$

By inspection, $K = 4$

$$\therefore \boxed{y = 9}$$

$$2x + 5y \equiv 7 \pmod{13}$$

$$2x + 45 \equiv 7 \pmod{13}$$

$$13 \mid 2x + 45 - 7$$

$$13 \mid 2x + 38$$

$$2x + 38 = 13K$$

$$x = \frac{13K - 38}{2}$$

By inspection, $K = 3$

$$\therefore \boxed{x = 7}$$

Polynomial Congruence

$$f(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_n \equiv 0 \pmod{m}$$

where $n \geq 1$ & $a_0, a_1, \dots, a_n \in \mathbb{Z}$

Roots of $f(x) \pmod{m}$

$x = c$ is called root of $f(x) \pmod{m}$
iff $f(c) \equiv 0 \pmod{m}$ i.e. iff $m \mid f(c)$

Problem:-

1) Find roots of the polynomial,
 $(x^2 + 2x - 3) \pmod{5}$ [$x^2 + 2x - 3 \equiv 0 \pmod{5}$
 $\pmod{5} \rightarrow \{0, 1, 2, 3, 4\}$
or $\{0, 1, 2, -2, -1\}$]

Soln:- $f(x) = x^2 + 2x - 3$

$$f(0) = -3 \not\equiv 0 \pmod{5}$$

$$f(1) = 0 \equiv 0 \pmod{5} \checkmark$$

$$f(2) = 5 \equiv 0 \pmod{5} \checkmark$$

$$f(3) = 12 \not\equiv 0 \pmod{5}$$

$$f(4) = 21 \not\equiv 0 \pmod{5}$$

$\therefore$ Roots are $x = 1$ & $x = 2$

2) Solve $x^3 + 5x + 1 \equiv 0 \pmod{27}$

Soln:- $f(x) = x^3 + 5x + 1$

$$27 = 3^3$$

$$\pmod{3} \rightarrow \{0, 1, 2\}$$

$$f(0) = 1 \not\equiv 0 \pmod{3}$$

$$f(1) = 7 \not\equiv 0 \pmod{3}$$

$$f(2) = 19 \not\equiv 0 \pmod{3}$$

$\therefore$ It has no solution

3) Find all the roots of the congruence
 $x^2 + x + 7 \equiv 0 \pmod{15}$

Soln:-

$$15 = 3 \times 5 \rightarrow 0, 1, 2, 3, 4 \\ \downarrow \\ 0, 1, 2$$

$$f(x) = x^2 + x + 7$$

$$f(0) = 7 \not\equiv 0 \pmod{5}$$

$$f(1) = 9 \not\equiv 0 \pmod{5}$$

$$f(2) = 13 \not\equiv 0 \pmod{5}$$

$$f(3) = 19 \not\equiv 0 \pmod{5} \Rightarrow f(x) \text{ has no solution}$$

$$f(4) = 27 \not\equiv 0 \pmod{5}$$

4)

Solve $x^5 - 3x^2 + 2 \equiv 0 \pmod{7}$

Soln:-

$$7 \rightarrow \{0, 1, 2, 3, 4, 5, 6\}$$

or

$$\{0, 1, 2, 3, -3, -2, -1\}$$

$$f(0) = 2 \not\equiv 0 \pmod{7}, f(-3) = -268 \not\equiv 0 \pmod{7}$$

$$f(1) = 0 \equiv 0 \pmod{7}, f(-2) = -42 \equiv 0 \pmod{7}$$

$$f(2) = 22 \not\equiv 0 \pmod{7}, f(-1) = -2 \not\equiv 0 \pmod{7}$$

$$f(3) = 218 \not\equiv 0 \pmod{7},$$

$$\therefore x = 1, -2 \pmod{7}$$

$$\text{i.e. } x = 1, 5$$

5) Find all the roots of the congruence $x^3 + 2x - 3 \equiv 0 \pmod{9}$

Soln: Mod 9 $\rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$

$$f(x) = x^3 + 2x - 3$$

$$f(0) \equiv -3 \not\equiv 0 \pmod{9}$$

$$f(1) \equiv 0 \equiv 0 \pmod{9}, \quad f(6) \equiv 225 \equiv 0 \pmod{9}$$

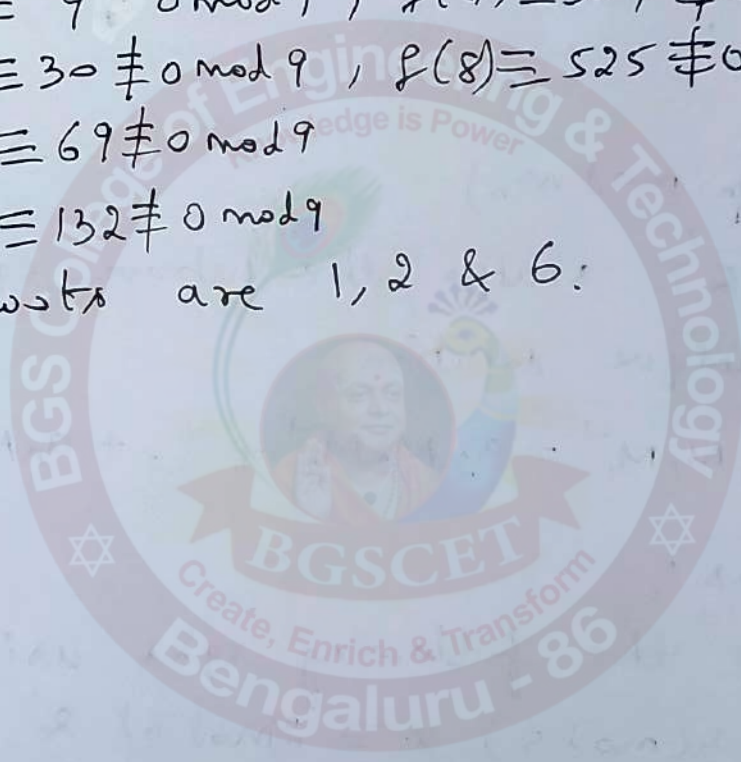
$$f(2) \equiv 9 \equiv 0 \pmod{9}, \quad f(7) \equiv 354 \not\equiv 0 \pmod{9}$$

$$f(3) \equiv 30 \not\equiv 0 \pmod{9}, \quad f(8) \equiv 525 \not\equiv 0 \pmod{9}$$

$$f(4) \equiv 69 \not\equiv 0 \pmod{9}$$

$$f(5) \equiv 132 \not\equiv 0 \pmod{9}$$

$\therefore$ Roots are 1, 2 & 6.



The Chinese Remainder Theorem

This theorem is used to solve a set of different congruent eqns with one variable but different moduli which are relatively prime as shown below,

$$x \equiv a_1 \pmod{m_1}$$

$$x \equiv a_2 \pmod{m_2}$$

⋮

$$x \equiv a_n \pmod{m_n}$$

$$\left[\begin{array}{l} \text{Here } M = m_1 \cdot m_2 \cdot m_3 \cdots m_n \\ M_1 = \frac{M}{m_1}, M_2 = \frac{M}{m_2} \cdots \end{array} \right]$$

CRT states the above eqns have a unique solution,

$$x = (a_1 M_1 M_1^{-1} + a_2 M_2 M_2^{-1} + \cdots + a_n M_n M_n^{-1}) \pmod{M}$$

Problems:-

1) ~~***~~ Model QP Solve the following eqns using CRT

$$x \equiv 3 \pmod{5} \quad x \equiv 2 \pmod{6} \quad \&$$

$$x \equiv 4 \pmod{7}$$

Soln:- $a_1 = 3, a_2 = 2, a_3 = 4$

$$m_1 = 5, m_2 = 6, m_3 = 7$$

$$\text{Clearly } (m_1, m_2) = (m_2, m_3) = (m_1, m_3) = 1$$

$\therefore$ Solution exists as m_1, m_2, m_3 are pairwise prime

$$M = m_1 \cdot m_2 \cdot m_3 = 5 \cdot 6 \cdot 7 = 210$$

$$M_1 = \frac{M}{m_1} = \frac{210}{5} = 42, \quad M_2 = \frac{210}{6} = 35$$

$$M_3 = \frac{210}{7} = 30.$$

To find inverses

$$M_1 \times M_1^{-1} \equiv 1 \pmod{m_1}$$

$$42 \times M_1^{-1} \equiv 1 \pmod{5}$$

By inspection $M_1^{-1} = 3$ $\left[\because 126 \equiv 1 \pmod{5} \right]$

$$M_2 \times M_2^{-1} \equiv 1 \pmod{m_2}$$

$$35 \times M_2^{-1} \equiv 1 \pmod{6}$$

By inspection $M_2^{-1} = 5$

$$M_3 \times M_3^{-1} \equiv 1 \pmod{m_3}$$

$$30 \times M_3^{-1} \equiv 1 \pmod{7}$$

By inspection $M_3^{-1} = 4$

$\therefore$ Soln is

$$x = [a_1 M_1 M_1^{-1} + a_2 M_2 M_2^{-1} + a_3 M_3 M_3^{-1}] \pmod{M}$$

$$= [(3 \times 42 \times 3) + (2 \times 35 \times 5) + (4 \times 30 \times 4)] \pmod{210}$$

$$x = 1208 \pmod{210}$$

$$x = 158 \pmod{210}$$

2) Solve the following eqns using CRT
 $x \equiv 2 \pmod{3}$
 $x \equiv 3 \pmod{5}$ OR
 $x \equiv 2 \pmod{7}$

[Find a no having remainder 2, 3, 2 when divided by 3, 5 & 7 respectively]

Soln: $a_1 = 2, a_2 = 3, a_3 = 2$
 $m_1 = 3, m_2 = 5, m_3 = 7$

clearly $(m_1, m_2) = (m_2, m_3) = (m_1, m_3) = 1$

$\therefore$ soln exists

$M = m_1 \cdot m_2 \cdot m_3 = 105$

$M_1 = \frac{M}{m_1} = 35, M_2 = \frac{M}{m_2} = 21, M_3 = \frac{M}{m_3} = 15$

To find inverses,

$M_1 \times M_1^{-1} \equiv 1 \pmod{m_1}; M_2 \times M_2^{-1} \equiv 1 \pmod{m_2}$
 $35 \times M_1^{-1} \equiv 1 \pmod{3} \quad 21 \times M_2^{-1} \equiv 1 \pmod{5}$

$M_1^{-1} = 2$

$M_2^{-1} = 4$

$M_3 \times M_3^{-1} \equiv 1 \pmod{m_3}$

$15 \times M_3^{-1} \equiv 1 \pmod{7}$

$M_3^{-1} = 1$

$\therefore$ soln is given by

$$x = (a_1 M_1 M_1^{-1} + a_2 M_2 M_2^{-1} + a_3 M_3 M_3^{-1}) \pmod{M}$$

$$= (2 \times 35 \times 2 + 3 \times 21 \times 1 + 2 \times 15 \times 1) \pmod{105}$$

$$= 233 \pmod{105}$$

$x = 23$

3) Solve the following eqns using CRT
 $x \equiv 5 \pmod{3}$, $x \equiv 2 \pmod{5}$ & $x \equiv 1 \pmod{11}$

Soln: $a_1 = 5$, $a_2 = 2$, $a_3 = 1$

$m_1 = 3$, $m_2 = 5$, $m_3 = 11$

Clearly $(m_1, m_2) = (m_2, m_3) = (m_1, m_3) = 1$

∴ soln exists

$M = m_1 \cdot m_2 \cdot m_3 = 165$

$M_1 = \frac{M}{m_1} = 55$, $M_2 = \frac{M}{m_2} = 33$, $M_3 = \frac{M}{m_3} = 15$

$M_1 \times M_1^{-1} \equiv 1 \pmod{m_1}$; $M_2 \times M_2^{-1} \equiv 1 \pmod{m_2}$
 $55 \times M_1^{-1} \equiv 1 \pmod{3}$ $M_2^{-1} \times 33 \equiv 1 \pmod{5}$

$M_1^{-1} = 1$

$M_2^{-1} = 2$

$M_3 \times M_3^{-1} \equiv 1 \pmod{m_3}$

$15 \times M_3^{-1} \equiv 1 \pmod{11} \Rightarrow M_3^{-1} = 3$

Soln is,

$x \equiv (a_1 M_1 M_1^{-1} + a_2 M_2 M_2^{-1} + a_3 M_3 M_3^{-1}) \pmod{M}$

$\equiv (5 \times 55 \times 1 + 2 \times 33 \times 2 + 1 \times 15 \times 3) \pmod{165}$

$\equiv 452 \pmod{165}$

$x = 122$

RSA Algorithm

Ron Rivest, Adi Shamir & Leonard Adleman
in 1978.

It is an asymmetric cryptographic algorithm
(2 keys) i.e. public & private key.

Public key → Known to all users in network.

Private key → Kept secret, not sharable to all.

If public key of user 'A' is used for encryption, private key of same user should be used for decryption.

The RSA scheme is a block cipher (secret or code) in which the plain text & cipher text are integers b/w 0 & $n-1$ for some 'n'.

① Key Generation [RSA Algorithm]

- (i) Select 2 large prime numbers 'p' & 'q'.
for higher security
- (ii) Calculate $n = p \times q$
- (iii) Calculate $\phi(n) = (p-1) \times (q-1)$
- (iv) Choose value of 'e' s.t. $1 < e < \phi(n)$ & $(\phi(n), e) = 1$.
- (v) Calculate, $d \equiv e^{-1} \pmod{\phi(n)}$
 $ed \equiv 1 \pmod{\phi(n)}$
- (vi) Public Key = $\{e, n\}$ → used in encryption
- (vii) Private Key = $\{d, n\}$ → used in decryption.

② Encryption

$$C \equiv M^e \pmod{n}$$

Plain text = $M < n$

$C \rightarrow$ Cipher text.

③ Decryption

$$M \equiv C^d \pmod{n}$$

Problem

1) In RSA algorithm, if $p=7$, $q=11$ & $e=13$ then what will be the value of d ?

Soln:- $p=7$, $q=11$

$$n = p \cdot q = 77, \quad \phi(n) = (p-1)(q-1)$$

$$\phi(77) = 6 \times 10 = 60$$

$$\text{Given } e = 13 \quad [1 < 13 < 60, (13, 60) = 1]$$

To find 'd'

$$d \equiv e^{-1} \pmod{\phi(n)}$$

$$ed \equiv 1 \pmod{\phi(n)}$$

$$13d \equiv 1 \pmod{60}$$

$$60 \mid 13d - 1$$

$$\Rightarrow 13d - 1 = 60k$$

$$d = \frac{60k + 1}{13}$$

By inspection,

$$k = 8$$

$$\boxed{d = 37}$$

$\therefore$ public key =

$$\{e, n\} = \{13, 77\}$$

private key =

$$\{d, n\} = \{37, 77\}$$

Inverse method

$\Rightarrow$ Alternative method

$$13d \equiv 1 \pmod{60} \rightarrow (*)$$

$$13 \mid 60(4) \rightarrow 8 = 60 - 4(13)$$

$$52 \rightarrow 5 = 13 - 1(8)$$

$$8 \mid 13(1) \rightarrow 3 = 8 - 1(5)$$

$$8 \rightarrow 2 = 5 - 1(3)$$

$$5 \mid 8(1) \rightarrow 1 = 3 - 1(2)$$

$$5 \rightarrow 1 = [8 - 1(5)] - 1[5 - 1(3)]$$

$$3 \mid 5(1) \Rightarrow 1 = [60 - 4(13)] - 2[13 - 1(8)] + [8 - 1(5)]$$

$$2 \mid 3(1) \Rightarrow 1 = 60 - 6(13) + 3(8) + 1(5)$$

$$GCD \leftarrow 1 \mid 2(2) \Rightarrow 1 = 60 - 6(13) + 3[60 - 4(13)] + 1[13 - 1(8)]$$

$$2 \Rightarrow 1 = 4(60) - 19(13) + [60 - 4(13)]$$

$$1 = 60(5) + 13(-23)$$

$$1 \bmod 60 \equiv 60(5) \bmod 60 + 13(-23) \bmod 60$$

$$1 \equiv 0 + 13(37) \bmod 60$$

$$13(37) \equiv 1 \bmod 60$$

Comparing with $(*)$, $d=37$

- 2) ~~xxxx~~ Using RSA algorithm, find public key & private key w.r.t $p=3$, $q=11$ & $M=31$. Also Encrypt & decrypt the plain text M.

0	1, 2, 3	25
A, B, C, D	...	Z

Soln:- $n = pq = 33$

$$\phi(n) = (p-1)(q-1) = 2 \times 10 = 20$$

Choose 'e' s.t. $1 < e < \phi(n)$ & $(\phi(n), e) = 1$

'e' can be taken as any one of these no's $\rightarrow 3, 7, 9, 11, 13, 17, 19$.

$$d \equiv e^{-1} \bmod [\phi(n)]$$

$$ed \equiv 1 \bmod 20 \Rightarrow 20 \mid 7d - 1$$

$$7d \equiv 1 \bmod 20$$

By inspection, $d=3$

$$\text{Public Key} = \{e, n\} = \{7, 33\}$$

$$\text{Private Key} = \{d, n\} = \{3, 33\}$$

Encryption: $C \equiv M^e \bmod n$, $M=31 = DB$
 $C \equiv (31)^7 \bmod 33 \rightarrow 1$

$$(31)^5 \equiv 1 \bmod 33 \text{ \& \& } (31)^2 \equiv 4 \bmod 33$$

$$\therefore (31)^7 \equiv 4 \bmod 33$$

$$4 \equiv (31)^7 \pmod{33} \rightarrow (2)$$

$$(1) \& (2) \Rightarrow C = 4 = 04$$

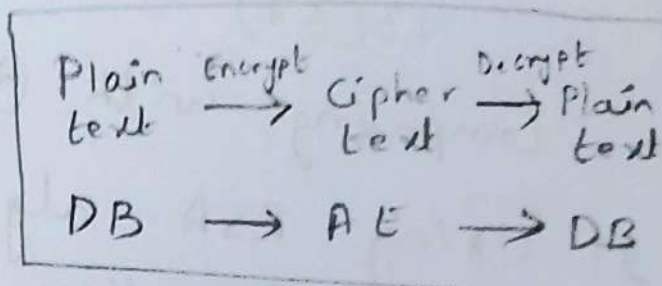
$$C = AE$$

Decryption

$$\begin{aligned} M &\equiv c^d \pmod{n} \\ &\equiv 4^3 \pmod{33} \\ &\equiv 64 \pmod{33} \end{aligned}$$

$$M \equiv 31 \pmod{33}$$

$$M = 31 = DB$$



3) Encode STOP using RSA algorithm with key $(2537, 13)$ & $p=43, q=59$

Soln: Public key = $(2537, 13)$
 n, e

$$\begin{aligned} n = pq &= 2537, \quad \phi(n) = \phi(43 \times 59) \\ &= 42 \times 58 = 2436 \end{aligned}$$

$$\text{Given } e=13, \quad 1 < e < 2436 \quad \& \quad (2436, 13) = 1$$

$$M = \text{STOP} = 18191415$$

$$\begin{aligned} \text{Let } M_1 &= 1819 \quad \& \quad 2 \\ M_2 &= 1415 \end{aligned}$$

$$\begin{bmatrix} 0, 1 & 14, 15 \\ A, B & 0, P \\ 18, 19 & 25 \\ S, T & 2 \end{bmatrix}$$

Encryption:

$$C \equiv M^e \pmod{n}$$

$$C_1 \equiv M_1^e \pmod{n}$$

$$C_1 \equiv (1819)^{13} \pmod{2537} \rightarrow (1)$$

$$(1819)^3 \equiv 2068 \pmod{2537} \rightarrow (2)$$

$$[(1819)^3]^3 \equiv (2068)^3 \pmod{2537}$$

$$(1819)^9 \equiv 322 \pmod{2537} \rightarrow (3)$$

$$(1819)^1 \equiv 1819 \pmod{2537} \rightarrow (4)$$

(2), (3) & (4) gives

$$(1819)^{13} \equiv (2068 \times 322 \times 1819) \pmod{2537}$$

$$(1819)^{13} \equiv 2081 \pmod{2537}$$

$$2081 \equiv (1819)^{13} \pmod{2537} \rightarrow (5)$$

$$(1) \text{ \& } (5) \Rightarrow \boxed{C_1 = 2081}$$

$$C_2 \equiv M_2^e \pmod{n}$$

$$C_2 \equiv (1415)^{13} \pmod{2537} \rightarrow (6)$$

$$(1415)^3 \equiv 1828 \pmod{2537} \rightarrow (7)$$

$$[(1415)^3]^3 \equiv (1828)^3 \pmod{2537}$$

$$(1415)^9 \equiv 2005 \pmod{2537} \rightarrow (8)$$

$$(1415)^1 \equiv 1415 \pmod{2537} \rightarrow (9)$$

(7), (8) & (9) gives

$$(1415)^{13} \equiv (1828 \times 2005 \times 1415) \pmod{2537}$$

$$(1415)^{13} \equiv 2182 \pmod{2537}$$

$$2182 \equiv (1415)^{13} \pmod{2537} \rightarrow (10)$$

$$(6) \& (10) \Rightarrow C_2 = 2182$$

$$\therefore C = C_1 C_2$$

$$= \boxed{2081} \boxed{2182}$$

$$C = U I B V I C$$

Curve Fitting

Curve fitting is a method of finding a specific relation connecting the dependent & independent variables for a given data so as to satisfy the data as accurately as possible.

Applications :- used for utilization of resources in industry, business management, economy planning, agriculture.

Fitting of a st line $y = ax + b$

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

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

(interchange the place of a & b in st form)

$$\begin{aligned} y &= a + bx \\ b \sum x^2 + a \sum x &= \sum xy \\ b \sum x + na &= \sum y \end{aligned}$$

These equations are called normal equations. (By solving this we get b & a)

By solving these we get values of a & b.

If fit a st line $y = ax + b$ for the foll data :-

x	1	3	4	6	8	9	11	14	→ Count these for values of n
y	1	2	4	4	5	7	8	9	

x	y	xy	x^2
1	1	1	1
3	2	6	9
4	4	16	16
6	4	24	36
8	5	40	64
9	7	63	81
11	8	88	121
14	9	126	196
$\sum x = 56$	$\sum y = 40$	$\sum xy = 364$	$\sum x^2 = 524$

$$\underline{n = 8}$$

The normal eqⁿ for $y = ax + b$ are
 $a \sum x^2 + b \sum x = \sum xy$
 $a \sum x + nb = \sum y$
 $a 524 + b 56 = 364$
 $a 56 + 8b = 40$

In calculator goto
 Equation
 ↓
 unknowns
 ↓
 2
 ↓
 enter values
 ↓
 get a & b.

On solving $a = 0.64$
 $b = 0.55$

$$\therefore y = 0.64x + 0.55$$

If find the equation of the best fitting st line for the following data & hence estimate the value of the dependent variable corresponding to the value 30 of the independent variable.

Ind x	5	10	15	20	25
Dep y	16	19	23	26	30

You can either consider $y = ax + b$ or $y = a + bx$. Since ^{type} not specified better consider $y = ax + b$

x	y	xy	x^2
5	16	80	25
10	19	190	100
15	23	345	225
20	26	520	400
25	30	750	625
$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$
$= 75$	$= 114$	$= 1885$	$= 1375$

$$\underline{n = 5}$$

The normal eqⁿ for $y = ax + b$ are

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

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

$$1375a + 75b = 1885$$

$$75a + 5b = 114$$

On solving $a = 0.7$

$$b = 12.3$$

$$y = 0.7x + 12.3$$

At $x = 30$, $y = 0.7(30) + 12.3 = \underline{\underline{33.3}}$

3. A simply supported beam carries a concentrated load P at its midpoint corresponding to various values of P the maximum deflection Y is measured and is given in the foll table.

P	100	120	140	160	180	200
Y	0.45	0.55	0.60	0.70	0.80	0.85

Find a law of the form $Y = a + bP$ & hence estimate Y when P is 150.

$$\underline{Y = a + bP}$$

Here $y = Y$

$$a = a$$

$$b = b$$

$$\underline{x = P}$$

P	Y	PY	P^2
100	0.45	45	10000
120	0.55	66	14400
140	0.60	84	19600
160	0.70	112	25600
180	0.80	144	32400
200	0.85	170	40000
$\sum P$	$\sum Y$	$\sum PY$	$\sum P^2$
$= 900$	$= 3.95$	$= 621$	$= 142000$

$$\underline{n = 6}$$

The normal eqⁿ for $y = a + bx$ are

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

$$b \sum x + na = \sum y$$

Now, the normal eqⁿ for $Y = a + bP$ are

$$b \sum P^2 + a \sum P = \sum PY$$

$$b \sum P + na = \sum Y$$

$$b 142000 + a 900 = 621$$

$$b 900 + 6a = 3.95$$

On solving

$$b = 0.004, a = 0.048$$

$$\therefore Y = 0.047 + 0.004P$$

$$\text{At } P = 150, Y = \underline{\underline{0.648}}$$

4. Fit a straight line to the foll data

Year	1961	1971	1981	1991	2001
Production (in tons)	8	10	12	10	16

Also find the expected production in the year
1. 2006.

1. Let the year and production be represented by
2. the variables x & y . We can fit the st line
2. either in the form $y = a + bx$ or $y = ax + b$.

3. The normal eqⁿ for $y = a + bx$ are

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

$$b \sum x + na = \sum y$$

x	y	xy	x^2
1961	8	15688	3845521
1971	10	19710	3884841
1981	12	23772	3924361
1991	10	19910	3964081
2001	16	32016	4004001
Σx = 9905	Σy = 56	Σxy = 111096	Σx^2 = 19622805

$$\underline{\underline{n=5}}$$

$$b \mid 19622805 + a \mid 9905 = 111096$$

$$b \mid 9905 + 5a = 56$$

On solving,

$$\underline{\underline{a = -305.76}}$$

$$\underline{\underline{b = 0.16}}$$

$$y = -305.76 + 0.16x$$

$$\text{when } x = 2006, \quad \underline{\underline{y = 15.2 \text{ tons}}}$$

Fitting of a second degree parabola

$$y = ax^2 + bx + c$$

(Interchange place of a, b, c but not x² std for)

$$a \sum x^4 + b \sum x^3 + c \sum x^2 = \sum x^2 y$$

$$a \sum x^3 + b \sum x^2 + c \sum x = \sum xy$$

$$a \sum x^2 + b \sum x + nc = \sum y$$

These eqⁿ are called normal eqⁿ.

By solving these we get the value of a, b, c.

$$y = a + bx + cx^2$$

$$c \sum x^4 + b \sum x^3 + a \sum x^2 = \sum x^2 y$$

$$c \sum x^3 + b \sum x^2 + a \sum x = \sum xy$$

$$c \sum x^2 + b \sum x + na = \sum y$$

↓
(solving this we get c, b, a)

If fit a second degree parabola $y = ax^2 + bx + c$ in the least square sense & hence estimate y at x=6.

x	1	2	3	4	5
y	10	12	13	16	19

x	y	x ²	x ³	x ⁴	xy	x ² y
1	10	1	1	1	10	10
2	12	4	8	16	24	48
3	13	9	27	81	39	117
4	16	16	64	256	64	256
5	19	25	125	625	95	475
$\sum x$	$\sum y$	$\sum x^2$	$\sum x^3$	$\sum x^4$	$\sum xy$	$\sum x^2 y$
=15	=70	=55	=225	=979	=232	=906

$$n = 5$$

In calculator goto
Equation

↓

Unknowns

↓

3

↓

enter values

↓

get a, b, c

The normal equations are

$$a \sum x^4 + b \sum x^3 + c \sum x^2 = \sum x^2 y$$

$$a \sum x^3 + b \sum x^2 + c \sum x = \sum xy$$

$$a \sum x^2 + b \sum x + nc = \sum y$$

$$a 979 + b 225 + c 55 = 906$$

$$a 225 + b 55 + c 15 = 232$$

$$a 55 + b 15 + 5c = 70$$

On solving

$$a = 0.285 / 0.29$$

$$b = 0.485 / 0.49$$

$$c = 9.4 / 9.40$$

$$y = 0.285x^2 + 0.485x + 9.4 \quad | \quad 0.29x^2 + 0.49x + 9.40$$

At $x=6$, $y = 22.612 / 22.78$

if fit a parabola $y = a + bx + cx^2$ for

$$x \quad 0 \quad 1 \quad 2 \quad 3 \quad 4$$

$$y \quad 1 \quad 1.8 \quad 1.3 \quad 2.5 \quad 2.3$$

$$y \quad 1$$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	1.8	1	1	1	1.8	1.8
2	1.3	4	8	16	2.6	5.2
3	2.5	9	27	81	7.5	22.5
4	2.3	16	64	256	9.2	36.8
$\sum x$	$\sum y$	$\sum x^2$	$\sum x^3$	$\sum x^4$	$\sum xy$	$\sum x^2y$
= 10	= 8.9	= 30	= 100	= 354	= 21.1	= 66.3

$$\underline{n=5}$$

The normal eq are

$$c \sum x^4 + b \sum x^3 + a \sum x^2 = \sum x^2 y$$

$$c \sum x^3 + b \sum x^2 + a \sum x = \sum x y$$

$$c \sum x^2 + b \sum x + na = \sum y$$

$$c 354 + b 100 + a 30 = 66.3$$

$$c 100 + b 30 + a 10 = 21.1$$

$$c 30 + b 10 + 5a = 8.9$$

On solving,

$$c = -0.021 / 0.02$$

$$b = 0.415 / 0.42$$

$$a = 1.077 / 1.08$$

$$y = 1.077 + 0.415x - 0.021x^2 / 1.08 + (0.42)x - (0.02)x^2$$

3/ find the best values of a, b, c if the equation $y = a + bx + cx^2$ is to fit most closely to the foll observations.

$$x \quad -2 \quad -1 \quad 0 \quad 1 \quad 2$$

$$y \quad -3.150 \quad -1.390 \quad 0.620 \quad 2.880 \quad 5.378$$

x	y	x^2	x^3	x^4	xy	x^2y
-2	-3.150	4	-8	16	6.3	-12.6
-1	-1.390	1	-1	1	1.390	-1.390
0	0.620	0	0	0	0	0
1	2.880	1	1	1	2.880	2.880
2	5.378	4	8	16	10.756	21.512
$\sum x$	$\sum y$	$\sum x^2$	$\sum x^3$	$\sum x^4$	$\sum xy$	$\sum x^2y$
= 0	= 4.338	= 10	= 0	= 34	= 21.326	= 10.402

$$\underline{n=5}$$

The normal eqⁿ are

$$\begin{aligned} c \sum x^4 + b \sum x^3 + a \sum x^2 &= \sum x^2 y \\ c \sum x^3 + b \sum x^2 + a \sum x &= \sum x y \\ c \sum x^2 + b \sum x + na &= \sum y \end{aligned}$$

$$c 34 + b 0 + a 10 = 10 \cdot 402$$

$$c 0 + b 10 + a 0 = 21 \cdot 326$$

$$c 10 + b 0 + 5a = 4 \cdot 338$$

On solving

$$\underline{c} = 0.123$$

$$\underline{b} = 2.132$$

$$\underline{a} = 0.621$$

$$y = 0.621 + 2.132x + 0.123x^2$$

$a + bx = \text{---} cx^2$

Fitting of Exponential Curves (ae^{bx} , ax^b , ab^x)

If fit a curve of the form $y = ae^{bx}$ for the foll data :-

x	0	2	4
y	8.120	10	31.820

Let $y = ae^{bx}$
 $\log y = \log ae^{bx}$
 $\log y = \log a + \log e^{bx}$
 $\log y = \log a + bx (\log e) = 1$

$Y = A + bx$ where $Y = \log y$, $A = \log a$

The normal eqⁿ of $y = a + bx$ are

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

$$b \sum x + na = \sum y$$

here $y = Y$, $a = A$, $b = b$, $x = x$.

The normal eqⁿ are

$$b \sum x^2 + A \sum x = \sum xY$$

$$b \sum x + nA = \sum Y$$

ln(ln) on your calculator

x	y	$Y = \log_e y$	xY	x^2
0	8.12	2.094	0	0
2	10	2.303	4.606	4
4	31.82	3.460	13.84	16
$\sum x$ = 6		$\sum Y$ = 7.857	$\sum xY$ = 18.446	$\sum x^2$ = 20

$n = 3$

$$b \cdot 20 + A \cdot 6 = 18.446$$

$$b \cdot 6 + 3A = 7.857$$

On solving

x	y	$Y = \log_e y$	xy	x^2
1	87	$4.465/4.47$	4.47	1
2	97	$4.574/4.57$	9.14	4
3	113	$4.727/4.73$	14.19	9
4	129	$4.859/4.86$	19.44	16
5	202	$5.308/5.31$	26.55	25
6	195	$5.273/5.27$	31.62	36
7	193	$5.262/5.26$	36.82	49
Σx = 28		ΣY = 34.471 34.47	Σxy = 142.25	Σx^2 = 140

$$\underline{n = 7}$$

$$140B + 28A = 142.25$$

$$28B + 7A = 34.47$$

On solving,

$$B = 0.16$$

$$A = 4.3$$

$$A = \log_e a \Rightarrow a = e^A = e^{4.3} = 73.70$$

$$B = \log_e b \Rightarrow b = e^B = e^{0.156} = 1.17$$

$$\therefore y = (73.70)(1.17)^x$$

$$\text{At } x=8, \quad y = \underline{\underline{258.79}}$$



$$b = 0.342$$

$$A = 1.936$$

$$A = \log_e a \Rightarrow a = e^A = e^{1.936} = 6.932$$

$$\therefore y = ae^{bx} \\ = (6.932) e^{(0.342)x}$$

fit a curve of the form $y = ab^x$ for the data & hence find y when $x=8$.

x	1	2	3	4	5	6	7
y	87	97	113	129	202	195	193

$$\text{let } y = ab^x \\ \log y = \log ab^x \\ \log y = \log a + \log b^x \\ \log y = \log a + x \log b$$

$$Y = A + Bx \quad \text{where } Y = \log y, A = \log a, B = \log b$$

The normal eqⁿ of $y = a + bx$ are

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

$$b \sum x + na = \sum y$$

here $y = Y$, $a = A$, $b = B$, $x = x$

The normal eqⁿ are

$$B \sum x^2 + A \sum x = \sum xY$$

$$B \sum x + nA = \sum Y$$

3) Fit a least square geometric curve $y = ax^b$ for the data

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

Let $y = ax^b$ $\log y = \log ax^b$

$\log y = \log a + \log x^b$

$\log y = \log a + b \log x$

$Y = A + bX$

where $Y = \log y$, $A = \log a$, $b = b$, $X = \log x$.

The normal eqⁿ of $y = a + bx$ are

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

$b \sum x + na = \sum y$

here $y = Y$, $a = A$, $b = b$, $x = X$

The normal eqⁿ are

$b \sum X^2 + A \sum X = \sum XY$

$b \sum X + nA = \sum Y$

x	y	$X = \log x$	$Y = \log y$	XY	X^2
1	0.5	0	-0.693	0	0
2	2	0.693	0.693	0.480	0.480
3	4.5	1.098	1.504	1.651	1.206
4	8	1.386	2.079	2.881	1.921
5	12.5	1.609	2.525	4.063	2.589
		$\sum X = 4.787$	$\sum Y = 6.109$	$\sum XY = 9.080$	$\sum X^2 = 6.197$
		4.79	6.11	9.09	6.21

$n = 5$

$$b6 \cdot 199 + A4 \cdot 787 = 9 \cdot 080$$

$$b4 \cdot 787 + A5 = 6 \cdot 109$$

$$b6 \cdot 21 + A4 \cdot 79 = 9 \cdot 09$$

$$b4 \cdot 79 + 5A = 6 \cdot 11$$

$$b = 2$$

$$A = -0.69$$

On solving,

$$b = 2$$

$$A = -0.693$$

$$A = \log_e a \Rightarrow a = e^A = e^{-0.693} = 0.5$$

$$\therefore y = 0.5 x^2$$

If An experiment on lifetime 't' of cutting tool at different cutting speeds 'v' are given below:-

Speed (v)	350	400	500	600
Life (t)	61	26	7	2.6

fit a relation of the form $v = at^b$

$$\text{Let } v = at^b$$

$$\log v = \log at^b$$

$$\log v = \log a + \log t^b$$

$$\log v = \log a + b \log t$$

$$V = A + bT$$

where $V = \log v$, $A = \log a$, $b = b$, $T = \log t$

The normal equation of $y = a + bx$ are

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

$$b \sum x + na = \sum y$$

here $y = V$, $a = A$, $b = b$, $x = T$

$$b \sum T^2 + A \sum T = \sum VT$$

$$b \sum T + nA = \sum V$$

V	t	$V = \log_e v$	$T = \log_e t$	VT	T^2
350	61	5.857 5.86	4.111 4.11	24.078 24.08	16.9 16.89
400	26	5.991 5.99	3.258 3.26	19.519 19.52	10.615 10.63
500	7	6.214 6.21	1.946 1.95	12.092 12.11	3.787 3.80
600	2.6	6.396 6.40	0.956 0.96	6.115 6.12	0.914 0.92
		ΣV = 24.460 24.46	$\Sigma T = 10.270$ 10.28	ΣVT = 61.807 61.83	ΣT^2 = 32.214 32.24

$$\underline{n=4}$$

$$b 32.214 + A 10.270 = 61.807$$

$$b 10.270 + 4A = 24.460$$

$$b 32.24 + A 10.28 = 61.83$$

$$b 10.28 + 4A = 24.46$$

On solving,

$$b = -0.171$$

$$A = 6.554$$

$$b = -0.18$$

$$A = 6.57$$

$$A = \log_e a \Rightarrow a = e^A = e^{6.554} = \cancel{701.976} 706.27$$

$$v = (706.27)^{\underline{\underline{t}}} \quad (-0.18 \dots)$$

Correlation & Regression

Co-variation of 2 independent magnitudes is known as correlation.

The numerical measure of correlation b/w 2 variables x & y is known as Pearson's co-efficient of correlation denoted by ' r '.

The value of ' r ' does not exceed unity ($-1 \leq r \leq 1$)
If $r = \pm 1$ we say that x & y are perfectly correlated
and if $r = 0$ we say that x & y are non correlated.

Regression is an estimation of one independent variable in terms of the other. The best fitting straight line of the form $y = ax + b$ is called regression line of y on x & $x = ay + b$ is called regression line of x on y .

Type 1 :- First to find co-eff of correlation (r) & then lines of regression ($y = ax + b$) & ($x = ay + b$)

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y} \quad \text{--- (1)}$$

Identify ' n '

$$\bar{x} = \frac{\sum x}{n}, \quad \bar{y} = \frac{\sum y}{n}, \quad \bar{z} = \frac{\sum z}{n} \quad (\text{where } z = x - y)$$

$$\sigma_x^2 = \frac{\sum x^2}{n} - (\bar{x})^2, \quad \sigma_y^2 = \frac{\sum y^2}{n} - (\bar{y})^2, \quad \sigma_z^2 = \sigma_{x-y}^2 = \frac{\sum z^2}{n} - (\bar{z})^2$$

• sub in eq ① to get 'r'.

• lines of regressions are given by

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x}) \quad \& \quad x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$\downarrow y = ax + b$ $\downarrow x = ay + b$

• Fix to 2 decimal places.

Type 2 :- First lines of regression and then co-eff of correlation (r).

→ Identify 'n'

→ $\bar{x} = \frac{\sum x}{n}$, $\bar{y} = \frac{\sum y}{n}$

→ $X = x - \bar{x}$, $Y = y - \bar{y}$

→ lines of regression are given by

$$Y = \frac{\sum XY}{\sum X^2} X \quad , \quad X = \frac{\sum XY}{\sum Y^2} Y$$

→ $r = \sqrt{(\text{Co-eff of } x)(\text{Co-eff of } y)}$

→ Fix to 2 decimal places.

Type 1 - Table

x	y	$z = x - y$	x^2	y^2	z^2
$\sum x$	$\sum y$	$\sum z$	$\sum x^2$	$\sum y^2$	$\sum z^2$

Type 2 - Table

x	y	X	Y	XY	x^2	y^2
$\sum x$	$\sum y$			$\sum XY$	$\sum x^2$	$\sum y^2$

Type-1

4) Compute the co-eff of correlation^(r) and the equation of the lines of regression for the data

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

x	y	$z = x - y$	x^2	y^2	z^2
1	9	-8	1	81	64
2	8	-6	4	64	36
3	10	-7	9	100	49
4	12	-8	16	144	64
5	11	-6	25	121	36
6	13	-7	36	169	49
7	14	-7	49	196	49
$\sum x$ = 28	$\sum y$ = 77	$\sum z = -49$	$\sum x^2 = 140$	$\sum y^2 = 875$	$\sum z^2 = 347$

$n = 7$

$$\bar{x} = \frac{\sum x}{n} = \frac{28}{7} = 4, \quad \bar{y} = \frac{\sum y}{n} = \frac{77}{7} = 11, \quad \bar{z} = \frac{\sum z}{n} = \frac{-49}{7} = -7$$

$$\begin{aligned} \sigma_x^2 &= \frac{\sum x^2}{n} - (\bar{x})^2, & \sigma_y^2 &= \frac{\sum y^2}{n} - (\bar{y})^2, & \sigma_z^2 &= \frac{\sum z^2}{n} - (\bar{z})^2 \\ &= \frac{140}{7} - (4)^2, & &= \frac{875}{7} - (11)^2, & &= \frac{347}{7} - (-7)^2 \\ &= 4, & &= 4, & &= 0.57 \end{aligned}$$

$$\sigma_x = \sqrt{4} = 2, \quad \sigma_y = \sqrt{4} = 2$$

$$\therefore r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}$$

$$= \frac{4 + 4 - 0.57}{2(2)(2)} = 0.93$$

The lines of regression are

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

$$y - 11 = \frac{(0.93)(2)}{2} (x - 4)$$

$$y - 11 = 0.93(x - 4)$$

$$\underline{y = 0.93x + 7.28}$$

$$x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$x - 4 = \frac{(0.93)2}{2} (y - 11)$$

$$x - 4 = 0.93(y - 11)$$

$$\underline{x = 0.93y - 6.23}$$

=

Find the correlation coefficient for the two groups

A = 92 89 87 86 83 77 71 63 53 50

B = 86 83 91 77 68 85 52 82 37 57

x	y	$z = x - y$	x^2	y^2	z^2
92	86	6	8464	7396	36
89	83	6	7921	6889	36
87	91	-4	7569	8281	16
86	77	9	7396	5929	81
83	68	15	6889	4624	225
77	85	-8	5929	7225	64
71	52	19	5041	2704	361
63	82	-19	3969	6724	361
53	37	16	2809	1369	256
50	57	-7	2500	3249	49
Σx = 751	Σy = 718	Σz = 33	Σx^2 = 58487	Σy^2 = 54390	Σz^2 = 1485

$$\underline{n = 10}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{751}{10} = 75.1, \bar{y} = \frac{\sum y}{n} = \frac{718}{10} = 71.8, \bar{z} = \frac{\sum z}{n} = \frac{33}{10} = 3.3$$

$$\begin{aligned}\sigma_x^2 &= \frac{\sum x^2}{n} - (\bar{x})^2 & \sigma_y^2 &= \frac{\sum y^2}{n} - (\bar{y})^2 & \sigma_z^2 &= \frac{\sum z^2}{n} - (\bar{z})^2 \\ &= \frac{58487}{10} - (75.1)^2 & &= \frac{54390}{10} - (71.8)^2 & &= \frac{1485}{10} - (3.3)^2 \\ &= 208.69 & &= 283.76 & &= 137.61\end{aligned}$$

$$\sigma_x = 14.45, \sigma_y = 16.85$$

$$\begin{aligned}r &= \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y} \\ &= \frac{208.69 + 283.76 - 137.61}{2(14.45)(16.85)} \\ &= 0.73\end{aligned}$$

①
Find the correlation co-eff and equation of the lines of regression for the following values of x & y

x	1	2	3	4	5
y	2	5	3	8	7

x	y	$z = x - y$	x^2	y^2	z^2
1	2	-1	1	4	1
2	5	-3	4	25	9
3	3	0	9	9	0
4	8	-4	16	64	16
5	7	-2	25	49	4
$\sum x = 15$	$\sum y = 25$	$\sum z = -10$	$\sum x^2 = 55$	$\sum y^2 = 151$	$\sum z^2 = 30$

$$n = 5$$

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{5} = 3, \quad \bar{y} = \frac{\sum y}{n} = \frac{25}{5} = 5, \quad \bar{z} = \frac{\sum z}{n} = \frac{-10}{5} = -2$$

$$\begin{aligned}\sigma_x^2 &= \frac{\sum x^2}{n} - (\bar{x})^2, & \sigma_y^2 &= \frac{\sum y^2}{n} - (\bar{y})^2, & \sigma_z^2 &= \frac{\sum z^2}{n} - (\bar{z})^2 \\ &= \frac{55}{5} - (3)^2, & &= \frac{151}{5} - (5)^2, & &= \frac{30}{5} - (-2)^2 \\ &= 2, & &= 5.2, & &= 2\end{aligned}$$

$$\sigma_x = \sqrt{2} = \cancel{1.41}, \quad \sigma_y = \sqrt{5.2} = 2.28$$

$$\begin{aligned}r &= \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x\sigma_y} \\ &= \frac{2 + 5.2 - 2}{2(\cancel{1.41})(2.28)} = 0.81\end{aligned}$$

The equation of lines of regression are

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

$$y - 5 = (0.81) \left(\frac{2.28}{\cancel{1.41}} \right) (x - 3)$$

$$y - 5 = 1.31 (x - 3)$$

$$y = 1.31x + 1.07$$

$$x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$x - 3 = (0.81) \left(\frac{\cancel{1.41}}{2.28} \right) (y - 5)$$

$$x - 3 = 0.50 (y - 5)$$

$$x = 0.50y + 0.50$$

4) Find the correlation co-effⁿ b/w x & y for the foll data. Also obtain the regression lines.

x	1	2	3	4	5	6	7	8	9	10
y	10	12	16	28	25	36	41	49	40	50

x	y	$z = x - y$	x^2	y^2	z^2
1	10	-9	1	100	81
2	12	-10	4	144	100
3	16	-13	9	256	169
4	28	-24	16	784	576
5	25	-20	25	625	400
6	36	-30	36	1296	900
7	41	-34	49	1681	1156
8	49	-41	64	2401	1681
9	40	-31	81	1600	961
10	50	-40	100	2500	1600
$\sum x = 55$	$\sum y = 307$	$\sum z = -252$	$\sum x^2 = 385$	$\sum y^2 = 11381$	$\sum z^2 = 7624$

$n = 10$

$$\bar{x} = \frac{\sum x}{n} = \frac{55}{10} = 5.5, \quad \bar{y} = \frac{\sum y}{n} = \frac{307}{10} = 30.7, \quad \bar{z} = \frac{\sum z}{n} = \frac{-252}{10} = -25.2$$

$$\begin{aligned} \sigma_x^2 &= \frac{\sum x^2}{n} - (\bar{x})^2 = \frac{385}{10} - (5.5)^2 = 8.25 \\ \sigma_y^2 &= \frac{\sum y^2}{n} - (\bar{y})^2 = \frac{11381}{10} - (30.7)^2 = 196.21 \\ \sigma_z^2 &= \frac{\sum z^2}{n} - (\bar{z})^2 = \frac{7624}{10} - (-25.2)^2 = 127.36 \end{aligned}$$

$$\sigma_x = 2.87, \quad \sigma_y = 14.01$$

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}$$

$$r = \frac{8 \cdot 25 + 196 \cdot 21 - 127 \cdot 36}{2(2.87)(14.01)}$$

$$= 0.96$$

Eq of lines of regression are

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$(y - 30.7) = 0.96 \left(\frac{14.01}{2.87} \right) (x - 5.5) \quad (x - 5.5) = 0.96 \left(\frac{2.87}{14.01} \right) (y - 30.7)$$

$$y = 4.69x + 4.93$$

$$x = 0.20y - 0.55$$

==

4) Find the co-eff of correlation for the following data

x 10 14 18 22 26 30

y 18 12 24 6 30 36

x	y	$x - y$	x^2	y^2	xy
10	18	-8	100	324	180
14	12	2	196	144	168
18	24	-6	324	576	432
22	6	16	484	36	132
26	30	-4	676	900	780
30	36	-6	900	1296	1080
$\Sigma x = 120$	$\Sigma y = 126$	$\Sigma (x - y) = -6$	$\Sigma x^2 = 2680$	$\Sigma y^2 = 3276$	$\Sigma xy = 412$

$$n = 6$$

We have lines of regression

$$Y = \frac{\sum XY}{\sum x^2} x$$

$$x = \frac{\sum xy}{\sum y^2} y$$

$$y - 15 = \frac{360}{204} (x - 7)$$

$$y - 15 = 1.76 (x - 7)$$

$$y = 1.76x + 2.65$$

$$x - 7 = \frac{360}{818} (y - 15)$$

$$x - 7 = 0.44 (y - 15)$$

$$x = 0.44y + 0.40$$

$$r = \sqrt{(\text{Co-eff of } x)(\text{Co-eff of } y)}$$

$$= \sqrt{(1.76)(0.44)}$$

$$= 0.88$$

(r is +ve because both co-eff of x & y are +ve)

Obtain the lines of regression & hence find co-eff of correlation for the data.

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

x	y	$x = (x - 4)$	$y = (y - 11)$	x^2	y^2	xy
1	9	-3	-2	9	4	6
2	8	-2	-3	4	9	6
3	10	-1	-1	1	1	1
4	12	0	1	0	1	0
5	11	1	0	1	0	0
6	13	2	2	4	4	4
7	14	3	3	9	9	9
$\sum x = 28$	$\sum y = 77$			$\sum x^2 = 28$	$\sum y^2 = 28$	$\sum xy = 26$

$$n = 7$$

$$\bar{x} = \frac{\sum x}{n} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{\sum y}{n} = \frac{77}{7} = 11$$

$$x = x - \bar{x} = x - 4$$

$$y = y - \bar{y} = y - 11$$

$$\bar{x} = \frac{\sum x}{n} = \frac{120}{6} = 20, \quad \bar{y} = \frac{\sum y}{n} = \frac{126}{6} = 21, \quad \bar{z} = \frac{\sum z}{n} = \frac{-6}{6} = -1$$

$$\begin{aligned}\sigma_x^2 &= \frac{\sum x^2}{n} - (\bar{x})^2, & \sigma_y^2 &= \frac{\sum y^2}{n} - (\bar{y})^2, & \sigma_z^2 &= \frac{\sum z^2}{n} - (\bar{z})^2 \\ &= \frac{2680}{6} - (20)^2, & &= \frac{3276}{6} - (21)^2, & &= \frac{412}{6} - (-1)^2 \\ &= 46.67, & &= 105, & &= 67.67\end{aligned}$$

$$\sigma_x = 6.83, \quad \sigma_y = 10.25$$

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y} = \frac{46.67 + 105 - 67.67}{2(6.83)(10.25)} = \frac{84}{140.02} = 0.60$$

Type-2

Obtain the lines of regression and hence find the co-eff of correlation for the data

x	1	3	4	2	5	8	9	10	13	15
y	8	6	10	8	12	16	16	10	32	32

x	y	X = (x-7)	Y = (y-15)	X ²	Y ²	XY
1	8	-6	-7	36	49	42
3	6	-4	-9	16	81	36
4	10	-3	-5	9	25	15
2	8	-5	-7	25	49	35
5	12	-2	-3	4	9	6
8	16	1	1	1	1	1
9	16	2	1	4	1	2
10	10	3	-5	9	25	-15
13	32	6	17	36	289	102
15	32	8	17	64	289	136
$\sum x = 70$	$\sum y = 150$			$\sum X^2 = 204$	$\sum Y^2 = 818$	$\sum XY = 360$

$$\begin{aligned}n &= 10 \\ \bar{x} &= \frac{\sum x}{n} = \frac{70}{10} = 7 \\ \bar{y} &= \frac{\sum y}{n} = \frac{150}{10} = 15 \\ X &= x - \bar{x} = x - 7 \\ Y &= y - \bar{y} = y - 15\end{aligned}$$

We have lines of regression

$$Y = \frac{\sum XY}{\sum x^2} x$$

$$x = \frac{\sum XY}{\sum y^2} y$$

$$(y-11) = \frac{26}{28} (x-4)$$

$$y-11 = 0.93 (x-4)$$

$$y = 0.93x + 7.28$$

$$(x-4) = \frac{26}{28} (y-11)$$

$$x-4 = 0.93 (y-11)$$

$$x = 0.93y - 6.23$$

$$r = \sqrt{(\text{Co-eff of } x)(\text{Co-eff of } y)}$$

$$= \sqrt{(0.93)(0.93)}$$

$$= 0.93$$

(r is +ve because both co-eff of x & y are +ve)

3/ The following data gives the age of husband (x) and the age of wife (y) in years. form the two regression lines and calculate the age of husband corresponding to 16 years age of wife.

x	36	23	27	28	28	29	30	31	33	35
y	29	18	20	22	27	21	29	27	29	28

x	y	$X = (x-30)$	$Y = (y-25)$	x^2	y^2	xy
36	29	6	4	36	16	24
23	18	-7	-7	49	49	49
27	20	-3	-5	9	25	15
28	22	-2	-3	4	9	6
28	27	-2	2	4	4	-4
29	21	-1	-4	1	16	4
30	29	0	4	0	16	0
31	27	1	2	1	4	2
33	29	3	4	9	16	12
35	28	5	3	25	9	15
$\sum x$ = 300	$\sum y$ = 250			$\sum x^2$ = 138	$\sum y^2$ = 164	$\sum xy$ = 122

$$n = 10$$

$$\bar{x} = \frac{\sum x}{n} = \frac{300}{10} = 30$$

$$\bar{y} = \frac{\sum y}{n} = \frac{250}{10} = 25$$

$$x = x - \bar{x} = x - 30$$

$$y = y - \bar{y} = y - 25$$

We have lines of regression

$$Y = \frac{\sum XY}{\sum X^2} X$$

$$, X = \frac{\sum XY}{\sum Y^2} Y$$

$$(y-25) = \frac{124}{138} (x-30)$$

$$(x-30) = \frac{124}{164} (y-25)$$

$$(y-25) = 0.89(x-30)$$

$$x-30 = 0.75(y-25)$$

$$y-25 = 0.89x - 26.70$$

$$x-30 = 0.75y - 18.75$$

$$y = 0.89x - 26.70 + 25$$

$$x = 0.75y - 18.75 + 30$$

$$y = 0.89x - 1.70$$

$$x = 0.75y + 11.25$$

To find x given $y = 16$

$$x = 0.75(16) + 11$$

$$= 23.25$$

$$x \approx 23.$$

The age of husband is 23 years corresponding to wife's age of 16 years.

Type-3

If $8x - 10y + 66 = 0$ and $40x - 18y = 214$ are the two regression lines. Find the mean of x 's, y 's and the correlation coefficient. Find σ_y if $\sigma_x = 3$.

WKT regression lines passes through $\bar{x}$ & $\bar{y}$

$$\left. \begin{array}{l} 8\bar{x} - 10\bar{y} = -66 \\ 40\bar{x} - 18\bar{y} = 214 \end{array} \right\} \text{ By solving we get } \bar{x} = 13, \bar{y} = 17$$

We shall rewrite the equation of regression lines to find the regression co-eff.

$$10y = 8x + 66 \quad \text{or} \quad y = 0.8x + 6.6 \quad \text{--- (1)}$$

$$40x = 18y + 214 \quad \text{or} \quad x = 0.45y + 5.35 \quad \text{--- (2)}$$

from (1) & (2)

$$\text{or } \frac{\sigma_y}{\sigma_x} = 0.8, \quad \text{or } \frac{\sigma_x}{\sigma_y} = 0.45$$

$$\therefore r = \sqrt{(\text{co-eff of } x)(\text{co-eff of } y)} = \sqrt{(0.8)(0.45)} = 0.6$$

(r is +ve because both co-eff of x & y are +ve)

$$0.6 \frac{\sigma_y}{3} = 0.8$$

$$\therefore \sigma_y = \frac{0.8 \times 3}{0.6}$$

$$\sigma_y = 4$$

If Given the equation of the regression lines $x = 19.13 - 0.87y$, $y = 11.64 - 0.5x$. Compute mean of x 's, mean of y 's and the co-eff of correlation.

WKT regression lines passes through $\bar{x}$ & $\bar{y}$

$$\bar{x} = 19.13 - 0.87\bar{y}$$

$$\bar{y} = 11.64 - 0.5\bar{x}$$

(r is -ve because both co-eff of x & y are -ve)

$$\left. \begin{array}{l} \bar{x} + 0.87\bar{y} = 19.13 \\ 0.5\bar{x} + \bar{y} = 11.64 \end{array} \right\} \text{ By solving we get } \bar{x} = 15.93, \bar{y} = 3.67$$

$$\therefore r = \sqrt{(\text{co-eff of } x)(\text{co-eff of } y)} = \sqrt{(-0.5)(-0.87)} = -0.66$$

3. Compute $\bar{x}$, $\bar{y}$ and r from the following equation of the regression lines $2x+3y+1=0$ & $x+6y-4=0$

The regression lines pass through $\bar{x}$ & $\bar{y}$

$$\begin{aligned} 2\bar{x} + 3\bar{y} + 1 &= 0 \\ \bar{x} + 6\bar{y} - 4 &= 0 \end{aligned} \quad \left\{ \begin{array}{l} \text{By solving} \\ \begin{array}{l} 2\bar{x} + 3\bar{y} = -1 \\ \bar{x} + 6\bar{y} = 4 \end{array} \end{array} \right. \quad \begin{array}{l} \bar{x} = -2 \\ \bar{y} = 1 \end{array}$$

Rewriting the equation

from $2x+3y+1=0$ we have $y = -\frac{2}{3}x - \frac{1}{3}$

from $x+6y-4=0$ we have $x = -6y + 4$

$$\therefore r = \sqrt{(\text{co-eff of } x)(\text{co-eff of } y)} = \sqrt{\left(\frac{2}{3}\right)(-6)} = -2$$

(r is -ve as co-eff of both x & y are -ve)

But value of r should lie b/w $+1$ & -1

So we must again rewrite the equation

from $2x+3y+1=0$ we have $x = -\frac{3}{2}y - \frac{1}{2}$

from $x+6y-4=0$ we have $y = -\frac{1}{6}x + \frac{2}{3}$

$$\therefore r = \sqrt{(\text{co-eff of } x)(\text{co-eff of } y)} = \sqrt{\left(-\frac{1}{6}\right)\left(-\frac{3}{2}\right)} = -\frac{1}{2}$$

(r is -ve as co-eff of both x & y are -ve)

The value of r lies b/w $+1$ & -1

$$\therefore \bar{x} = -2, \quad \bar{y} = 1, \quad r = -\frac{1}{2} = -0.5$$

4# Given

	x-series	y-series
Mean	18	100
SD	14	20

and $r = 0.8$ write down the equation of the lines of regression and hence find the most probable value of y when $x = 70$.

By data $\bar{x} = 18$, $\bar{y} = 100$, $\sigma_x = 14$, $\sigma_y = 20$

We have equation of the regression lines

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

$$y - 100 = (0.8) \frac{(20)}{(14)} (x - 18) \quad x - 18 = (0.8) \left(\frac{14}{20} \right) (y - 100)$$

$$y - 100 = 1.14 (x - 18) \quad x - 18 = 0.56 (y - 100)$$

$$y = 1.14x + 79.48$$

$$y = ax + b$$

when $x = 70$,

$$y = 1.14(70) + 79.48$$

$$y = 159.28$$

$$x = 0.56y - 38$$

$$x = ay + b$$

5# Show that if θ is the angle b/w the lines of regression then

$$\tan \theta = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \left(\frac{1 - r^2}{r} \right)$$

WKT if θ is acute, ^{then} the angle between the lines $y = m_1x + c_1$ and $y = m_2x + c_2$ is given by

$$\tan \theta = \frac{m_2 - m_1}{1 + m_1 m_2} \quad \text{--- (*)}$$

We have lines of regression

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x}) \quad \& \quad x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y}) \quad \text{--- (2)}$$

$$\text{--- (1)}$$

We write eq (2) as

$$y - \bar{y} = \frac{\sigma_y}{r\sigma_x} (x - \bar{x}) \quad \text{--- (3)}$$

slopes of (1) & (3) are

$$m_1 = r \frac{\sigma_y}{\sigma_x} \quad \& \quad m_2 = \frac{\sigma_y}{r\sigma_x}$$

Subs in (*)

$$\tan \theta = \frac{\frac{\sigma_y}{r\sigma_x} - r \frac{\sigma_y}{\sigma_x}}{1 + r \frac{\sigma_y}{\sigma_x} \frac{\sigma_y}{r\sigma_x}}$$

$$= \frac{\frac{\sigma_y}{\sigma_x} \left(\frac{1}{r} - r \right)}{1 + \frac{\sigma_y^2}{\sigma_x^2}} = \frac{\frac{\sigma_y}{\sigma_x} \left(\frac{1-r^2}{r} \right)}{\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2}} = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \left(\frac{1-r^2}{r} \right)$$

Q. If the co-eff of correlation between two variables x & y is 0.5 and the acute angle between their lines of regression is $\tan^{-1} 3/5$. Show that $\sigma_y = 2\sigma_x$ or $\sigma_x = 2\sigma_y$

By data $r = 0.5$, $\theta = \tan^{-1}(3/5)$
 $\tan \theta = 3/5$

The angle b/w the lines of regression is

$$\tan \theta = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \left(\frac{1-r^2}{r} \right)$$

$$\frac{3}{5} = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \left(\frac{1 - (1/2)^2}{1/2} \right)$$

$$\frac{3}{5} = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \left(\frac{3}{4} \times \frac{2}{1} \right)$$

$$\frac{1}{5} = \frac{\sigma_x \sigma_y}{2(\sigma_x^2 + \sigma_y^2)}$$

$$2\sigma_x^2 + 2\sigma_y^2 = 5\sigma_x \sigma_y$$

$$2\sigma_x^2 - 5\sigma_x \sigma_y + 2\sigma_y^2 = 0$$

On factorizing, $(2\sigma_x - \sigma_y)(\sigma_x - 2\sigma_y) = 0$
 $2\sigma_x = \sigma_y$ or $\sigma_x = 2\sigma_y$

$$\begin{aligned} 2\sigma_x^2 - 4\sigma_x \sigma_y - \sigma_x \sigma_y + 2\sigma_y^2 &= 0 \\ 2\sigma_x(\sigma_x - 2\sigma_y) - \sigma_y(\sigma_x - 2\sigma_y) &= 0 \\ (2\sigma_x - \sigma_y)(\sigma_x - 2\sigma_y) &= 0 \end{aligned}$$

If $\sigma_x = \sigma_y$ and the angle between the regression lines is $\tan^{-1}(3)$. Find the correlation coefficient.

The angle between the lines of regression is

$$\tan \theta = \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \left(\frac{1 - r^2}{r} \right)$$

By data $\theta = \tan^{-1} 3$, $\tan \theta = 3$ & $\sigma_x = \sigma_y$

$$3 = \frac{\sigma_x^2}{2\sigma_x^2} \left(\frac{1 - r^2}{r} \right)$$

$$6r = 1 - r^2$$

$$r^2 + 6r - 1 = 0$$

$$r = \frac{-6 \pm \sqrt{40}}{2} = \frac{-6 \pm 2\sqrt{10}}{2} = -3 \pm \sqrt{10} = +0.16, -6.16$$

Since $|r| \leq 1$ the required $r = 0.16$

If $x = 4y + 5$ and $y = Kx + 4$ are the regression lines of x on y and y on x respectively. Prove that $0 \leq K \leq 1/4$.

WKT Co-eff of correlation $\Rightarrow r = \sqrt{(\text{Co-eff of } x)(\text{Co-eff of } y)}$

$$r = \sqrt{K(4)}$$

$$r = \sqrt{4K}$$

$$r^2 = 4K$$

Also $0 \leq r^2 \leq 1$
 $0 \leq 4K \leq 1$ i.e. $0 \leq K \leq 1/4$

Q1. If x & y are random variables with S.D σ_x & σ_y it was found that random variables $(x+y)$, $(x-y)$, $(2x+y)$ respectively have variance 15, 11, 29. Compute S.D of x , y and also the co-eff of correlation.

By data $\sigma_{x+y}^2 = 15$, $\sigma_{x-y}^2 = 11$, $\sigma_{2x+y}^2 = 29$ — (1)

We have a formula $\sigma_{ax+by}^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y$

Take (a, b) equal to $(1, 1)$, $(1, -1)$, $(2, 1)$ we have

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y$$

$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$$

$$\sigma_{2x+y}^2 = 4\sigma_x^2 + \sigma_y^2 + 4\rho\sigma_x\sigma_y$$

Using (1) in these eqⁿ

$$\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y = 15 \quad \text{--- (2)}$$

$$\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y = 11 \quad \text{--- (3)}$$

$$4\sigma_x^2 + \sigma_y^2 + 4\rho\sigma_x\sigma_y = 29 \quad \text{--- (4)}$$

(2) + (3)

$$2\sigma_x^2 + 2\sigma_y^2 = 26 \quad \text{ie } \sigma_x^2 + \sigma_y^2 = 13 \quad \text{--- (5)}$$

$$2 \times (3) + (4) \text{ gives } 6\sigma_x^2 + 3\sigma_y^2 = 51 \quad \text{--- (6)}$$

On solving (5) & (6) we get $\sigma_x^2 = 4$, $\sigma_y^2 = 9$

$$\therefore \sigma_x = 2 \text{ \& } \sigma_y = 3$$

Subs these in (2), $4 + 9 + 12\rho = 15$
 $\therefore \rho = 1/6 = 0.17$

Q2. The S.D's of x & y are 2 & 3 respectively. Find the S.D's of $x+y$ & $x-y$ given that 0.4 is the correlation co-eff b/w x & y .

By data $\sigma_x = 2$, $\sigma_y = 3$, $\rho = 0.4$

We have a formula $\sigma_{ax+by}^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y$

Take (a, b) equal to $(1, 1)$, $(1, -1)$ we have

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y$$

$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$$

$$\sigma_{x+y}^2 = 2^2 + 3^2 + 2(0.4)(2)(3) = 17.80$$

$$\sigma_{x-y}^2 = 2^2 + 3^2 - 2(0.4)(2)(3) = 8.20$$

$$\sigma_{x+y} = 4.22 \quad \& \quad \sigma_{x-y} = \underline{\underline{2.86}}$$

11/ A person while calculating the co-eff of correlation between 2 variables x, y from a set of 25 observations obtained the following results

$$\sum x = 125, \sum y = 100, \sum xy = 508, \sum x^2 = 650, \sum y^2 = 160$$

But it was later found that the pair of values

$x : 8 \ 6$ correct were wrongly copied as $x : 6 \ 8$ wrong value
 $y : 12 \ 8$ values $y : 14 \ 6$

Obtain the correct value for the correlation co-eff.

Computation for wrong values is as follows :-

x	y	xy	x^2	y^2
6	14	84	36	196
8	6	48	64	36
$\sum x = 14$	$\sum y = 20$	$\sum xy = 132$	$\sum x^2 = 100$	$\sum y^2 = 232$

Computation for correct values is as follows :-

x	y	xy	x^2	y^2
8	12	96	64	144
6	8	48	36	64
$\sum x = 14$	$\sum y = 20$	$\sum xy = 144$	$\sum x^2 = 100$	$\sum y^2 = 208$

It is observed that $\sum x$, $\sum y$, $\sum x^2$ are unchanged, even after the correction. However corrected value of

$$\sum xy = 508 - 132 + 144 = 520$$

$$\sum y^2 = 460 - 232 + 208 = 436$$

(Given data value $\uparrow$ - wrong value + right value)

$\therefore$ Corrected values of the mean and S.D of x & y are as follows

$$\bar{x} = \frac{\sum x}{n} = \frac{125}{25} = 5$$

$$\sigma_x^2 = \frac{\sum x^2}{n} - (\bar{x})^2 = \frac{650}{25} - 5^2 = 1$$

$$\bar{y} = \frac{\sum y}{n} = \frac{100}{25} = 4$$

$$\sigma_y^2 = \frac{\sum y^2}{n} - (\bar{y})^2 = \frac{436}{25} - 4^2 = 1.44$$

$$\sigma_x = 1, \sigma_y = 1.2$$

$$\text{We have } r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y}$$

$$= \frac{\sum xy - \sum \bar{x}y - \sum \bar{x}\bar{y} + \sum \bar{x}\bar{y}}{n \sigma_x \sigma_y}$$

$$r = \frac{1}{\sigma_x \sigma_y} \left(\frac{\sum xy}{n} - \bar{y} \left(\frac{\sum \bar{x}}{n} \right) - \bar{x} \left(\frac{\sum \bar{y}}{n} \right) + \frac{\sum \bar{x}\bar{y}}{n} \right)$$

$$= \frac{1}{\sigma_x \sigma_y} \left(\frac{\sum xy}{n} - \bar{y} \bar{x} - \bar{x} \bar{y} + \bar{x} \bar{y} \right)$$

$$r = \frac{\frac{\sum xy}{n} - \bar{x} \bar{y}}{\sigma_x \sigma_y}$$

$$r = \frac{\frac{520}{25} - (5)(4)}{(1)(1.2)} = 0.67$$

$$\begin{cases} n\bar{x} = \sum x \\ n\bar{y} = \sum y \\ n\bar{x}\bar{y} = \sum \bar{x}\bar{y} \end{cases}$$

$$= n\bar{x}\bar{y}$$

Rank Correlation (Spearman's rank Correlation)

The qualitative characteristics like honesty, beauty, intelligence etc cannot be measured quantitatively but can be arranged in serial order. We can find the correlation co-eff b/w the ranks of n individuals in 2 attributes under study known as rank correlation.

Let x & y denote the ranks of ' n ' individuals in 2 qualitative characteristics A & B, then Spearman's rank correlation co-eff denoted by ρ is given by

$$\rho = 1 - \frac{6 \sum d^2}{n^3 - n} \quad \text{where } d = x - y$$

$n = \text{no. of pairs}$

or

$$\rho = 1 - \frac{6 \sum (x - y)^2}{n(n^2 - 1)}$$

or

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Cases

1. When ranks are given

Compute d , diff. of ranks , d^2 , $\sum d^2$ & use above formula.

2. When ranks are not given

When actual values are given assign the ranks to ' n ' individuals in descending or ascending order w.r.t 2 variables. The same approach should be followed for all the variables under consideration.

Compute d , diff. of ranks , d^2 , $\sum d^2$ and use above formula

Note :- If The rank correlation co-eff has a value from -1 to 1

2) If $+1$ = perfect +ve correlation b/w ranks

3) If -1 = perfect -ve " " "

4) If 0 = no correlation " "

5) If ranking of x, y are entirely in same order like $x=1, 2, 3, 4, 5$ & $y=1, 2, 3, 4, 5$ then

$6 \leq d^2 = 6 \leq (x-y)^2 = 0$. This will give $S=1$ & is called perfect direct or +ve correlation

6) If ranking of x, y are entirely in opp order like $x=1, 2, 3, 4, 5$ & $y=5, 4, 3, 2, 1$ then

$6 \leq d^2 = 6 \leq (x-y)^2 = 240$. This will give $S=-1$ & is called perfect inverse or negative correlation

x	1	2	3	4	5	$n=5$
y	1	2	3	4	5	

x	y	$d = x - y$	d^2
1	1		
2	2	0	0
3	3		
4	4		
5	5		
			$\sum d^2 = 0$

$$S = 1 - \frac{6 \sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(0)}{5(5^2-1)}$$

$$= 1$$

x	1	2	3	4	5
y	5	4	3	2	1

$n=5$

x	y	$d = x - y$	d^2
1	5	-4	16
2	4	-2	4
3	3	0	0
4	2	2	4
5	1	4	16
			$\sum d^2 = 40$

$$S = 1 - \frac{6 \sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(40)}{5(5^2-1)}$$

$$= 1 - \frac{240}{120}$$

$$= 1 - 2$$

$$= -1$$

1/10 Ten participants in a contest are ranked by 2 judges as follows. Compute their rank correlation co-eff.

x	1	6	5	10	3	2	4	9	7	8
y	6	4	9	8	1	2	3	10	5	7

$n=10$, we shall compute $s(x, y)$

x	y	$d = x - y$	d^2
1	6	-5	25
6	4	2	4
5	9	-4	16
10	8	2	4
3	1	2	4
2	2	0	0
4	3	1	1
9	10	-1	1
7	5	2	4
8	7	1	1
			$\Sigma d^2 = 60$

$$\therefore s = 1 - \frac{6 \Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6(60)}{10(10^2 - 1)} = 1 - \frac{360}{990} = 0.64$$

$\downarrow$
 $s(x, y)$

$\therefore$ rank correlation is 0.64

2) Three judges A, B, C give the foll ranks. Find which pair of judges has common approach.

A	1	6	5	10	3	2	4	9	7	8
B	3	5	8	4	7	10	2	1	6	9
C	6	4	9	8	1	2	3	10	5	7

we shall compute $s(x, y)$, $s(y, z)$, $s(z, x)$.

d is the diff in rank where $x=A$, $y=B$, $z=C$

$$\underline{n=10}$$

$x(A)$	$y(B)$	$z(C)$	$d_1 = x - y$	$d_2 = y - z$	$d_3 = z - x$	d_1^2	d_2^2	d_3^2
1	3	6	-2	-3	5	4	9	25
6	5	4	1	1	-2	1	1	4
5	8	9	-3	-1	4	9	16	16
10	4	8	6	-4	-2	36	16	4
3	7	1	-4	6	-2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	-1	4	1	1
9	1	10	8	-9	1	64	81	1
7	6	5	1	1	-2	1	1	4
8	9	7	-1	2	-1	1	4	1
						$\sum d_1^2$ = 200	$\sum d_2^2$ = 214	$\sum d_3^2$ = 60

$$s(x, y) = 1 - \frac{6 \sum d_1^2}{n(n^2-1)} = 1 - \frac{6(200)}{10(10^2-1)} = -0.21$$

$$s(y, z) = 1 - \frac{6 \sum d_2^2}{n(n^2-1)} = 1 - \frac{6(214)}{10(10^2-1)} = -0.30$$

$$s(z, x) = 1 - \frac{6 \sum d_3^2}{n(n^2-1)} = 1 - \frac{6(60)}{10(10^2-1)} = 0.64$$

It may be observed that $s(x, y)$ & $s(y, z)$ are neg which means their tests are opposite. But $s(z, x)$ is +ve & near to 1. Thus $s(z, x)$ is max^m, the pair of judges A & C have nearest common approach.

3. Calculate the rank correlation co-eff. from the foll data showing ranks of 10 students in 2 subjects

Mathe x	3	8	9	2	7	10	4	6	1	5
Physics y	5	9	10	1	8	7	3	4	2	6

$n=10$ we shall find $\rho(x, y)$

x	y	$d = x - y$	d^2
3	5	-2	4
8	9	-1	1
9	10	-1	1
2	1	+1	1
7	8	-1	1
10	7	3	9
4	3	1	1
6	4	-1	1
1	2	-1	1
5	6		
			$\sum d^2 = 24$

$$\rho(x, y) = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} = 1 - \frac{6(24)}{10(10^2 - 1)} = \underline{\underline{0.855}}$$

4. If Ten students got the foll percentage of marks in 2 subjects x & y. Compute their rank correlation co-eff for the foll data.

Marks in X	78	36	98	25	75	82	90	62	65	39
Marks in Y	84	51	91	60	68	62	86	58	53	47

We shall give ranks according to their order of magnitude. We shall start from 98 as rank 1, 90 as rank 2 and so on.
 ↳ as 98 is highest from both subjects

$n = 10$

Marks in X	Rank X	Marks in Y	Rank Y	$d = X - Y$	d^2
78	4	84	3	1	1
36	9	51	9	0	0
98	1	91	1	0	0
25	10	60	6	4	16
75	5	68	4	1	1
82	3	62	5	-2	4
90	2	86	2	0	0
62	7	58	7	0	0
65	6	53	8	-2	4
39	8	47	10	-2	4
					$\Sigma d^2 = 30$

$$r_{xy} = 1 - \frac{6 \Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6(30)}{10(10^2 - 1)} = 1 - \frac{180}{990} = 0.82$$

(Since rank is not specified as in earlier problems we assign rank as it is given in percentage)

Probability Distribution

Exhaustive event :- An event consisting of all the various possibilities is called exhaustive event.

Mutually exclusive events :- Two or more events are said to be mutually exclusive if the happening of one event prevents the simultaneous happening of the others.

Independent events :- Two or more events are said to be independent if the happening or non happening of one event does not prevent the happening or non-happening of the others.

Probability :- If the outcome of a trial consists of n exhaustive, mutually exclusive, equally possible cases of which m of them are favourable cases to an event E , then the probability of the happening of the event E denoted by $P(E)$ is

$$P(E) = p = \frac{\text{number of favourable cases} = m}{\text{number of possible cases} = n}$$

The probability can be atmost equal to 1.

Probability of non-happening of the event denoted by $P(\bar{E})$ is given by

$$P(\bar{E}) = q = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - P(E) = 1 - p$$

$$\therefore p + q = 1$$

Equivalently $P(E) + P(\bar{E}) = 1$

If $P(E) = 1$, E is called a sure event

$P(E) = 0$, E is called an impossible event.

Ex:- The probability of getting

by a no greater than 2

by an odd no

by an even no when a die is thrown

Sol:- No of possible cases / outcomes = $n = 6$ (1, 2, 3, 4, 5, 6)

by $m = 4$ (3, 4, 5, 6)

$$\therefore P = \frac{m}{n} = \frac{4}{6} = \frac{2}{3}$$

by $m = 3$ (1, 3, 5)

$$\therefore P = \frac{m}{n} = \frac{3}{6} = \frac{1}{2}$$

by $m = 3$ (2, 4, 6)

$$\therefore P = \frac{m}{n} = \frac{3}{6} = \frac{1}{2}$$

Addition theorem of probability

$$P(A \text{ or } B) = P(A) + P(B)$$

Product theorem of probability

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

Random variables

A random variable is a function that assigns a real number to every sample point in the sample space of a random experiment.

The set of all real nos of a random variable X is called range of X .

Eg:- Suppose a coin is tossed twice, we shall associate two different random variables X, Y

$$S = \{HH, HT, TH, TT\}$$

$X = \underline{\text{No}}$ of heads in the outcome

Outcome	HH	HT	TH	TT
Random Variable X	2	1	1	0

$$\text{Range of } X = \{0, 1, 2\}$$

$Y = \underline{\text{No}}$ of tails in the outcome

Outcome	HH	HT	TH	TT
Random Variable Y	0	1	1	2

$$\text{Range of } Y = \{0, 1, 2\}$$

Discrete random variables

If a random variable takes finite or countably infinite number of values then it is called discrete random variables.

Eg:- Tossing coins and observing the number of heads turning up.

Continuous random variables

If a random variable takes non countable infinite number of values then it is called continuous random variable.

Eg:- Weight of articles, conducting a survey on the life of electric bulbs.

Probability function and Discrete probability distribution

If for each value x_i of a discrete random variable X we assign a real number $p(x_i)$ such that

$$p(x_i) \geq 0, \quad \sum_i p(x_i) = 1$$

then the func $p(x)$ is called a probability function.

If the probability that X take the values x_i is p_i then $P(X=x_i) = p_i$ or $p(x_i)$.

The set of values $[x_i, p(x_i)]$ is called a discrete probability distribution.

The function $P(x)$ is called the probability density function (p.d.f) or probability mass function (p.m.f)

The distribution function $F(x)$ is

$F(x) = P(X \leq x) = \sum_{i=1}^x p(x_i)$ x being an integer is called the cumulative distribution function (C.d.f).

The mean and variance of discrete probability distribution is defined as

$$\text{Mean}(\mu) = \sum_i x_i \cdot p(x_i)$$

$$\text{Variance}(\sigma^2) = \sum_i (x_i - \mu)^2 \cdot p(x_i)$$

$$\text{Standard deviation}(\sigma) = \sqrt{\sigma^2}$$

If a coin is tossed twice. A random Variable X represent the number of heads turning up. find the discrete probability distribution for X . Also find mean & variance.

Sol :- $S = \{HH, HT, TH, TT\}$

$$X = \{2, 1, 1, 0\}$$

$$P(HH) = \frac{1}{4}, P(HT) = \frac{1}{4}, P(TH) = \frac{1}{4}, P(TT) = \frac{1}{4}$$

$$P(X=0, \text{ no head}) = P(TT) = \frac{1}{4}$$

$$P(X=1, 1 \text{ head}) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X=2, 2 \text{ heads}) = P(HH) = \frac{1}{4}$$

The discrete probability distribution for X is

$X = x_i$	0	1	2
$P(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

We have $p(x_i) > 0$ & $\sum p(x_i) = 1$

$$\text{Mean} = \mu = \sum x_i \cdot p(x_i) = 0 + \frac{1}{2} + \frac{1}{2} = 1$$

$$\begin{aligned} \text{Variance} = V &= \sum (x_i - \mu)^2 \cdot p(x_i) \\ &= (0-1)^2 \frac{1}{4} + (1-1)^2 \frac{1}{2} + (2-1)^2 \frac{1}{4} \\ &= \frac{1}{2} \end{aligned}$$

If a random experiment of tossing a die twice is performed. Random Variable X on this sample space is defined to be the sum of 2 nos turning up on the toss. Find the discrete probability distribution for the random variable X and compute the corresponding mean and standard deviation.

$S = \{(x, y) \text{ where } x = 1, 2, \dots, 6; y = 1, 2, \dots, 6\}$

$S = \{(1, 1) (1, 2) (1, 3) \dots (6, 4) (6, 5) (6, 6)\}$

No. of elements in $S = n(S) = 36$.

The set of values of the random variable X defined as the sum of 2 nos on the face of the die are

$\{2, 3, 4, \dots, 10, 11, 12\}$

Elements of S or Events E	$X = x_i$	Total no. $n(E)$ of events
1 $\{ (1, 1) \}$	$x_1 = 2$	1
2 $\{ (1, 2) (2, 1) \}$	$x_2 = 3$	2
3 $\{ (1, 3) (3, 1) (2, 2) \}$	$x_3 = 4$	3
4 $\{ (1, 4) (4, 1) (2, 3) (3, 2) \}$	$x_4 = 5$	4
5 $\{ (1, 5) (5, 1) (2, 4) (4, 2) (3, 3) \}$	$x_5 = 6$	5
6 $\{ (1, 6) (6, 1) (2, 5) (5, 2) (4, 3) (3, 4) \}$	$x_6 = 7$	6
7 $\{ (2, 6) (6, 2) (3, 5) (5, 3) (4, 4) \}$	$x_7 = 8$	5
8 $\{ (3, 6) (6, 3) (4, 5) (5, 4) \}$	$x_8 = 9$	4
9 $\{ (4, 6) (6, 4) (5, 5) \}$	$x_9 = 10$	3
10 $\{ (5, 6) (6, 5) \}$	$x_{10} = 11$	2
11 $\{ (6, 6) \}$	$x_{11} = 12$	1
		36

$$P(x_1) = \frac{n(E)}{n(S)} = \frac{1}{36}, \quad P(x_2) = \frac{2}{36} \dots$$

The discrete probability distribution for X

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(x_i)$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

$P(x_i) > 0 \quad \sum P(x_i) = 1$

Mean = $\mu = \sum x_i p(x_i) = \frac{252}{36} = 7$

Variance = $V = \sum_i (x_i - \mu)^2 p(x_i) = \frac{210}{36} = \frac{35}{6}$

S.D = $\sqrt{V} = \sqrt{35/6}$

3.6 S.T the foll distribution represents a discrete probability distribution & find mean variance.

x	10	20	30	40
$p(x)$	$1/8$	$3/8$	$3/8$	$1/8$

We have $p(x_i) > 0$ & $\sum p(x_i) = 1$

Mean = $\mu = \sum p(x_i) \cdot x_i = \frac{10 + 60 + 90 + 40}{8} = 25$

Variance = $V = \sum (x_i - \mu)^2 \cdot p(x_i)$

$= 225 \cdot 1/8 + 25 \cdot 3/8 + 25 \cdot 3/8 + 225 \cdot 1/8$

$= \frac{600}{8}$

$= 75$

4. find the value of K such that the following is a discrete probability distribution. Hence find its mean and standard deviation
Also find $P(X \leq 1)$, $P(X > 1)$ & $P(-1 < X \leq 2)$

x	-3	-2	-1	0	1	2	3
$P(x)$	K	2K	3K	4K	3K	2K	K

We must have $p(x) \geq 0$ & $\sum p(x) = 1$. The first condⁿ is satisfied if $k \geq 0$ & second condition requires

$$k + 2k + 3k + 4k + 3k + 2k + k = 1$$

$$16k = 1$$

$$k = 1/16$$

The d.p.d is as follows

x	-3	-2	-1	0	1	2	3
$p(x)$	$1/16$	$2/16$	$3/16$	$4/16$	$3/16$	$2/16$	$1/16$

$$\text{Mean} = \mu = \sum x_i \cdot p(x_i) = \frac{1}{16}(-3 - 2 - 1 + 0 + 1 + 2 + 3) = 0$$

$$\begin{aligned} \text{Variance} = \sigma^2 &= \sum (x_i - \mu)^2 \cdot p(x_i) \\ &= \frac{1}{16}(9 + 8 + 3 + 0 + 3 + 8 + 9) \\ &= 40/16 = 5/2 \end{aligned}$$

$$\therefore k = 1/16, \text{ Mean} = 0, \text{ S.D} = \sqrt{5/2}$$

$$\begin{aligned} \text{Also } P(x \leq 1) &= P(-3) + P(-2) + P(-1) + P(0) + P(1) \\ &= 13/16 \end{aligned}$$

$$P(x \geq 1) = P(2) + P(3) = 3/16$$

$$P(-1 < x \leq 2) = P(0) + P(1) + P(2) = 9/16$$

5. The p.d.f of a variate x is given by the foll table

x	0	1	2	3	4	5	6
$P(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

for what value of k this represents a valid probability distribution? Also find $P(x \geq 5)$, $P(3 < x \leq 6)$

The probability distribution is valid if $P(x_i) \geq 0$ & $\sum P(x_i) = 1$

Hence we must have $K > 0$ & $K + 3K + 5K + 7K + 9K + 11K + 13K = 1$
 $49K = 1$ or $K = 1/49$

Also $P(x > 5) = P(5) + P(6) = 11K + 13K = 24K = 24/49$

$P(3 < x \leq 6) = P(4) + P(5) + P(6) = 33K = 33/49$
 $= 9K + 11K + 13K$

Ex 6 The probability distribution of a finite random variable X is given by

x_i	-2	-1	0	1	2	3
$P(x_i)$	0.1	K	0.2	2K	0.3	K

find K , mean & variance.

We must have $P(x_i) \geq 0$ & $\sum P(x_i) = 1$ for a probability distribution

$\sum P(x_i) = 1$ requires $4K + 0.6 = 1 \therefore K = 0.1$

Mean (μ) = $\sum x_i P(x_i) = -0.2 - 0.1 + 0.2 + 0.6 + 0.3$
 $= 0.8$

Variance (σ^2) = $\sum x_i^2 P(x_i) - \mu^2$
 $= (0.4 + 0.1 + 0.2 + 1.2 + 0.9) - (0.8)^2 = 2.160$

7 If A random variable X takes the values -3, -2, -1, 0, 1, 2, 3 such that $P(X=0) = P(X < 0)$ and $P(X=-3) = P(X=-2) = P(X=-1) = P(X=1) = P(X=2) = P(X=3)$

find the probability distribution.

Let the distribution $[X, P(X)]$ be as follows

X	-3	-2	-1	0	1	2	3
$P(X)$	P_1	P_2	P_3	P_4	P_5	P_6	P_7

By data $P(X=0) = P(X < 0)$
 $\Rightarrow P(X=0) = P(X=-1) + P(X=-2) + P(X=-3)$
 $P_4 = P_3 + P_2 + P_1$ ——— (1)

Also by data
 $P_1 = P_2 = P_3 = P_5 = P_6 = P_7$ ——— (2)

Further we must have
 $P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 = 1$ ——— (3)

Using (2) in (1) we get $3P_1 = P_4$

Using (2) in (3) we get $6P_1 + P_4 = 1$ But $P_4 = 3P_1$

$\therefore 9P_1 = 1$ or $P_1 = 1/9$

Hence $P_4 = 3(1/9) = 1/3$

Thus the probability distribution is as follows:-

X	-3	-2	-1	0	1	2	3
$P(X)$	$1/9$	$1/9$	$1/9$	$1/3$	$1/9$	$1/9$	$1/9$

8/9/ the random Variable X take the values 1, 2, 3, 4
 such that $2P(X=1) = 3P(X=2) = P(X=3) = 5P(X=4)$
 find the probability distribution function of X .
 and cumulative distribution of X
 Let the distribution $[X, P(X)]$ be as follows

X	1	2	3	4
$P(X)$	P_1	P_2	P_3	P_4

By data $2p_1 = 3p_2 = p_3 = 5p_4$ — (1)

We also have $p_1 + p_2 + p_3 + p_4 = 1$ — (2)

from (1) $p_2 = \frac{2}{3}p_1$, $p_3 = 2p_1$, $p_4 = \frac{2}{5}p_1$

Hence (2) becomes $p_1 + \frac{2}{3}p_1 + 2p_1 + \frac{2}{5}p_1 = 1$

$$\frac{61}{15} p_1 = 1$$

$$\therefore p_1 = \frac{15}{61}$$

Hence $p_2 = \frac{10}{61}$, $p_3 = \frac{30}{61}$, $p_4 = \frac{6}{61}$

The probability distribution $P(x)$ & cumulative distribution $F(x) = \sum_{i=1}^x p(x_i)$ is as follows

$X = x_i$	1	2	3	4
$P(x)$	$\frac{15}{61}$	$\frac{10}{61}$	$\frac{30}{61}$	$\frac{6}{61}$
$F(x)$	$\frac{15}{61}$	$\frac{25}{61}$	$\frac{55}{61}$	$\frac{61}{61} = 1$

If X is a discrete random variable having $p(x)$ defined as follows

$$p(x) = \begin{cases} x/15 & \text{if } 1 \leq x \leq 5 \\ 0 & \text{if } x > 5 \end{cases}$$

show that $p(x)$ is a probability function and find $P(X=1 \text{ or } 2)$.

The probability distribution is as follows:-

x	1	2	3	4	5	6, 7, ...
$p(x)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{5}{15}$	0

We have $p(x_i) \geq 0$ and $\sum_i p(x_i) = 1$

$\therefore p(x)$ is a probability function

$$\begin{aligned}\text{Now } P(x=1 \text{ or } 2) &= P(x=1) + P(x=2) \\ &= \frac{1}{15} + \frac{2}{15} \\ &= \frac{3}{15} \\ &= \frac{1}{5}\end{aligned}$$

10. A random variable x has the foll probability function for values of x

x	0	1	2	3	4	5	6	7
$P(x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

find k , $P(x < 6)$, $P(x > 6)$, $P(3 < x \leq 6)$

also find probability distribution & distribution function of x .

We must have $P(x) \geq 0$ & $\sum P(x) = 1$

The first condⁿ is satisfied for $k \geq 0$ & we have to find k such that $\sum P(x) = 1$

$$0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$

$$10k^2 + 9k - 1 = 0$$

$$(10k^2 - 1)(k + 1) = 0$$

$$k = \frac{1}{10} \text{ or } k = -1$$

$k = -1$ fails & $k \neq -1$

$$\therefore k = \frac{1}{10}$$

Hence

x	0	1	2	3	4	5	6	7
$P(x)$	0	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{100}$	$\frac{1}{50}$	$\frac{17}{100}$

$$\begin{aligned}P(x < 6) &= P(0) + P(1) + P(2) + P(3) + P(4) + P(5) \\ &= 0 + \frac{1}{10} + \frac{1}{5} + \frac{1}{5} + \frac{3}{10} + \frac{1}{100} \\ &= \frac{81}{100} = 0.81\end{aligned}$$

$$P(x \geq 6) = P(6) + P(7) \\ = 1/50 + 17/100 = 19/100 = 0.19$$

$$P(3 < x \leq 6) = P(4) + P(5) + P(6) \\ = 3/10 + 1/100 + 1/50 \\ = 33/100 \\ = 0.33$$

The probability distribution is as follows.

x	0	1	2	3	4	5	6	7
$p(x)$	0	0.1	0.2	0.2	0.3	0.01	0.02	0.17

The cumulative distribution function is as follows

x	0	1	2	3	4	5	6	7
$F(x)$	0	$0+0.1$ $=0.1$	$0.1+0.2$ $=0.3$	$0.3+0.2$ $=0.5$	$0.5+0.3$ $=0.8$	$0.8+0.01$ $=0.81$	$0.81+0.02$ $=0.83$	$0.83+0.17$ $=1$

==

Binomial Distribution

If p is the probability of success and q is the probability of failure the probability of x successes out of n trials is given by

$$P(x) = {}^nC_x p^x q^{n-x}$$

Mean and standard deviation

$$\text{Mean} = \mu = \sum_{x=0}^n x P(x)$$

$$= \sum_{x=0}^n x {}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^n x \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$= \sum_{x=0}^n \frac{n(n-1)!}{(x-1)!(n-x)!} p^x q^{n-x}$$

$$= np \sum_{x=1}^n \frac{(n-1)!}{[(n-1)-(x-1)]!(x-1)!} p^{x-1} q^{(n-1)-(x-1)}$$

$$= np \sum_{x=1}^n {}^{n-1}C_{(x-1)} p^{x-1} q^{(n-1)-(x-1)}$$

$$= np (q+p)^{n-1}$$

$$= np$$

$$\text{Variance} = V = \sum_{x=0}^n x^2 P(x) - \mu^2$$

$$\begin{aligned}
\sum_{x=0}^n x^2 P(x) &= \sum_{x=0}^n [x(x-1) + x] P(x) \\
&= \sum_{x=0}^n x(x-1) P(x) + \sum_{x=0}^n x P(x) \\
&= \sum_{x=0}^n x(x-1) {}^n C_x p^x q^{n-x} + np \\
&= \sum_{x=0}^n x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + np \\
&= \sum_{x=2}^n \frac{n(n-1)(n-2)!}{(x-2)!(n-x)!} p^2 p^{x-2} q^{n-x} + np \\
&= n(n-1) p^2 \sum_{x=2}^n \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{(n-2)-(x-2)} + np \\
&= n(n-1) p^2 \sum_{x=2}^n {}^{(n-2)} C_{x-2} p^{x-2} q^{(n-2)-(x-2)} + np \\
&= n(n-1) p^2 (q+p)^{n-2} + np \\
&= n(n-1) p^2 + np
\end{aligned}$$

$$\therefore \sum_{x=0}^n x^2 P(x) = n(n-1) p^2 + np$$

$$\begin{aligned}
\therefore V &= \{n(n-1) p^2 + np\} - (np)^2 \\
&= n^2 p^2 - np^2 + np - n^2 p^2 \\
&= np(1-p) \\
&= npq
\end{aligned}$$

$$S.D = \sigma = \sqrt{V} = \sqrt{npq}$$

The prob distribution of $\overline{[x, P(x)]}$ where $x = 0, 1, 2, \dots, n$

x	0	1	2	...	n
$P(x)$	q^n	$nq q^{n-1} p$	$n {}^2 C_2 q^{n-2} p^2$	...	p^n

$\therefore \sum P(x) = q^n + nq q^{n-1} p + n {}^2 C_2 q^{n-2} p^2 + \dots + p^n = (q+p)^n = 1$

Discrete Probability Distribution

Binomial Distribution | Bernoulli's distribution

$$\text{Mean } (\mu) = np$$

$$P(x) = {}^nC_x p^x q^{n-x}$$

$$\text{Variance } (v) = npq$$

$$\text{Standard (S.D) } (\sigma) = \sqrt{npq}$$

deviation

Poisson distribution

$$\text{Mean } (\mu) = m$$

$$P(x) = \frac{m^x e^{-m}}{x!}$$

$$\text{Variance } (v) = m$$

$$\text{Standard (SD) } (\sigma) = \sqrt{m}$$

deviation

If find the binomial probability distribution which has mean 2 and variance $4/3$.

$$\text{WKT mean} = np = 2$$

$$\text{Variance} = npq = 4/3$$

$$\therefore 2q = 4/3$$

$$q = 2/3$$

$$p = 1 - q = 1 - 2/3 = 1/3$$

$$\text{Also } np = 2 \Rightarrow n(1/3) = 2 \Rightarrow n = 6.$$

The binomial probability function $P(x) = {}^nC_x p^x q^{n-x}$ becomes $P(x) = {}^6C_x (1/3)^x (2/3)^{6-x}$

The distribution probability function $P(x)$ is as follows:-

x	0	1	2	3	4
$P(x)$	$\binom{6}{0} \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6$	$\binom{6}{1} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^5$	$\binom{6}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^4$	$\binom{6}{3} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^3$	$\binom{6}{4} \left(\frac{1}{3}\right)^4 \left(\frac{2}{3}\right)^2$

5	6
$\binom{6}{5} \left(\frac{1}{3}\right)^5 \left(\frac{2}{3}\right)$	$\left(\frac{1}{3}\right)^6$

If when a coin is tossed 4 times, find the probability of getting
 a) exactly one head
 b) at most 3 heads
 c) at least 2 heads.

$$p = \frac{1}{2} = 0.5$$

$$\text{We have } P(x) = {}^nC_x p^x q^{n-x}$$

$$q = 1 - p = 0.5$$

$$n = 4$$

$$\text{a) } P(1 \text{ head}) = P(x=1) = {}^4C_1 (0.5)^1 (0.5)^3 = 0.25$$

$$\begin{aligned} \text{b) } P(\text{at most 3 heads}) &= P(x=0) + P(x=1) + P(x=2) + P(x=3) \\ &= {}^4C_0 (0.5)^0 (0.5)^4 + {}^4C_1 (0.5)^1 (0.5)^3 \\ &\quad + {}^4C_2 (0.5)^2 (0.5)^2 + {}^4C_3 (0.5)^3 (0.5)^1 \\ &= 0.9375 \end{aligned}$$

$$\begin{aligned} \text{c) } P(\text{at least 2 heads}) &= 1 - [P(x=0) + P(x=1)] \\ &= 1 - [(0.5)^4 + {}^4C_1 (0.5) (0.5)^3] \\ &= 0.6875 \end{aligned}$$

$$\begin{aligned} &P(x=2) + P(x=3) + P(x=4) \\ &= 0.688 \end{aligned}$$

- 3/ The probability that a pen manufactured by a factory be defective is $\frac{1}{10}$. If 12 such pens are manufactured what is the probability that
- a/ exactly 2 are defective
 - b/ at least 2 are defective
 - c/ none of them are defective

Probability of defective pen is $p = \frac{1}{10} = 0.1$
 " of non-defective pen $= q = 1 - p = 1 - 0.1 = 0.9$

We have $P(x) = {}^nC_x p^x q^{n-x}$, $n=12$

a/ $P(x=2) = {}^{12}C_2 (0.1)^2 (0.9)^{10} = 0.2301$

b/ $1 - [P(x=0) + P(x=1)] = 1 - [{}^{12}C_0 (0.1)^0 (0.9)^{12} + {}^{12}C_1 (0.1)^1 (0.9)^{11}]$
 $= 0.3410$

c/ $P(x=0) = {}^{12}C_0 (0.1)^0 (0.9)^{12} = (0.9)^{12} = 0.2824$

- 4/ In a consignment of electric lamps 5% are defective. If a random sample of 8 lamps are inspected what is the probability that one or more lamps are defective?

Probability of a defective lamp $= p = \frac{5}{100} = 0.05$
 $\therefore q = 1 - p = 1 - 0.05 = 0.95$

$P(x) = {}^nC_x p^x q^{n-x}$, $n=8$

(one or more lamps defective) $= P(1) + P(2) + \dots + P(8)$
 $= 1 - P(0)$
 $= 1 - [(0.95)^8]$
 $= 0.3366$

5) The probability that a person aged 60 years will live upto 70 is 0.65. What is the probability that out of 10 persons aged 60 atleast 7 of them will live upto 70.

Let x be the number of persons aged 60 years living upto 70 years.

$$p = 0.65 \quad q = 1 - p = 1 - 0.65 = 0.35$$

$$P(x) = {}^nC_x p^x q^{n-x}, \quad n = 10$$

$$\begin{aligned} \text{To find } P(x \geq 7) &= P(7) + P(8) + P(9) + P(10) \\ &= {}^{10}C_7 (0.65)^7 (0.35)^3 + {}^{10}C_8 (0.65)^8 (0.35)^2 \\ &\quad + {}^{10}C_9 (0.65)^9 (0.35) + (0.65)^{10} \\ &= 0.5138 \end{aligned}$$

$$1 - [P(0) + P(1) + \dots + P(6)]$$

$$1 - [{}^{10}C_0 (0.65)^0 (0.35)^{10} + \dots + {}^{10}C_6 (0.65)^6 (0.35)^4]$$

6) The number of telephone lines busy at an instant of time is a binomial variate with prob 0.1 that a line is busy. If 10 lines are chosen at random, what is the prob that at no line is busy or all lines are busy or atleast one line is busy or at most 2 lines are busy. Let x denote the number of telephone lines busy. $p = 0.1$, $q = 1 - p = 1 - 0.1 = 0.9$, $n = 10$

$$P(x) = {}^nC_x p^x q^{n-x}$$

a) $P(0) = (0.9)^{10} = 0.3487$

b) $P(10) = (0.1)^{10}$

c) $1 - \text{Probability of no line busy} = \text{Prob that atleast one line is busy}$

$$\begin{aligned} &= 1 - P(0) \\ &= 1 - 0.3487 \\ &= 0.6513 \end{aligned}$$

$$\begin{aligned} \text{Prob that at most 2 lines are busy} &= P(0) + P(1) + P(2) \\ &= (0.9)^{10} + 10C_1(0.1)(0.9)^9 \\ &\quad + 10C_2(0.1)^2(0.9)^8 \\ &= 0.9298 \end{aligned}$$

7. In a quiz contest of answering Yes or No what is the probability of guessing atleast 6 answers correctly out of 10 questions asked? Also find the probability of the same if there are 4 options for a correct answer. Let x denote the correct answer

$P = 1/2 \text{ then } q = 1 - P = 1 - 1/2 = 1/2$

$$P(x) = {}^nC_x p^x q^{n-x} = {}^{10}C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = {}^{10}C_x \left(\frac{1}{2}\right)^{10}$$

To find $P(x \geq 6)$

$$\begin{aligned} P(x \geq 6) &= P(6) + P(7) + P(8) + P(9) + P(10) \\ &= \left(\frac{1}{2}\right)^{10} [{}^{10}C_6 + {}^{10}C_7 + {}^{10}C_8 + {}^{10}C_9 + {}^{10}C_{10}] \\ &= \frac{386}{1024} = 0.3770 \end{aligned}$$

When there are 4 options

$P = 1/4 \text{ then } q = 1 - 1/4 = 3/4, n = 10$

$$P(x) = {}^{10}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}$$

$$\begin{aligned} P(x \geq 6) &= P(6) + P(7) + P(8) + P(9) + P(10) \\ &= 0.0197 \end{aligned}$$

8/ In sampling a large number of parts manufactured by a company, the mean number of defectives in samples of 20 is 2. Out of 1000 such samples how many would be expected to contain at least 3 defective parts.

$$\text{Mean } (\mu) = np = 2 \quad \text{but } n = 20$$

$$\therefore 20p = 2, \quad p = \frac{1}{10} = 0.1$$

$$\text{hence } q = 1 - p = 1 - 0.1 = 0.9$$

Let x denote the defective part

$$P(x) = {}^nC_x p^x q^{n-x}$$

Probability of at least 3 defective parts

$$P(x \geq 3) = P(3) + P(4) + \dots + P(20)$$

$$= 1 - [P(0) + P(1) + P(2)]$$

$$= 1 - \left[(0.9)^{20} + 20 {}^1C_1 (0.1)(0.9)^{19} + 20 {}^2C_2 (0.1)^2 (0.9)^{18} \right]$$

$$= 0.3231$$

Thus the number of defectives in 1000 samples is $1000 \times 0.3231 = 3231$

9/ S.T for a binomial distribution

$$P(x+1) = \frac{n-x}{x+1} \frac{p}{q} P(x)$$

$$\text{We have } P(x) = {}^nC_x p^x q^{n-x}$$

$$P(x+1) = {}^nC_{x+1} p^{x+1} q^{n-(x+1)}$$

$$= \frac{n!}{(x+1)! [n-(x+1)]!} p^{x+1} q^{n-(x+1)}$$

$${}^nC_x = \frac{n!}{x! (n-x)!}$$

x/y & divide by $(n-x)$ in the RHS

$$\begin{aligned}
 P(x+1) &= \frac{n!(n-x)}{(x+1)!(n-x)[n-(x+1)]!} p^{x+1} q^{n-(x+1)} \\
 &= \frac{n!(n-x)}{(x+1)x!(n-x)(n-x-1)!} p^{x+1} q^{n-(x+1)} \\
 P(x+1) &= \frac{(n-x)}{(x+1)} \frac{n!}{x!(n-x)!} p p^x \frac{q^{n-x}}{q} \\
 &= \frac{n-x}{x+1} \frac{p}{q} \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
 &= \frac{n-x}{x+1} \frac{p}{q} {}^nC_x p^x q^{n-x} \\
 P(x+1) &= \frac{n-x}{x+1} \frac{p}{q} P(x)
 \end{aligned}$$

10. If the mean and S.D of the number of correctly answered questions in a test given to 4096 students are 2.5 and $\sqrt{1.875}$. find an estimate of the number of candidates answering correctly
 a) 8 or more questions
 b) 2 or less
 c) 5 questions

$$\text{Mean} = np = 2.5$$

$$\text{S.D} = \sigma = \sqrt{npq} = \sqrt{1.875}$$

$$\text{or } npq = 1.875$$

$$2.5q = 1.875$$

$$q = 0.75$$

$$p = 1 - q = 1 - 0.75 = 0.25$$

$$\therefore (0.25)n = 2.5 \quad \therefore n = 10$$

Let x denote the number of correctly answered questions

$$P(x) = {}^{10}C_x p^x q^{10-x} = {}^{10}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}$$

$$4096 P(x) = [{}^{10}C_x (0.25)^x (0.75)^{10-x}] \cdot 4096 \quad 4096 P(x) = 4096 {}^{10}C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}$$

$$= f(x) \quad = 4096 {}^{10}C_x \left(\frac{1}{4}\right)^x \frac{3^{10-x}}{4^{10-x}}$$

$$= 4096 {}^{10}C_x \frac{1}{4^{10}} 3^{10-x}$$

$$= \frac{1}{256} [{}^{10}C_x (3)^{10-x}]$$

a) $f(8) + f(9) + f(10)$

$$= \frac{1}{256} [{}^{10}C_8 3^2 + {}^{10}C_9 \cdot 3 + 1] = 1.7031 \approx 2$$

b) $f(2) + f(1) + f(0)$

$$= \frac{1}{256} [{}^{10}C_2 \cdot 3^8 + {}^{10}C_1 \cdot 3^9 + 3^{10}] = 2152.82 = 2153$$

$$c) f(5) = \frac{1}{256} [{}^{10}C_5 \cdot 3^5] = 239.2931 = 239$$

Based on this concept

$$4096 [{}^{10}C_8 (0.25)^8 (0.75)^2] + 4096 [{}^{10}C_9 (0.25)^9 (0.75)^1] + 4096 [{}^{10}C_{10} (0.25)^{10} (0.75)^0]$$

$$\{4096 [{}^{10}C_2 (0.25)^2 (0.75)^8] + 4096 [{}^{10}C_1 (0.25)^1 (0.75)^9] + 4096 [{}^{10}C_0 (0.25)^0 (0.75)^{10}]\}$$

$$4096 [{}^{10}C_5 (0.25)^5 (0.75)^5]$$

11) An airline knows that 5% of the people making reservations on a certain flight will not turn up. Consequently their policy is to sell 52 tickets for a flight that can only hold 50 people. What is the probability that there will be a seat for every passenger who turns up? The probability p that a passenger will not turn up is $p = 0.05$

$$q = 1 - p = 1 - 0.05 = 0.95, \quad n = 52$$

Let x denote the no. of passenger who will not turn up

$$P(x) = {}^{n}C_x p^x q^{n-x} = {}^{52}C_x (0.05)^x (0.95)^{52-x}$$

A seat is assured for every passenger who turns up ^{only} if the number of passengers who fail to turn up is more than or equal to 2.

Hence we have to find $P(x \geq 2)$

$$\begin{aligned} P(x \geq 2) &= 1 - [P(x < 2)] \\ &= 1 - [P(x=0) + P(x=1)] \\ &= 1 - [(0.95)^5 + 5(0.05)(0.95)^4] \\ &= 1 - 0.2592 \\ &= \underline{\underline{0.7405}} \end{aligned}$$

12. In 800 families with 5 children each how many families would be expected to have at least 3 boys by 5 girls if either 2 or 3 boys at most 2 girls by assuming probabilities for boys and girls to be equal.

p = prob of having a boy = $\frac{1}{2}$

q = prob of having a girl = $\frac{1}{2}$

Let x denotes the number of boys in the family

$$P(x) = {}^nC_x p^x q^{n-x}, \quad n=5$$

$$P(x) = {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} = \frac{1}{2^5} {}^5C_x = \frac{1}{32} {}^5C_x$$

Since we have to find the expected number in respect of 800 families we have

$$800 P(x) = \frac{800}{32} {}^5C_x = 25 \cdot {}^5C_x$$

a) We have find $f(3)$

$$f(3) = 25 \cdot {}^5C_3 = 250$$

b) We have find $f(0)$

$$f(0) = 25 \cdot {}^5C_0 = 25$$

c) We have $f(2) + f(3)$

$$= 25 \cdot {}^5C_2 + 25 \cdot {}^5C_3$$

$$= \cancel{500} = 500$$

If at most 2 girls means that families can have 5 boys and 0 girls or 4 boys and 1 girl or 3 boys and 2 girls.

$$\begin{aligned}
 &= f(5) + f(4) + f(3) \\
 &= 25 \cdot {}^5C_5 + 25 \cdot {}^5C_4 + 25 \cdot {}^5C_3 \\
 &= 25 \times 16 \\
 &= 400
 \end{aligned}$$

13. Five dice were thrown 96 times and the no. of times an odd number actually turned out in the experiment is given. Fit a binomial distribution to this data and calculate the expected frequencies.

No. of dice showing 1 or 3 or 5	0	1	2	3	4	5
Observed frequency	1	10	24	35	18	8

$$\begin{aligned}
 P &= \text{probability of getting 1 or 3 or 5} = 3/6 = 1/2 \\
 q &= 1 - P = 1 - 1/2 = 1/2
 \end{aligned}$$

Let x denote the number of times an odd number turning out.

$$P(x) = {}^nC_x P^x q^{n-x}, \text{ where } n=5$$

$$P(x) = {}^5C_x (1/2)^x (1/2)^{5-x} = \frac{1}{2^5} {}^5C_x = \frac{1}{32} {}^5C_x$$

This is the binomial probability distribution function.

Since 5 dice were thrown 96 times, expected frequencies are obtained from $f(x) = 96 P(x)$ where $x = 0, 1, 2, 3, 4, 5$.

$$\begin{aligned}
 96 P(x) &= 96 \cdot \frac{1}{2^5} {}^5C_x = 3 \cdot {}^5C_x \\
 &= f(x)
 \end{aligned}$$

$$\begin{aligned}
 f(0) &= 3 \cdot {}^5C_0 = 3 & f(3) &= 3 \cdot {}^5C_3 = 30 \\
 f(2) &= 3 \cdot {}^5C_2 = 30 & f(1) &= 3 \cdot {}^5C_1 = 15 \\
 f(4) &= 3 \cdot {}^5C_4 = 15 & f(5) &= 3 \cdot {}^5C_5 = 3
 \end{aligned}$$

The expected frequencies are 3, 15, 30, 30, 15, 3

14/4 4 coins are tossed 100 times and the following results were obtained. Fit a binomial distribution for the data and calculate the theoretical frequencies.

No. of heads	0	1	2	3	4
frequency	5	29	36	25	5

Let x denote the no. of heads and f the corresponding frequency. Since the data is in the form of a frequency distribution we shall first calculate the mean.

$$\text{Mean } (\mu) = \frac{\sum fx}{\sum f} = \frac{0 + 29 + 72 + 75 + 20}{100} = \frac{196}{100} = 1.96$$

$$\mu = np, \quad n = 4$$

$$1.96 = 4p$$

$$p = 0.49$$

$$q = 1 - p = 0.51$$

$$P(x) = {}^nC_x p^x q^{n-x} = {}^4C_x (0.49)^x (0.51)^{4-x}$$

$$100 P(x) = 100 \cdot {}^4C_x (0.49)^x (0.51)^{4-x} = F(x)$$

where $x = 0, 1, 2, 3, 4$

$$F(0) = 100 (0.51)^4 = 6.7652 \approx 7$$

$$F(1) = 100 \cdot {}^4C_1 (0.49)(0.51)^3 = 25.9996 \approx 26$$

$$F(2) = 100 \cdot {}^4C_2 (0.49)^2 (0.51)^2 = 37.47 \approx 37$$

$$F(3) = 100 \cdot {}^4C_3 (0.49)^3 (0.51) = 24.0004 \approx 24$$

$$F(4) = 100 \cdot {}^4C_4 (0.49)^4 = 5.7648 \approx 6$$

Thus the reqd theoretical frequencies are 7, 26, 37, 24, 6

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Date

--	--	--

Poisson Distribution

Page No.

$$P(x) = \frac{m^x e^{-m}}{x!}$$

Mean and S.D for poisson distribution

$$\text{Mean} = \mu = \sum_{x=0}^{\infty} x \cdot P(x)$$

$$= \sum_{x=0}^{\infty} x \cdot \frac{m^x e^{-m}}{x!} \rightarrow x(x-1)!$$

$$= \sum_{x=1}^{\infty} \frac{x m^x e^{-m}}{x(x-1)!}$$

$$= \sum_{x=1}^{\infty} \frac{m^x e^{-m}}{(x-1)!} \quad m^{x-1+1}$$

$$= m e^{-m} \sum_{x=1}^{\infty} \frac{m^{x-1}}{(x-1)!}$$

$$= m e^{-m} \left[1 + \frac{m}{1!} + \frac{m^2}{2!} + \frac{m^3}{3!} + \dots \right]$$

$$= m e^{-m} e^m = m e^0 = m$$

$$\text{Variance} = V = \sum_{x=0}^{\infty} x^2 P(x) - \mu^2$$

$$\sum x^2 P(x) = \sum [x(x-1) + x] \frac{m^x e^{-m}}{x!}$$

$$= \sum_{x=2}^{\infty} \frac{m^x e^{-m}}{(x-2)!} + \sum_{x=1}^{\infty} \frac{m^x e^{-m}}{(x-1)!}$$

$$= m^2 e^{-m} \sum_{x=2}^{\infty} \frac{m^{x-2}}{(x-2)!} + m$$

$$= m^2 e^{-m} \left[1 + \frac{m}{1!} + \frac{m^2}{2!} + \dots \right] + m$$

$$= m^2 e^{-m} e^m + m$$

$$= m^2 + m$$

$$\therefore V = m^2 + m - (m)^2 = m$$

$$S.D = \sigma = \sqrt{V} = \underline{\underline{\sqrt{m}}}$$

1/2 Fit a poisson distribution for the foll data and calculate the theoretical frequencies.

x	0	1	2	3	4
$f(x)$	122	60	15	2	1

$$\text{Mean} = \mu = \frac{\sum fx}{\sum f} = \frac{0 + 60 + 30 + 6 + 4}{200} = 0.5$$

$$\text{Mean} = \mu = m = 0.5$$

$$P(x) = \frac{m^x e^{-m}}{x!} = \frac{(0.5)^x e^{-0.5}}{x!}$$

$$\text{Let } f(x) = 200 P(x) = 200 \frac{(0.5)^x e^{-0.5}}{x!}$$

$$f(0) = 121.3061 \approx 121$$

$$f(1) = 60.6531 \approx 61$$

$$f(2) = 15.1633 \approx 15$$

$$f(3) = 2.5272 \approx 3$$

$$f(4) = 0.3159 \approx 0$$

$$= 200 \frac{(0.5)^x (0.6065)}{x!} = \frac{121.3 (0.5)^x}{x!}$$

Put $x = 0, 1, 2, 3, 4$ in $f(x)$ we have

121, 61, 15, 3, 0

2/2 The number of accidents per day (x) as recorded in a textile industry over a period of 400 days is given. Fit a Poisson distribution for the data and calculate the theoretical frequencies.

x	0	1	2	3	4	5
$f(x)$	173	168	37	18	3	1

$$\text{Mean} = \mu = \frac{\sum fx}{\sum f} = \frac{0 + 168 + 74 + 54 + 12 + 5}{400} = 0.7825$$

$$\mu = m = 0.7825$$

$$P(x) = \frac{m^x e^{-m}}{x!} = \frac{(0.7825)^x e^{-0.7825}}{x!}$$

$$\begin{aligned}
 \text{Let } f(x) &= 400 P(x) \\
 &= \frac{400 (0.7825)^x e^{-0.7825}}{x!} \\
 &= \frac{400 (0.7825)^x (0.4573)}{x!} \\
 f(x) &= 183.306 \frac{(0.7825)^x}{x!}
 \end{aligned}$$

$$\begin{aligned}
 f(0) &= 182.9046 \approx 183 \\
 f(1) &= 143.1228 \approx 143 \\
 f(2) &= 55.9968 \approx 56 \\
 f(3) &= 14.6058 \approx 15 \\
 f(4) &= 2.8573 \approx 3 \\
 f(5) &= 0.4472 \approx 0
 \end{aligned}$$

Put $x=0, 1, 2, 3, 4, 5$ in $f(x)$ we have
183, 143, 56, 15, 3, 0

3/6 P.T $P(x+1) = \frac{m}{x+1} P(x)$

We have $P(x) = \frac{m^x e^{-m}}{x!}$

Rep x by $x+1$
 $P(x+1) = \frac{m^{x+1} e^{-m}}{(x+1)!}$

$$= \frac{m m^x e^{-m}}{(x+1) x!} \underbrace{(n+3)(n+2)(n+1) \dots n(n-1)(n-2) \dots 3 \cdot 2 \cdot 1}_{(n+1)!}$$

$$= \frac{m}{x+1} \frac{m^x e^{-m}}{x!}$$

$$\underline{\underline{P(x+1) = \frac{m}{x+1} P(x)}}$$

4/6 The probabilities of a Poisson Variate takes the value 3 and 4 are equal. Calculate the probabilities of the variate taking the values 0 and 1.

$$P(x) = \frac{m^x e^{-m}}{x!}$$

By data,

$$P(3) = P(4)$$

$$\frac{m^3 e^{-m}}{3!} = \frac{m^4 e^{-m}}{4!}$$

$$\frac{m^3}{6} = \frac{m^4}{24}$$

$$6m = 24 \Rightarrow m = 4$$

$$\therefore P(x) = \frac{e^{-4} 4^x}{x!}$$

$$\text{Now } P(0) = e^{-4} = \frac{1}{e^4} = 0.0183$$

$$P(1) = e^{-4} \cdot 4 = \frac{4}{e^4} = 0.0733$$



Ex 9 If X follows a Poisson law such that $P(X=2) = \frac{2}{3} P(X=1)$ find $P(X=0)$ & $P(X=3)$

$$P(X=x) = \frac{m^x e^{-m}}{x!}$$

By data,

$$\frac{m^2 e^{-m}}{2!} = \frac{2}{3} \frac{m e^{-m}}{1!}$$

$$\frac{m^2}{2} = \frac{2m}{3} \Rightarrow m = \frac{4}{3}$$

$$P(X=x) = e^{-4/3} \frac{(4/3)^x}{x!}$$

$$P(X=0) = e^{-4/3} = 0.2636$$

$$P(X=3) = e^{-4/3} \frac{(4/3)^3}{3!} = 0.1041$$

Ex 10 If x is a Poisson Variate such that $P(x=2) = 9P(x=4) + 90P(x=6)$. Compute the mean and Variance of the Poisson distribution.

$$P(x) = \frac{m^x e^{-m}}{x!}$$

By data,

$$\frac{m^2 e^{-m}}{2!} = 9 \frac{m^4 e^{-m}}{4!} + 90 \frac{m^6 e^{-m}}{6!}$$

$$\frac{m^2}{2} = \frac{9}{24} m^4 + \frac{90}{720} m^6$$

$$\frac{m^2}{2} = \frac{3}{8} m^4 + \frac{1}{8} m^6$$

$$1 = \frac{3}{4} m^2 + \frac{1}{4} m^4$$

$$m^4 + 3m^2 - 4 = 0$$

$$(m^2 - 1)(m^2 + 4) = 0$$

$$m = \pm 1, m = \pm 2i$$

Neglecting imaginary, negative value we have
 $m = 1$

$$\therefore \text{Mean} = \text{Variance} = \underline{\underline{1 = m}}$$

By insp $m=1$

1	0	3	0	-4
0	1	1	4	4
1	3	1	4	4

$m^2 + m + 4m + 4 = 0$
 $m^2(m+1) + 4(m+1) = 0$
 $m^2 + 4 = 0$
 $m^2 = -4$
 $m = \pm \sqrt{-4}$
 $= \pm \sqrt{-1} \sqrt{4}$
 $= \pm i \cdot 2$
 $m = \pm 2i$

If In a certain factory turning ^{out razor} blade there is a small probability of $1/500$ for any blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution to calculate the approxⁿ no of packets containing a) no defective, b) one defective, c) 2 defective blades, in a consignment of 10,000 packets.
 $p = \text{prob of defective blade} = 1/500 = 0.002$

In a packet of 10, the mean defective blades is

$$m = np = 10 \times 0.002 = 0.02$$

$$\mu = m$$

$$np = m$$

$$P(x) = \frac{m^x e^{-m}}{x!} = \frac{e^{-0.02} (0.02)^x}{x!}$$

$$f(x) = 10,000 P(x) = 10,000 \frac{e^{-0.02} (0.02)^x}{x!}$$

$$= \frac{10,000 (0.9802) (0.02)^x}{x!} = \frac{9802 (0.02)^x}{x!}$$

a) $f(0) = 9802$

b) $f(1) = 196.04 = 196$

c) $f(2) \approx 2$

8) The number of accidents in a year to taxi drivers in a city follows a P.D with mean 3. Out of 1000 taxi drivers find ^{the} approx ~~the~~ number of drivers with
 a) no accident in a year
 b) more than 3 accidents in a year.

By data $\mu = 3$ & for Poisson distribution $\mu = m = 3$

$$P(x) = \frac{m^x e^{-m}}{x!} = \frac{e^{-3} 3^x}{x!}$$

$$f(x) = 1000 P(x) = 1000 \frac{e^{-3} 3^x}{x!} = 1000 (0.05)^{\frac{3}{x!}} = \frac{49.7871 \cdot x}{3} = f(x)$$

a) $f(0) \approx 50$

b) $P(x > 3) = 1 - [P(0) + P(1) + P(2) + P(3)]$
 $= 0.3527$

$\therefore$ No of drivers out of 1000 with more than 3 accidents in a year $= 1000 \times 0.3527 = 352.7 \approx 353$

9) 2% of the fuses manufactured by a firm are found to be defective. Find the probability that a box containing 200 fuses contains a) no defective fuses b) 3 or more defective fuses.

$P =$ probability of defective fuse $= 2/100 = 0.02$

Mean $= \mu = m = np = 200 \times 0.02 = 4$

$$\begin{aligned}
 P(x) &= \frac{m^x e^{-m}}{x!} \\
 &= \frac{4^x e^{-4}}{x!} \\
 &= \frac{4^x}{x!} \cdot 0.0183
 \end{aligned}$$

∴ $P(0) = 0.0183$

∴ $1 - [P(0) + P(1) + P(2)] = 0.7621$

10. If the probability of a bad reaction from a certain injection is 0.001. Determine the chance that out of 2000 individuals more than two will get a bad reaction.

$$P(x) = \frac{m^x e^{-m}}{x!}$$

Mean = $m = np = 2000 \times 0.001 = 2 \therefore P(x) = \frac{2^x e^{-2}}{x!}$

$$\begin{aligned}
 P(x > 2) &= 1 - P(x \leq 2) \\
 &= 1 - [P(x=0) + P(x=1) + P(x=2)]
 \end{aligned}$$

$$= 0.3235$$

11. A communication channel receives independent pulses at the rate of 12 pulses per micro second. The probability of transmission error is 0.001 for each micro second. Compute the probabilities of
 a) no error during a micro second
 b) one error per micro second
 c) at least one error per micro second
 d) two errors
 e) at most 2 errors.

Mean number of errors in one micro second

$$\mu = np = m = 12 \times 0.001 = 0.012$$

$$P(x) = \frac{m^x e^{-m}}{x!} = \frac{e^{-0.0120} (0.0120)^x}{x!}$$

$$a) P(0) = 0.9881$$

$$b) P(1) = 0.0119$$

$$c) \begin{aligned} & \rightarrow P(x > 1) = 1 - [P(x \leq 1)] \\ & 1 - P(0) = 1 - e^{-0.0120} = 0.0119 \end{aligned}$$

$$d) P(2) = 0.0001$$

$$e) \begin{aligned} & \rightarrow P(x \leq 2) \\ & P(0) + P(1) + P(2) = 1.0001 \end{aligned}$$

12) A shop has 4 diesel generator sets which it hires everyday. The demand for a gen. set on an average is a poisson variate with value $5/2$. Obtain the prob that on a particular day if there was no demand by a demand had to be refused.

By data $m = 5/2 = 2.5$

$$P(x) = \frac{m^x e^{-m}}{x!} = \frac{e^{-2.5} (2.5)^x}{x!}$$

$$a) P(0) = 0.0821$$

b) If a demand had to be refused, there should have been a demand for more than 4 generators. We have to find $P(x > 4)$.

$$\begin{aligned} P(x > 4) &= 1 - P(x \leq 4) \\ &= 1 - [P(0) + P(1) + P(2) + P(3) + P(4)] \\ &= 0.1088 \end{aligned}$$

13) The probability that a news reader commits no mistake in reading the news is $1/e^3$. Find the probability that on a particular news broadcast he commits
a) only 2 mistakes b) more than 3 mistakes c) at most 3 mistakes

We consider $1/e^3 = e^{-3}$

$$P(x) = \frac{m^x e^{-m}}{x!}$$

By data $P(x=0) = e^{-3}$

$$e^{-m} = e^{-3} \Rightarrow m = 3$$

$$\therefore P(x) = e^{-3} \frac{3^x}{x!}$$

a) $P(2) = 0.224$

b) $P(x > 3) = 1 - P(x \leq 3)$
 $= 1 - [P(0) + P(1) + P(2) + P(3)]$
 $= 0.3528$

c) $P(x \leq 3) = P(0) + P(1) + P(2) + P(3)$
 $= 0.6472$

Normal Distribution

The continuous probability distribution having the P.d.f $f(x)$ given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

where $-\infty < x < \infty$, $-\infty < \mu < \infty$ & $\sigma > 0$ is known as the normal distribution.

Evidently $f(x) > 0$, $\int_{-\infty}^{\infty} f(x) dx = 1$

Mean and S.D of the normal distribution

Mean = μ

S.D = σ

Variance = σ^2

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-z^2/2} dz$$

Note :- $\int_{-\infty}^{\infty} \phi(z) dz = 1$

$$\int_{-\infty}^0 \phi(z) dz = \int_0^{\infty} \phi(z) dz = 1/2$$

3) $P(-\infty \leq z \leq \infty) = 1$

4) $P(-\infty \leq z \leq 0) = 1/2$

5) $P(0 \leq z \leq \infty)$ or $P(z \geq 0) = 1/2$

6) $P(z \leq z_1) = 0.5 + \phi(z_1) = P(z \leq 0) + P(z \leq z_1)$

7) $P(z > z_2) = 0.5 - \phi(z_2) = P(z \geq 0) - P(z \leq z_2)$

8) $P(|z| < z_1) = P(-z_1 \leq z \leq z_1) = 2\phi(z_1)$

If Evaluate the foll probabilities with the help of normal probability tables.

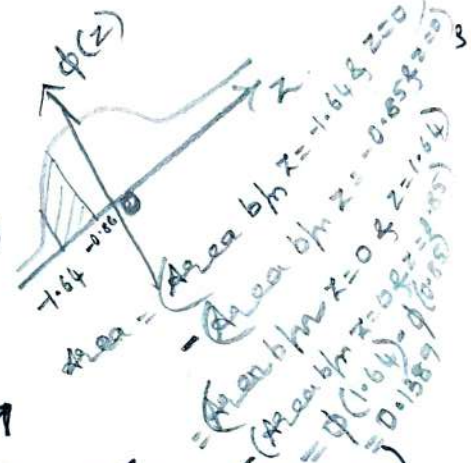
a) $P(Z > 0.85)$

b) $P(-1.64 \leq Z \leq -0.88)$

c) $P(Z \leq -2.43)$

d) $P(|Z| \leq 1.94)$

$$\begin{aligned} \text{a) } P(Z > 0.85) &= P(Z > 0) - P(Z \leq 0.85) \\ &= 0.5 - \phi(0.85) \\ &= 0.5 - 0.3023 \\ &= 0.1977 \end{aligned}$$



$$\begin{aligned} \text{b) } P(-1.64 \leq Z \leq -0.88) &= P(0 \leq Z \leq 1.64) - P(0 \leq Z \leq 0.88) \\ &= \phi(1.64) - \phi(0.88) \\ &= 0.4495 - 0.3106 \\ &= 0.1389 \end{aligned}$$

$$\begin{aligned} \text{c) } P(Z \leq -2.43) &= P(Z > 2.43) \\ &= P(Z > 0) - P(Z \leq 2.43) \\ &= 0.5 - \phi(2.43) \\ &= 0.5 - 0.4925 \\ &= 0.0075 \end{aligned}$$

$$\begin{aligned} \text{d) } P(|Z| \leq 1.94) &= P(-1.94 \leq Z \leq 1.94) \\ &= 2P(0 \leq Z \leq 1.94) \\ &= 2\phi(1.94) \\ &= 2(0.4738) \\ &= 0.9476 \end{aligned}$$

$$\begin{aligned} &= P(-1.94 \leq Z \leq 0) \\ &\quad + P(0 \leq Z \leq 1.94) \\ &= P(0 \leq Z \leq 1.94) \\ &\quad + P(0 \leq Z \leq 1.94) \\ &= 2P(0 \leq Z \leq 1.94) \\ &= 2\phi(1.94) \\ &= 0.9476 \end{aligned}$$

If X is a normal variate with mean 30 and standard deviation 5 find the probability that
a) $26 \leq X \leq 40$ b) $X > 45$

Standard normal variate $z = \frac{x - \mu}{\sigma} = \frac{x - 30}{5}$
(S.N.V)

a) $P(26 \leq X \leq 40)$

If $X = 26$, $z = -0.8$, If $X = 40$, $z = 2$

$$\begin{aligned}\therefore P(-0.8 \leq Z \leq 2) &= P(-0.8 \leq Z \leq 0) + P(0 \leq Z \leq 2) \\ &= P(0 \leq Z \leq 0.8) + P(0 \leq Z \leq 2) \\ &= \phi(0.8) + \phi(2) \\ &= 0.2881 + 0.4772 \\ &= 0.7653\end{aligned}$$

b) $P(X > 45)$

If $X = 45$, $z = 3$

$$\begin{aligned}P(Z > 3) &= P(Z > 0) - P(Z \leq 3) \\ &= 0.5 - \phi(3) \\ &= 0.5 - 0.4987 \\ &= 0.0013\end{aligned}$$

==

3. If x is normally distributed with mean 12
and S.D 4 find
a) $P(x > 20)$
b) $P(x \leq 20)$

S.n.v $z = \frac{x - \mu}{\sigma} = \frac{x - 12}{4}$

If $x = 20$, $z = 2$

$$\begin{aligned} P(z > 2) &= P(z > 0) - P(z \leq 2) \\ &= 0.5 - \phi(2) \\ &= 0.5 - 0.4772 \\ &= 0.0228 \end{aligned}$$

$$\begin{aligned} P(z \leq 2) &= P(-\infty \leq z \leq 0) + P(0 \leq z \leq 2) \\ &= 0.5 + \phi(2) \\ &= 0.5 + 0.4772 \\ &= 0.9772 \end{aligned}$$



4) The marks of 1000 students in an examination follow a normal distribution with mean 70 and S.D 5. Find the number of students whose marks will be

a) less than 65

b) more than 75

c) b/w 65 & 75

Let x represent the marks of students

$$\mu = 70, \sigma = 5 \quad \text{Hence s.d.v. } z = \frac{x - \mu}{\sigma} = \frac{x - 70}{5}$$

a) If $x = 65, z = -1$

$$\begin{aligned} P(z < -1) &= P(z > 1) \\ &= P(z > 0) - P(0 < z < 1) \\ &= 0.5 - \phi(1) \\ &= 0.5 - 0.3413 \\ &= 0.1587 \end{aligned}$$

no. of students scoring less than 65 marks
 $= 1000 \times 0.1587 = 158.7 \approx 159$

b) If $x = 75, z = 1$

$$\begin{aligned} P(z > 1) &= P(z > 0) - P(0 < z < 1) \\ &= 0.5 - \phi(1) \\ &= 0.5 - 0.3413 \\ &= 0.1587 \end{aligned}$$

no. of students scoring more than 75 marks
 $= 1000 \times 0.1587 = 158.7 \approx 159$

c) $P(-1 < z < 1) = 2 P(0 < z < 1)$
 $= 2 \phi(1) = 2(0.3413) = 0.6826$

no. of students scoring marks b/w 65 & 75
 $1000 \times 0.6826 = 682.6 \approx 683$

5. In a test on electric bulbs, it was found that the life time of a particular brand was distributed normally with an average life of 2000 hours & SD of 60 hrs. If a firm purchases 2500 bulbs find the no. of bulbs that last for
 a) more than 2100 hrs b) less than 1950 hrs c) b/w 1900 to 2100 hrs.
 By data $\mu = 2000$, $\sigma = 60$

S.N.V
$$z = \frac{x - \mu}{\sigma} = \frac{x - 2000}{60}$$

a) $P(x > 2100)$

If $x = 2100$, $z = 100/60 \approx 1.67$

$$\begin{aligned} P(x > 2100) &= P(z > 1.67) \\ &= P(z > 0) - P(0 < z < 1.67) \\ &= 0.5 - \phi(1.67) \\ &= 0.5 - 0.4525 \\ &= 0.0475 \end{aligned}$$

No. of bulbs that are likely to last for more than 2100 hours is $2500 \times 0.0475 = 118.75 \approx 119$

b) $P(x < 1950)$

If $x = 1950$, $z = -5/6 = -0.83$

$$\begin{aligned} P(x < 1950) &= P(z < -0.83) \\ &= P(z > 0.83) \\ &= P(z > 0) - P(0 < z < 0.83) \\ &= 0.5 - \phi(0.83) \\ &= 0.5 - 0.2967 \\ &= 0.2033 \end{aligned}$$

No. of bulbs to last for less than 1950 hrs is $2500 \times 0.2033 = 508.25 \approx 508$

c) $P(1900 < x < 2100)$

If $x = 1900$, $z = -1.67$ & if $x = 2100$, $z = 1.67$

$$\begin{aligned} P(1900 < x < 2100) &= P(-1.67 < z < 1.67) \\ &= 2P(0 < z < 1.67) \\ &= 2\phi(1.67) \\ &= 2 \times 0.4525 = 0.905 \end{aligned}$$

No. of bulbs to last b/w 1900 & 2100 hours
 $= 2500 \times 0.905$
 $= 2262.5 \approx 2263$

7. In an examination 7% of students score less than 35% marks and 89% of students score less than 60% marks. Find the mean & S.D if the marks are normally distributed if it is given

$$P(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-z^2/2} dz \quad \text{then } P(1.2263) = 0.39$$

$$P(1.4757) = 0.43$$

Let μ & σ be the mean and S.D of the normal distribution.

By data $P(x < 35) = 0.07$
 $P(x < 60) = 0.89$

$$S.D. = \sigma = \frac{x - \mu}{z}$$

When $x = 35$, $z = \frac{35 - \mu}{\sigma} = z_1$

$x = 60$, $z = \frac{60 - \mu}{\sigma} = z_2$

Hence

$$P(z < z_1) = 0.07$$

$$0.5 + \phi(z_1) = 0.07$$

$$\phi(z_1) = -0.43$$

$$\phi(z_1) = -\phi(1.4757)$$

$$z_1 = -1.4757$$

$$\frac{35 - \mu}{\sigma} = -1.4757$$

$$\mu - 1.4757\sigma = 35$$

$$P(z < z_2) = 0.89$$

$$0.5 + \phi(z_2) = 0.89$$

$$\phi(z_2) = 0.39$$

$$\phi(z_2) = \phi(1.2263)$$

$$z_2 = 1.2263$$

$$\frac{60 - \mu}{\sigma} = 1.2263$$

$$\mu + 1.2263\sigma = 60$$

$$\therefore \mu = 48.6538$$

$$\sigma = 9.2524$$

6. In a normal distribution 31% of the items are under 45 and 8% of the items are over 64. Find the mean and S.D of the ^{normal} distribution.

Let μ and σ be the mean and S.D of the normal distribution.

By data $P(X < 45) = 0.31$ & $P(X > 64) = 0.08$

S. n. v $Z = \frac{X - \mu}{\sigma}$

$$X = 45, \quad Z = \frac{45 - \mu}{\sigma} = Z_1$$

$$X = 64, \quad Z = \frac{64 - \mu}{\sigma} = Z_2$$

$$P(Z < Z_1) = 0.31$$

$$P(Z > Z_2) = 0.08$$

$$0.5 + \phi(Z_1) = 0.31$$

$$0.5 - \phi(Z_2) = 0.08$$

$$\phi(Z_1) = -0.19$$

$$\phi(Z_2) = 0.42$$

Referring to normal tables

$$0.1915 (\approx 0.19) = \phi(0.5)$$

$$0.4192 (\approx 0.42) = \phi(1.4)$$

$$\phi(Z_1) = -\phi(0.5)$$

$$\phi(Z_2) = \phi(1.4)$$

$$Z_1 = -0.5 \quad Z_2 = 1.4$$

$$\frac{45 - \mu}{\sigma} = -0.5$$

$$\frac{64 - \mu}{\sigma} = 1.4$$

$$\mu - 0.5\sigma = 45$$

$$\mu + 1.4\sigma = 64$$

$$\mu = 50, \quad \sigma = 10$$

ie mean = 50, S.D = 10

9) Obtain the equation of the normal probability curve that may be fitted to the foll data

Variable	6	7	8	9	10	11	12
frequency	3	6	9	13	8	5	4

The equation of the best fitting normal probability curve is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$\mu = \frac{\sum fx}{\sum f} = \frac{18 + 42 + 72 + 117 + 80 + 55 + 48}{48} = 9$$

$\nearrow 432/48$
 $\nwarrow$

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - (\mu)^2$$

$$= \frac{3(36) + 6(49) + 9(64) + 13(81) + 8(100) + 5(121) + 4(144)}{48} - 9^2$$

$$= 2.5833$$

$$\sigma = 1.6073$$

$$\therefore f(x) = \frac{1}{(1.6073)\sqrt{2\pi}} e^{-\frac{(x-9)^2}{5.1667}}$$

$$= (0.2483) e^{-\frac{(x-9)^2}{5.1667}}$$

8/6 for the full normal distribution find c and also the mean & S.D of the frequency distribution.

$$f(x) = c e^{-1/24 (x^2 - 6x + 4)}$$

We have p.d.f of the normal distribution

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{--- (1)}$$

$$\begin{aligned} \text{Consider } f(x) &= c e^{-1/24 (x^2 - 6x + 4)} \\ &= c e^{-1/24 [(x-3)^2 - 5]} \\ &= c e^{5/24} e^{-(x-3)^2/24} \\ &= c e^{5/24} e^{-(x-3)^2/2(\sqrt{12})^2} \quad \text{--- (2)} \end{aligned}$$

Comparing (1) & (2)

$\mu = 3$, $\sigma = \sqrt{12}$ provided

$$c e^{5/24} = \frac{1}{\sigma \sqrt{2\pi}} = \frac{1}{\sqrt{12} \sqrt{2\pi}} = \frac{1}{\sqrt{24\pi}}$$

$$c = \frac{e^{-5/24}}{\sqrt{24\pi}}$$

$\therefore \text{Mean} = \mu = 3$

S.D = $\sigma = \sqrt{12}$

Joint Probability distributions

The distributions associated with two random variables is referred to as joint distributions.

If X and Y are 2 discrete random variables, we define the joint probability function of X and Y by

$$P(X=x_i, Y=y_j) = p(x_i, y_j) \text{ where } p(x, y) \text{ satisfy the conditions}$$

$$a) p(x, y) \geq 0$$

$$b) \sum_x \sum_y p(x, y) = 1$$

Joint Probability Table

$x \backslash y$	y_1	y_2	...	y_n	Sum
x_1	J_{11}	J_{12}	...	J_{1n}	$p(x_1)$
x_2	J_{21}	J_{22}	...	J_{2n}	$p(x_2)$
...	...	...	...	...	...
x_m	J_{m1}	J_{m2}	...	J_{mn}	$p(x_m)$
Sum	$q(y_1)$	$q(y_2)$	...	$q(y_n)$	1

Here $\{p(x_1), p(x_2), \dots, p(x_m)\}$ & $\{q(y_1), q(y_2), \dots, q(y_n)\}$ are called marginal probability distributions of X & Y

Note :- $\frac{1}{p} p(x_1) + p(x_2) + \dots + p(x_m) = 1$
 $q(y_1) + q(y_2) + \dots + q(y_n) = 1$

$$a) \sum_{i=1}^m \sum_{j=1}^n p(x_i, y_j) = \sum_{i=1}^m \sum_{j=1}^n J_{ij} = \sum_{i=1}^m \sum_{j=1}^n P(x_i, y_j) = 1$$

Independent random variables

X & Y are ind random variables if $P(X=x_i, Y=y_j)$

$$= P(X=x_i) \cdot P(Y=y_j)$$

$$\text{ie } f(x_i) g(y_j) = J_{ij}$$

ie if each entry J_{ij} in the table is equal to the product of its marginal entries.

Expectation, Variance, Covariance and Correlation

$$\text{Expectation of } X = E(X) = \mu_x = \sum_i x_i f(x_i)$$

$$\text{Expectation of } Y = E(Y) = \mu_y = \sum_j y_j g(y_j)$$

$$\text{Expectation of } XY = E(XY) = \sum_{i,j} x_i y_j J_{ij}$$

$$\text{Covariance} = \text{COV}(X, Y) = E(XY) - E(X) \cdot E(Y)$$

$$\text{Correlation} = \rho(X, Y) = \frac{\text{COV}(X, Y)}{\sigma_X \sigma_Y}$$

$$\text{where } \sigma_X^2 = E(X^2) - \mu_x^2 \quad \sigma_Y^2 = E(Y^2) - \mu_y^2$$

$$\text{and } E(X^2) = \sum_i x_i^2 f(x_i)$$

$$\text{and } E(Y^2) = \sum_j y_j^2 g(y_j)$$

This is called variance of X and Y .

Expectation of a random variable is the average of random variable.

Covariance is a term that quantifies the extent of dependence b/w 2 random variables.

MQP

If The joint probability of 2 R.V's X & Y is as follows

$X \backslash Y$	-4	2	7
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
5	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

find marginal distribution of X & Y

Compute $E(X)$, $E(Y)$, $E(XY)$, σ_X , σ_Y , $\text{Cov}(X, Y)$, $S(X, Y)$

Marginal distribution of X Marginal distribution of Y

x_i	1	5
$f(x_i)$	$\frac{1}{2}$	$\frac{1}{2}$

y_j	-4	2	7
$g(y_j)$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$

$$\begin{aligned} \text{a) } E(X) &= \sum x_i f(x_i) = \mu_X \\ &= 1\left(\frac{1}{2}\right) + 5\left(\frac{1}{2}\right) \\ &= 3 \end{aligned}$$

$$\begin{aligned} \text{b) } E(Y) &= \sum y_j g(y_j) = \mu_Y \\ &= -4\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 7\left(\frac{1}{4}\right) \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{c) } E(XY) &= \sum x_i y_j f_{ij} \\ &= 1(-4)\left(\frac{1}{8}\right) + 1(2)\left(\frac{1}{4}\right) + (1)(7)\left(\frac{1}{8}\right) \\ &\quad + 5(-4)\left(\frac{1}{4}\right) + 5(2)\left(\frac{1}{8}\right) + 5(7)\left(\frac{1}{8}\right) \\ &= \frac{3}{2} \end{aligned}$$

$$\text{df } \sigma_x^2 = E(x^2) - \mu_x^2$$

$$\begin{aligned} E(x^2) &= \sum x_i^2 p(x_i) \\ &= (1)^2(1/2) + (5)^2(1/2) \\ &= 13 \end{aligned}$$

$$\sigma_x^2 = 13 - 9 = 4$$

$$\therefore \sigma_x = 2$$

$$\text{df } \sigma_y^2 = E(y^2) - \mu_y^2$$

$$\begin{aligned} E(y^2) &= \sum y_i^2 q(y_i) \\ &= (-4)^2(3/8) + (2)^2(3/8) \\ &\quad + 7^2(1/4) \\ &= 79/4 \end{aligned}$$

$$\sigma_y^2 = 79/4 - 1 = 15/4$$

$$\therefore \sigma_y = 4.33$$

$$\begin{aligned} \text{df } \text{COV}(x, y) &= E(xy) - E(x) \cdot E(y) \\ &= 3/2 - 3(1) \\ &= -3/2 \end{aligned}$$

$$\begin{aligned} \text{df } \rho(x, y) &= \frac{\text{COV}(x, y)}{\sigma_x \sigma_y} \\ &= \frac{-3/2}{2(4.33)} = -0.1732 \end{aligned}$$

Jan 2024.

2) The joint probability distribution table for 2 R.V's X & Y is as follows

$X \backslash Y$	-2	-1	4	5
1	0.1	0.2	0	0.3
2	0.2	0.1	0.1	0

Determine marginal probability distributions of X & Y .
 Also find Expectation of X , Y & XY , S.D of X , Y ,
 Covariance of X & Y , Correlation of X & Y .
 further V.T X & Y are dependent R.V's
 Also find $P(X+Y > 0)$.

M.D of X

x_i	1	2
$f(x_i)$	0.6	0.4

M.D of Y

y_j	-2	-1	4	5
$g(y_j)$	0.3	0.3	0.1	0.3

$$\begin{aligned} \text{a) } E(X) &= \sum_i x_i f(x_i) \\ &= 1(0.6) + 2(0.4) = 1.4 \end{aligned}$$

$$\begin{aligned} \text{b) } E(Y) &= \sum_j y_j g(y_j) \\ &= -2(0.3) + (-1)(0.3) + 4(0.1) + 5(0.3) = 1 \end{aligned}$$

$$\begin{aligned} \text{c) } E(XY) &= \sum_{i,j} x_i y_j f_{ij} \\ &= 1(-2)(0.1) + 1(-1)(0.2) + 1(4)(0) + 1(5)(0.3) \\ &\quad + 2(-2)(0.2) + 2(-1)(0.1) + 2(4)(0.1) \\ &\quad + 2(5)(0) = 0.9 \end{aligned}$$

$$\text{Hf } \sigma_x^2 = E(x^2) - \mu_x^2 = 2.2 - (1.4)^2 = 0.24$$

$$E(x^2) = \sum_i x_i^2 f(x_i)$$

$$= 1(0.6) + 4(0.4) = 2.2$$

$$\text{Hf } \sigma_y^2 = E(y^2) - \mu_y^2 = 10.6 - 1^2 = 9.6$$

$$E(y^2) = \sum_j y_j^2 g(y_j)$$

$$= 4(0.3) + 1(0.3) + 16(0.1) + 25(0.3) = 10.6$$

$$\therefore \sigma_x = 0.49, \quad \sigma_y = 3.1$$

$$\text{Hf } \text{Cov}(x, y) = E(xy) - E(x)E(y) = 0.9 - 1.4(1) = -0.5$$

$$\text{Hf } \rho(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} = \frac{-0.5}{0.49 \times 3.1} = -0.3$$

for x & y to be ind. variables, $f(x_i)g(y_j) = J_{ij}$

$$\text{But } f(x_1)g(y_1) = 0.6 \times 0.3 = 0.18$$

$$J_{11} = 0.1$$

$$\therefore f(x_i)g(y_j) \neq J_{ij}$$

$\Rightarrow x$ & y are dependent R.V's

$$x = \{x_1, x_2\} = \{1, 2\}$$

$$y = \{y_1, y_2, y_3, y_4\} = \{-2, -1, 4, 5\}$$

$$J_{11} = 0.1, J_{12} = 0.2, J_{13} = 0, J_{14} = 0.3, J_{21} = 0.2, J_{22} = 0.1,$$

$$J_{23} = 0.1, J_{24} = 0$$

$X+Y > 0$ is possible when (X, Y) take the values
 $(x_1, y_3) = (1, 4)$; $(x_1, y_4) = (1, 5)$; $(x_2, y_2) = (2, -1)$;
 $(x_2, y_3) = (2, 4)$; $(x_2, y_4) = (2, 5)$
 $\therefore P(X+Y > 0) = J_{13} + J_{14} + J_{22} + J_{23} + J_{24}$
 $= 0.5$

Model paper

Bp A fair coin is tossed thrice. The R.V's X & Y are defined as follows. $X = 0$ or 1 according as head or tail occurs on the first toss
 $Y =$ Number of heads.

a) Determine the distributions of X & Y

b) Determine the joint distribution of X & Y

c) Obtain the expectation of X , Y & XY

Also find S.D of X & Y

d) Compute Covariance & Correlation of X & Y

S	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	0	0	0	0	1	1	1	1
Y	3	2	2	1	2	1	1	0

$$X = \{0, 1\}$$

$$Y = \{0, 1, 2, 3\}$$

$$P(X=0) = 4/8 = 1/2$$

$$P(Y=0) = 1/8$$

$$P(Y=2) = 3/8$$

$$P(X=1) = 4/8 = 1/2$$

$$P(Y=1) = 3/8$$

$$P(Y=3) = 1/8$$

a) M.O of X & Y

x_i	0	1
$f(x_i)$	$\frac{1}{2}$	$\frac{1}{2}$

y_j	0	1	2	3
$g(y_j)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

b) J.O of X & Y

$$J_{ij} = P(X=x_i, Y=y_j)$$

$$J_{11} = P(X=0, Y=0) = 0 \quad [\because X=0 \text{ means head on first toss, } Y=0 \text{ means no head which is not possible}]$$

$$J_{12} = P(X=0, Y=1) = \frac{1}{8} \quad [HTT]$$

$$J_{13} = P(X=0, Y=2) = \frac{2}{8} = \frac{1}{4} \quad [HHT \text{ \& } HTH]$$

$$J_{14} = P(X=0, Y=3) = \frac{1}{8} \quad [HHH]$$

$$J_{21} = P(X=1, Y=0) = \frac{1}{8} \quad [TTT]$$

$$J_{22} = P(X=1, Y=1) = \frac{2}{8} = \frac{1}{4} \quad [THT, TTH]$$

$$J_{23} = P(X=1, Y=2) = \frac{1}{8} \quad [THH]$$

$$J_{24} = P(X=1, Y=3) = 0 \quad [\text{Impossible}]$$

$X \backslash Y$	0	1	2	3	Sum
0	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0	$\frac{1}{2}$
Sum	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

$$\begin{aligned} \text{Q. } \mu_x = E(x) &= \sum_i x_i f(x_i) \\ &= 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \mu_y = E(y) &= \sum_j y_j g(y_j) \\ &= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{3}{2} \end{aligned}$$

$$\begin{aligned} E(xy) &= \sum_{i,j} x_i y_j J_{ij} \\ &= 0 + 0 + \frac{1}{4} + \frac{2}{8} + 0 = \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \text{Q. } \sigma_x^2 &= E(x^2) - \mu_x^2 & \sigma_y^2 &= E(y^2) - \mu_y^2 \\ E(x^2) &= \sum_i x_i^2 f(x_i) & E(y^2) &= \sum_j y_j^2 g(y_j) \\ &= 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2} & &= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 4 \times \frac{3}{8} + 9 \times \frac{1}{8} \end{aligned}$$

$$\sigma_x^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

$$\sigma_y^2 = 3 - \frac{9}{4} = \frac{3}{4}$$

$$\sigma_x = \frac{1}{2}$$

$$\sigma_y = 0.8660$$

$$\begin{aligned} \text{Q. } \text{cov}(x, y) &= E(xy) - E(x)E(y) \\ &= \frac{1}{2} - \frac{3}{4} = -\frac{1}{4} \end{aligned}$$

$$\text{Q. } \rho(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{-\frac{1}{4}}{\frac{1}{2} \times \frac{\sqrt{3}}{2}} = -\frac{1}{\sqrt{3}} = -0.5774$$

4. Consider the joint distribution of X & Y given below

$X \backslash Y$	0	1	2	3	Sum
0	0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	0	$\frac{1}{2}$
Sum	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Compute

a) $P(X=1, Y=2)$

b) $P(X \geq 1, Y \leq 2)$

c) $P(X \leq 1, Y \leq 2)$

d) $P(X+Y \geq 2)$

$X = \{x_1, x_2\} = \{0, 1\}$

$Y = \{y_1, y_2, y_3, y_4\} = \{0, 1, 2, 3\}$

$J_{11} = 0, J_{12} = \frac{1}{8}, J_{13} = \frac{1}{4}, J_{14} = \frac{1}{8}, J_{21} = \frac{1}{8}, J_{22} = \frac{1}{4}, J_{23} = \frac{1}{8}, J_{24} = 0$

a) $P(X=1, Y=2) = P(x_2, y_3) = J_{23} = \frac{1}{8}$

b) $P(X \geq 1, Y \leq 2) = P(x_2, y_1) + P(x_2, y_2) + P(x_2, y_3)$
 $= J_{21} + J_{22} + J_{23}$
 $= \frac{1}{8} + \frac{1}{4} + \frac{1}{8} = \frac{1}{2}$

c) $P(X \leq 1, Y \leq 2) = P(x_1, y_1) + P(x_1, y_2) + P(x_1, y_3) + P(x_2, y_1)$
 $+ P(x_2, y_2) + P(x_2, y_3)$
 $= J_{11} + J_{12} + J_{13} + J_{21} + J_{22} + J_{23}$
 $= 0 + \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8}$
 $= \frac{7}{8}$

$$\begin{aligned}
 & \text{d) } P(X+Y \geq 2) \\
 &= P(X_1, Y_3) + P(X_1, Y_4) + P(X_2, Y_2) + P(X_2, Y_3) + P(X_2, Y_4) \\
 &= J_{13} + J_{14} + J_{22} + J_{23} + J_{24} \\
 &= \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + 0 \\
 &= \frac{3}{4}
 \end{aligned}$$

Model paper

5) Let X be a random variable with the following distribution & Y is defined to be X^2

x_i	-2	-1	1	2
$f(x_i)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Determine

a) the distribution of Y

b) the joint distribution of X & Y

c) $E(X)$ $E(Y)$ $E(XY)$

d) $\text{Cov}(X, Y)$ & $\rho(X, Y)$

Given $Y = X^2$

When $X = -2$ $Y = 4$
 $X = -1$ $Y = 1$

$X = 2$ $Y = 4$
 $X = 1$ $Y = 1$

y_j	1	4
$g(y_j)$	$\frac{1}{2}$	$\frac{1}{2}$

Distribution of Y

$X \backslash Y$	1	4
-2	0	$\frac{1}{4}$
-1	$\frac{1}{4}$	0
1	$\frac{1}{4}$	0
2	0	$\frac{1}{4}$

joint distribution of X & Y

$$\text{If } E(X) = \sum_i x_i p(x_i)$$

$$= -2\left(\frac{1}{4}\right) + -1\left(\frac{1}{4}\right) + 1\left(\frac{1}{4}\right) + 2\left(\frac{1}{4}\right) = 0$$

$$\text{If } E(Y) = \sum_j y_j q(y_j)$$

$$= 1\left(\frac{1}{2}\right) + 4\left(\frac{1}{2}\right) = \frac{5}{2}$$

$$\text{If } E(XY) = \sum_{i,j} x_i y_j J_{ij}$$

$$= (-2 \times 1 \times 0) + (-2 \times 4 \times \frac{1}{4}) + (-1 \times 4) + 0 + \frac{1}{4} + 0 + 0 + 2$$

$$= 0$$

$$\text{If } \text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$0 - 0 = 0$$

$$\text{If } \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = 0$$

Jan 2024

If The joint Probability distribution of two discrete random variables x & y is given by $p(x, y) = k(2x + y)$ where x & y are integers such that $0 \leq x \leq 2$ & $0 \leq y \leq 3$

Find the value of 'k'

Find the M.D of X & Y

S.T random variables X & Y are dependent

$X \backslash Y$	0	1	2	3	$f(x_i)$
0	0	K	2K	3K	6K
1	2K	3K	4K	5K	14K
2	4K	5K	6K	7K	22K
$g(y_j)$	6K	9K	12K	15K	42K

$\rightarrow \sum_i \sum_j P(x_i, y_j)$

WKT $\sum_i \sum_j P(x_i, y_j) = 1$

$42K = 1$

$K = \frac{1}{42}$

$x \backslash y$	0	1	2	3
0	0	$\frac{1}{42}$	$\frac{2}{42}$	$\frac{3}{42}$
1	$\frac{2}{42}$	$\frac{3}{42}$	$\frac{4}{42}$	$\frac{5}{42}$
2	$\frac{4}{42}$	$\frac{5}{42}$	$\frac{6}{42}$	$\frac{7}{42}$

M.D of Y

M.D of X

x_i	0	1	2
$f(x_i)$	$\frac{1}{7}$	$\frac{1}{3}$	$\frac{11}{21}$

y_j	0	1	2	3
$g(y_j)$	$\frac{1}{7}$	$\frac{3}{14}$	$\frac{2}{7}$	$\frac{5}{14}$

for X & Y to be independent variables

$f(x_i) \cdot g(y_j) = J_{ij}$

But $f(x_1) \cdot g(y_1) = f(0) \cdot g(0) = \frac{1}{7} \times \frac{1}{7} = \frac{1}{49}$

whereas $J_{11} = 0$

$f(x_1) \cdot g(y_1) \neq J_{11}$

Hence X & Y are dependent variables

Model QP

If the joint probability distribution of X & Y is given by
 $f(x, y) = \frac{x+y}{30}$ for $x=0, 1, 2, 3$ & $y=0, 1, 2$

find a) $P[X \leq 2, Y=1]$

b) $P[X > 2, Y \leq 1]$

c) $P[X > Y]$

$Y \backslash X$	0	1	2	$f(x)$
0	0	$\frac{1}{30}$	$\frac{1}{15}$	$\frac{1}{10}$
1	$\frac{1}{30}$	$\frac{1}{15}$	$\frac{1}{10}$	$\frac{1}{5}$
2	$\frac{1}{15}$	$\frac{1}{10}$	$\frac{2}{15}$	$\frac{3}{10}$
3	$\frac{1}{10}$	$\frac{2}{15}$	$\frac{1}{6}$	$\frac{2}{5}$
$g(y)$	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{7}{15}$	1

a) $P[X \leq 2, Y=1]$

$$= P(0, 1) + P(1, 1) + P(2, 1)$$

$$= J_{12} + J_{22} + J_{32}$$

$$= \frac{1}{30} + \frac{1}{15} + \frac{1}{10}$$

$$= \frac{1}{5}$$

b) $P[X > 2, Y \leq 1]$

$$= P(3, 0) + P(3, 1)$$

$$= J_{41} + J_{42}$$

$$= \frac{1}{10} + \frac{2}{15} = \frac{7}{30}$$

c) $P[X > Y]$

$$= P(1, 0) + P(2, 0) + P(2, 1) + P(3, 0) + P(3, 1) + P(3, 2)$$

$$= J_{21} + J_{31} + J_{32} + J_{41} + J_{42} + J_{43}$$

$$= \frac{1}{30} + \frac{1}{15} + \frac{1}{10} + \frac{1}{10} + \frac{2}{15} + \frac{1}{6}$$

$$= \frac{3}{5}$$

Q8) The joint probability distribution of 2 random variables X & Y is

$Y \backslash X$	-3	2	4
1	0.1	0.2	0.2
3	0.3	0.1	0.1

a) Are X & Y independent?

b) Evaluate $P[Y \leq 2]$

c) Evaluate $P[X+Y \leq 2]$

d) Find M.O of X & Y

e) Find $E(X)$ & $E(Y)$

$Y \backslash X$	-3	2	4	$g(y)$
1	0.1	0.2	0.2	0.5
3	0.3	0.1	0.1	0.5
$f(x)$	0.4	0.3	0.3	1

a) $f(x_i) \cdot g(y_j) = J_{ij}$, for X & Y to be ind R.V's

$$f(x_1) \cdot g(y_1) = J_{11}$$

$$0.4 \times 0.5 = 0.1$$

$$0.2 \neq 0.1$$

$$\therefore f(x_i) \cdot g(y_j) \neq J_{ij}$$

$\therefore X$ & Y are dep^{random} variables

$$b) P[Y \leq 2] = P(Y=1) = 0.1 + 0.2 + 0.2 = 0.5$$

$$c) P[X+Y \leq 2] = P(X=-3, Y=1) + P(X=-3, Y=3) \\ = 0.1 + 0.3 \\ = 0.4$$

df M.D of X

x_i	-3	2	4
$f(x_i)$	0.4	0.3	0.3

M.D of Y

y_j	1	3
$g(y_j)$	0.5	0.5

$$E(X) = \sum x_i f(x_i) \\ = -3(0.4) + 2(0.3) + 4(0.3) \\ = 0.6$$

$$E(Y) = \sum y_j g(y_j) \\ = 1(0.5) + 3(0.5) \\ = 2$$

MQP

9) Determine the value of K so that the function $f(x, y) = K|x-y|$ for $x = -2, 0, 2$ & $y = -2, 3$ represents joint probability distribution of the random variables X & Y . Also find $\text{cov}(X, Y)$

$X \backslash Y$	-2	3	$f(x_i)$
-2	0	5K	5K
0	2K	3K	5K
2	4K	K	5K
$g(y_j)$	6K	9K	15K

$$\therefore$$

$X \backslash Y$	-2	3
-2	0	$\frac{1}{3}$
0	$\frac{2}{15}$	$\frac{1}{5}$
2	$\frac{4}{15}$	$\frac{1}{15}$

$$\sum_i \sum_j P(x_i, y_j) = 1$$

$$\Rightarrow 15K = 1 \Rightarrow K = \frac{1}{15}$$

M.D of X

x_i	-2	0	2
$f(x_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

M.D of Y

y_j	-2	3
$g(y_j)$	$\frac{2}{5}$	$\frac{3}{5}$

$$E(X) = \sum_i x_i f(x_i) = -2\left(\frac{1}{3}\right) + 0 + 2\left(\frac{1}{3}\right) = 0$$

$$E(Y) = \sum_j y_j g(y_j) = -2\left(\frac{2}{5}\right) + 3\left(\frac{3}{5}\right) = 1$$

$$E(XY) = \sum x_i y_j J_{ij} = (-2 \times -2 \times 0) + (-2 \times 3 \times \frac{1}{3}) + 0 + 0 + 2 \times -2 \times \frac{1}{15} + 2 \times 3 \times \frac{1}{15}$$

$$= -\frac{8}{3}$$

$$\text{COV}(X, Y) = E(XY) - E(X)E(Y) = -\frac{8}{3} - 0 = -\frac{8}{3}$$

Q.P

10 of Random Variable X & Y have full joint distribution

$X \backslash Y$	-3	2	4
1	0.1	0.2	0.2
2	0.3	0.1	0.1

find of M.D of X & Y

by COV(X, Y)

of Correlation of X & Y

of Evaluate $P[Y \leq 2]$, $P[X + Y \leq 2]$

of Are the variables X, Y independent random variable?

of M.D of X

x_i	1	2
$f(x_i)$	0.5	0.5

M.D of Y

y_j	-3	2	4
$g(y_j)$	0.4	0.3	0.3

$$\text{b)} E(X) = \sum_i x_i p(x_i) = 1(0.5) + 2(0.5) = 1.5$$

$$E(Y) = \sum_j y_j q(y_j) = -3(0.4) + 2(0.3) + 4(0.3) = 0.6$$

$$\begin{aligned} E(XY) &= \sum x_i y_j J_{ij} \\ &= (1 \times -3 \times 0.1) + (1 \times 2 \times 0.2) + (1 \times 4 \times 0.2) + (2 \times -3 \times 0.3) \\ &\quad + (2 \times 2 \times 0.1) + (2 \times 4 \times 0.1) \\ &= 0.3 \end{aligned}$$

$$\begin{aligned} \text{Cov}(XY) &= E(XY) - E(X)E(Y) \\ &= 0.3 - (1.5)(0.6) \\ &= -0.6 \end{aligned}$$

$$\text{c)} \sigma_x^2 = E(X^2) - [E(X)]^2 \quad \sigma_y^2 = E(Y^2) - [E(Y)]^2$$

$$\begin{aligned} E(X^2) &= \sum x_i^2 p(x_i) \\ &= 1^2 \times 0.5 + 2^2 \times 0.5 \\ &= 2.5 \end{aligned}$$

$$\begin{aligned} E(Y^2) &= \sum y_j^2 q(y_j) \\ &= (-3)^2(0.4) + 2^2(0.3) + 4^2(0.3) \\ &= 9.6 \end{aligned}$$

$$\begin{aligned} \sigma_x^2 &= 2.5 - (1.5)^2 \\ &= 0.25 \end{aligned}$$

$$\begin{aligned} \sigma_y^2 &= 9.6 - (0.6)^2 \\ &= 9.24 \end{aligned}$$

$$\sigma_x = 0.5$$

$$\sigma_y = 3.0397$$

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{-0.6}{0.5 \times 3.0397} = -0.3948$$

$$\begin{aligned} \text{d)} P(Y \leq 2) &= P(Y = -3) + P(Y = 2) \\ &= 0.4 + 0.3 = 0.7 \end{aligned}$$

$$\begin{aligned} P(X+Y \leq 2) &= P(X=1, Y=-3) + P(X=2, Y=-3) \\ &= 0.1 + 0.3 = 0.4 \end{aligned}$$

$$\text{e)} p(x_1) \cdot q(y_1) = 0.5 \times 0.4 = 0.2$$

$$J_{11} = 0.1$$

$$\therefore p(x_i) \cdot q(y_j) \neq J_{ij} \quad \therefore X \text{ \& } Y \text{ are dependent R.V's}$$

If X & Y are independent random variables X taking values 2, 5, 7 with the probability $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ respectively & Y taking values 3, 4, 5 with the probability $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ find a) Joint probability distribution
 b) S.T $\text{COV}(X, Y) = 0$

M. D of X

x_i	2	5	7
$p(x_i)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

M. D of Y

y_j	3	4	5
$q(y_j)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Given X & Y are ind R.V's

$$\therefore J_{ij} = p(x_i) q(y_j)$$

$$J_{11} = p(x_1) q(y_1) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

$$J_{12} = p(x_1) q(y_2) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

$$J_{13} = p(x_1) q(y_3) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$

$$J_{21} = p(x_2) q(y_1) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

$$J_{22} = p(x_2) q(y_2) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

$$J_{23} = p(x_2) q(y_3) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

$$J_{31} = p(x_3) q(y_1) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

$$J_{32} = p(x_3) q(y_2) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

$$J_{33} = p(x_3) q(y_3) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

Joint Probability distribution

$X \backslash Y$	3	4	5	$p(x)$
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{3}{6}$
5	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{3}{12}$
7	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{3}{12}$
$q(y)$	$\frac{4}{12}$	$\frac{4}{12}$	$\frac{4}{12}$	(1)

$$E(X) = \sum_i x_i f(x_i) = 2\left(\frac{1}{2}\right) + 5\left(\frac{1}{4}\right) + 7\left(\frac{1}{4}\right) = 4$$

$$E(Y) = \sum_j y_j g(y_j) = 3\left(\frac{1}{3}\right) + 4\left(\frac{1}{3}\right) + 5\left(\frac{1}{3}\right) = 4$$

$$\begin{aligned} E(XY) &= \sum_i x_i y_j J_{ij} \\ &= (2 \times 3 \times \frac{1}{6}) + (2 \times 4 \times \frac{1}{6}) + (2 \times 5 \times \frac{1}{6}) + (5 \times 3 \times \frac{1}{12}) \\ &\quad + (5 \times 4 \times \frac{1}{12}) + (5 \times 5 \times \frac{1}{12}) + (7 \times 3 \times \frac{1}{12}) + (7 \times 4 \times \frac{1}{12}) \\ &\quad + (7 \times 5 \times \frac{1}{12}) \\ &= 16 \end{aligned}$$

$$\begin{aligned} \text{COV}(X, Y) &= E(XY) - E(X)E(Y) \\ &= 16 - 4(4) \\ &= 0 \end{aligned}$$

Q7

Given the full joint distribution of R.V's X & Y find the corresponding M.D of X & Y. Compute Covariance & Correlation of the random variables X & Y.

X \ Y	1	3	9
2	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{12}$
4	$\frac{1}{4}$	$\frac{1}{4}$	0
6	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{12}$

M.D of X

x_i	2	4	6
$f(x_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

M.D of Y

y_j	1	3	9
$g(y_j)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

$$E(X) = \sum_i x_i f(x_i) \\ = 2\left(\frac{1}{4}\right) + 4\left(\frac{1}{2}\right) + 6\left(\frac{1}{4}\right) = 4$$

$$E(Y) = \sum_j y_j g(y_j) \\ = 1\left(\frac{1}{2}\right) + 3\left(\frac{1}{3}\right) + 9\left(\frac{1}{6}\right) = 3$$

$$E(XY) = \sum_i x_i y_i J_i \\ = (2 \times 1 \times \frac{1}{8}) + (2 \times 3 \times \frac{1}{24}) + (2 \times 9 \times \frac{1}{12}) + \\ (4 \times 1 \times \frac{1}{4}) + (4 \times 3 \times \frac{1}{4}) + (4 \times 9 \times 0) + \\ (6 \times 1 \times \frac{1}{8}) + (6 \times 3 \times \frac{1}{24}) + (6 \times 9 \times \frac{1}{12}) \\ = 12$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) \\ = 12 - 4 \times 3 \\ = 0$$

$$\sigma_X^2 = E(X^2) - [E(X)]^2$$

$$E(X^2) = \sum_i x_i^2 f(x_i) \\ = 2^2\left(\frac{1}{4}\right) + 4^2\left(\frac{1}{2}\right) + 6^2\left(\frac{1}{4}\right) \\ = 18$$

$$\sigma_X^2 = 18 - 4^2 = 2$$

$$\sigma_X = \sqrt{2}$$

$$\sigma_Y^2 = E(Y^2) - [E(Y)]^2$$

$$E(Y^2) = \sum_j y_j^2 g(y_j) \\ = 1^2\left(\frac{1}{2}\right) + 3^2\left(\frac{1}{3}\right) + 9^2\left(\frac{1}{6}\right) \\ = 17$$

$$\sigma_Y^2 = 17 - 3^2 = 8$$

$$\sigma_Y = \sqrt{8}$$

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{0}{\sqrt{2} \sqrt{8}} = 0$$

Stochastic Process

A real variable X associated with every outcome is called a random variable or stochastic variable.

The family of all such random variables namely the set $\{X(t), t \in T\}$ is called stochastic process defined on sample space S with t as parameter.

Here $X_0 = X(0)$ is called as the initial state of the system.

The values assumed by the random variable $X(t)$ are called states and the set of all possible values forms the state space of the process.

If the state space of a stochastic process is discrete/continuous then it is called a discrete/continuous state process or chain.

HQP

Probability Vector

A vector $V = (v_1, v_2, \dots, v_n)$ is called a probability vector if each one of its components are non negative & their sum is equal to unity.

Ex:- $u = (1, 0)$ $v = (1/4, 1/4, 1/2)$

MQP

Stochastic matrix

A square matrix $P = p_{ij}$ having every row in the form of a probability vector is called a stochastic matrix [ie $p_{ij} \geq 0$ & $\sum p_{ij} = 1, \forall i$]

eg:-
$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

MQP

Regular stochastic matrix

A stochastic matrix P is said to be a regular stochastic matrix if all the entries of some power P^n are positive. eg:- $A = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

$A^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$

$\therefore A$ is a regular stochastic matrix [$n=2$]

if which vectors are probability vectors? No neg entry
sum=1

eg $(\frac{1}{4}, \frac{3}{2}, -\frac{1}{4}, \frac{1}{2})$

Not a probability vector because negative entry

eg $(\frac{5}{2}, 0, \frac{8}{3}, \frac{1}{6}, \frac{1}{6})$

Not a prob vector because sum is not equal to 1

eg $(\frac{1}{2}, \frac{1}{2}, \frac{1}{6}, 0, \frac{1}{4})$

is a prob vector because all entries are non-negative & sum=1

df $(3, 0, 2, 5, 3)$

$$3 + 0 + 2 + 5 + 3 = 13$$

$\therefore (3/13, 0, 2/13, 5/13, 3/13)$

is a prob vector because all entries are non-negative and sum = 1.

df which matrices are stochastic?

[square matrix
No neg entry
sum of each row = 1]

af $\begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}$

Not stochastic because it is not a square matrix.

bf $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

stochastic because it is a square matrix, non-neg, sum of each row = 1.

cf $\begin{pmatrix} 0 & 1 \\ 1/2 & 1/4 \end{pmatrix}$

not stochastic because sum of each row $\neq 1$ (last row)

df $\begin{pmatrix} 0 & 1 \\ -1/2 & 3/2 \end{pmatrix}$

No stochastic as we have negative entry

df which of the stochastic matrices are regular?

[diagonal element should not be 1
No entries to be zero
No neg entries]

$$\text{ex } A = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

Not regular since 1 lie on the main diagonal

$$\text{ex } B = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

$$B^2 = B \cdot B = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 3/8 & 3/8 & 1/4 \end{pmatrix}$$

$$B^3 = B^2 \cdot B = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 7/16 & 7/16 & 1/8 \end{pmatrix}$$

Since entries of b_{13} b_{23} are zero, it is not regular

$$\text{ex } C = \begin{pmatrix} 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$$

$$C^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

$$C^3 = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

$$C^4 = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

$$C^5 = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/8 & 1/2 & 3/8 \end{pmatrix}$$

$\therefore$ All entries are +ve and non-zero.
C is regular stochastic matrix.

MQP

4) find the unique fixed probability vector for the regular stochastic matrix

$$A = \begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

let $v = (x, y, z)$ be the unique fixed prob vector associated with P

$$\Rightarrow x + y + z = 1 \quad \text{--- (1)}$$

We have by property of regular stochastic matrix

$$vP = v$$

$$[x \ y \ z] \begin{pmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (x, y, z)$$

$$\left(y/2, \frac{3x+y}{4} + z, x/4 \right) = (x, y, z)$$

$$y = 2x, \quad \frac{3x+y}{4} + z = y, \quad x = 4z$$

$$3x - 2y + 4z = 0$$

from (1)

$$\begin{aligned} z &= 1 - x - y \\ &= 1 - x - 2x \\ &= 1 - 3x \end{aligned}$$

$$\therefore 3x - 2y + 4z = 0$$

$$3x - 2(2x) + 4(1 - 3x) = 0$$

$$3x - 4x + 4 - 12x = 0$$

$$\Rightarrow x = 4/13$$

$$y = 8/13, \quad z = 1/13$$

Hence the reqd unique fixed probability vector is $v = (4/13, 8/13, 1/13)$

5. S.T $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$ is a regular stochastic

matrix. Also find the associated unique fixed probability vector.

$$P^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$P^3 = P \cdot P^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

$$P^4 = P \cdot P^3 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$P^5 = P \cdot P^4 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{2} \end{bmatrix}$$

$\therefore P$ is a regular stochastic matrix

Let $v = (x, y, z)$ be the unique fixed prob vector associated with P

$$\Rightarrow x + y + z = 1 \quad \text{--- (1)}$$

We have $vP = v$

$$(x \ y \ z) \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = (x \ y \ z)$$

$$\left(\frac{z}{2}, x + \frac{z}{2}, y\right) = (x, y, z)$$

$$\frac{z}{2} = x, \quad x + \frac{z}{2} = y, \quad y = z$$

$$z = 2x \quad y = z = 2x$$

from ①

$$x + y + z = 1$$

$$x + 2x + 2x = 1$$

$$x = \frac{1}{5}$$

$$y = z = \frac{2}{5}$$

$\therefore v = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}\right)$ is the reqd u.f.p. vector

Ex 24

Find the unique fixed probability vector for the regular stochastic matrix

$$A = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

Let $v = (x, y, z)$ be the unique fixed prob vector associated with P

$$\Rightarrow x + y + z = 1 \quad \text{--- ①}$$

We have $vP = v$

$$(x, y, z) \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix} = (x, y, z)$$

$$\left(\frac{y}{6}, x + \frac{y}{2} + \frac{2z}{3}, \frac{y}{3} + \frac{z}{3}\right) = (x, y, z)$$

$$\frac{y}{6} = x, \quad x + \frac{y}{2} + \frac{2z}{3} = y, \quad \frac{y}{3} + \frac{z}{3} = z$$

$$y = 6x, \quad 6x - 3y + 4z = 0, \quad y = 2z$$

$$6x = 2z$$

$$z = 3x$$

from ①

$$x + 6x + 3x = 1$$

$$x = \frac{1}{10}$$

$$y = \frac{6}{10}, \quad z = \frac{3}{10}$$

$\therefore \vec{v} = \left(\frac{1}{10}, \frac{6}{10}, \frac{3}{10}\right)$ is the reqd u.p. vector

HQP

For S.T the matrix $\begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$ is a regular

stochastic matrix

$$P = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$P^3 = P \cdot P^2 = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$P^4 = P \cdot P^3 = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

$$P^5 = P \cdot P^4 = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & \frac{1}{2} & \frac{3}{8} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

$\therefore P$ is a regular stochastic matrix

Markov Chains

Let the outcomes $x_1, x_2, \dots$ of a sequence of trials satisfy the following properties
of each outcome belong to the finite set of the outcomes $\{a_1, a_2, \dots, a_m\}$.

by The outcome of any trial depends almost upon the outcome of the immediate preceding trial.

Probability p_{ij} is associated with every pair of states (a_i, a_j) that a_j occurs immediately after a_i occurs. Such a stochastic process is called a finite Markov chain. p_{ij} are called transition probabilities & they form a square matrix of order 'm' called the transition probability matrix (t.p.m) denoted by P.

$$P = [p_{ij}] = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{bmatrix}$$

The transition matrix of a Markov chain is a stochastic matrix.

$$[\because 0 \leq p_{ij} \leq 1 \text{ \& } \sum p_{ij} = 1]$$

Q.P

Higher transition probabilities

The probability that the system changes from the state a_i to the state a_j in exactly ' n ' steps is denoted by $P_{ij}^{(n)}$

ie $a_i \rightarrow a_{r_1} \rightarrow a_{r_2} \rightarrow \dots \rightarrow a_{r_{n-1}} \rightarrow a_j$

The matrix formed by $P_{ij}^{(n)}$ is called the n -step transition matrix

$$[P^{(n)}] = [P_{ij}^{(n)}]$$

Let $p^{(0)} = [p_1^{(0)}, p_2^{(0)} \dots p_m^{(0)}]$ denote the initial probability distribution at the start of the process & let $p^{(n)} = [p_1^{(n)}, p_2^{(n)} \dots p_m^{(n)}]$ denote the n^{th} step probability distribution at the end of n steps.

$$\text{ie } p^{(1)} = p^{(0)} \cdot P, \quad p^{(2)} = p^{(0)} \cdot P^2, \quad \dots \quad p^{(n)} = p^{(0)} P^n.$$

MQP

Stationary distribution of regular Markov chains

A Markov chain is said to be regular if the associated transition probability matrix ' P ' is regular

if P is a regular stochastic matrix of the Markov chain, then the sequence of n step transition matrices $P^1, P^2, \dots, P^n$ approaches the matrix V whose rows are each the unique fixed probability vector v of P .

$$P^{(n)} = P^{(0)} P^n \quad \left[P^{(n)} = \{P_1^{(n)}, P_2^{(n)} \dots P_m^{(n)}\} \right]$$

further as $n \rightarrow \infty$, $P_i^{(n)} = v_i$ where $i = 1, 2, \dots, m$

This is called the stationary distribution of the Markov chain & $v = (v_1, v_2, \dots, v_m)$ is called stationary (fixed) probability vector of the Markov chain.

QP

Irreducible

A Markov chain is said to be irreducible if every state can be reached from every other state in a finite no. of steps.

HQP

Absorbing state of a Markov chain

In a Markov chain the process reaches to a certain state after which it continues to remain in the same state. Such a state is called an absorbing state of a Markov chain.

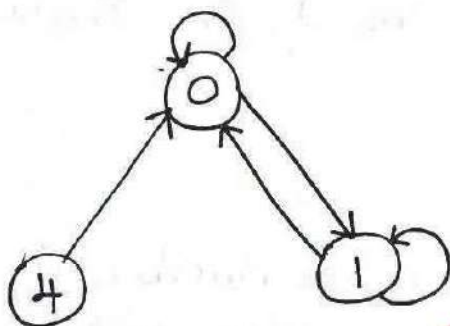
In this state, $P_{ij} = 1$ for $i = j$ & $P_{ij} = 0$ otherwise

Thus a state a_i of the Markov chain is absorbing if the i^{th} row of the t.p.m has 1 on the principal diagonal & zeroes elsewhere.

Eg:- $P = \begin{matrix} & \begin{matrix} a_1 & a_2 & a_3 & a_4 \end{matrix} \\ \begin{matrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix} \end{matrix}$

The states a_3 & a_4 are absorbing

- If Explain of Transient state
 of Recurrent state
 of Periodic state
 of absorbing state of Markov chain



	0	1	2	3	4
0	x	x	0	0	0
1	x	x	0	0	0
2	0	0	x	0	0
3	0	0	x	x	0
4	x	0	0	0	0



of A state is transient if there's a nonzero probability that the chain will leave this state and never return to it. Here 3 & 4 are transient states.

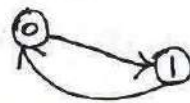
eg:- like a temporary city you might pass through once but not necessarily return

of A state is recurrent if anytime that state is left it is possible to return back to that state in future with probability 1. Here 0 & 1 are recurrent states.

eg:- like your home city you'll always return eventually.

of A state is periodic if the chain can only return to that state at multiples of some integer

starting from state 0 we only return to 0 at times $n = 2, 4, \dots$ is a periodic state with period 2.



If a state is absorbing if once entered it cannot be left.

Here state 2 is absorbing state.

(from 2 you stay in 2 with prob. 1.

Once you reach 2 you never leave

So 2 is absorbing)

Eg:- like a dead end once you're there you're stuck forever

2. The t.p.m of a Markov chain is given by

$$P = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$$

and the initial probability

distribution is $p^{(0)} = (1/2, 1/2, 0)$. find $p_{13}^{(2)}, p_{23}^{(2)}, p_{33}^{(2)}$ & $p_1^{(2)}$

$$P^2 = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix} = \begin{bmatrix} 3/8 & 1/4 & 3/8 \\ 1/2 & 0 & 1/2 \\ 11/16 & 1/8 & 3/16 \end{bmatrix}$$

$p_{13}^{(2)} = 3/8$ [It is the probability of moving from state a_1 to a_3 in 2 steps]

$p_{23}^{(2)} = 1/2$ [a_2 to a_3 in 2 steps]

$$p^{(2)} = p^{(0)} P^2 = [1/2, 1/2, 0] \begin{bmatrix} 3/8 & 1/4 & 3/8 \\ 1/2 & 0 & 1/2 \\ 11/16 & 1/8 & 3/16 \end{bmatrix}$$

$$= \left[\frac{7}{16}, \frac{1}{8}, \frac{7}{16} \right]$$

$$\therefore p_1^{(2)} = \frac{7}{16}$$

Model QP

3. P.T the Markov chain whose t.p.m is

$$P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

is irreducible. find the corresponding

stationary probability vector / fixed probability vector.

$$P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

$$P^2 = P \cdot P = \frac{1}{6} \begin{bmatrix} 0 & 4 & 2 \\ 3 & 0 & 3 \\ 3 & 3 & 0 \end{bmatrix} \cdot \frac{1}{6} \begin{bmatrix} 0 & 4 & 2 \\ 3 & 0 & 3 \\ 3 & 3 & 0 \end{bmatrix}$$

$$= \frac{1}{36} \begin{bmatrix} 18 & 6 & 12 \\ 9 & 21 & 6 \\ 9 & 12 & 15 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & 1/6 & 1/3 \\ 1/4 & 7/12 & 1/6 \\ 1/4 & 1/3 & 5/12 \end{bmatrix}$$

Since all the entries in P^2 are +ve, t.p.m 'P' is regular.

∴ The Markov chain is irreducible.

Let $v = (x, y, z)$ be the unique fixed prob vector associated with P.

$$x + y + z = 1 \quad \text{--- (1)}$$

We also have $vP = v$

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} x & y & z \end{bmatrix}$$

$$\left[\frac{y}{2} + \frac{z}{2}, \frac{2x}{3} + \frac{z}{2}, \frac{x}{3} + \frac{y}{2} \right] = [x, y, z]$$

$$\frac{y}{2} + \frac{z}{2} = x, \quad \frac{2x}{3} + \frac{z}{2} = y, \quad \frac{x}{3} + \frac{y}{2} = z$$

$$y + z = 2x$$

$$4x + 3z = 6y$$

$$2x + 3y = 6z$$

from ①

$$x + 2x = 1$$

$$3x = 1$$

$$x = \frac{1}{3}$$

sub $x = \frac{1}{3}$ in the other 2 eqⁿ

$$4\left(\frac{1}{3}\right) + 3z = 6y, \quad 2\left(\frac{1}{3}\right) + 3y = 6z$$

$$\frac{4}{3} + 3z = 6y, \quad \frac{2}{3} + 3y = 6z$$

$$4 + 9z = 18y, \quad 2 + 9y = 18z$$

$$18y - 9z = 4$$

$$-9y + 18z = 2$$

} on solving $z = \frac{8}{27}$

$$y = \frac{10}{27}$$

$$\therefore v = (x, y, z) = \left(\frac{1}{3}, \frac{10}{27}, \frac{8}{27} \right)$$

Model QP

Every year a man trades for his car for a new car. If he has Maruti, he trade it for a Ford. If he has a Ford, he trade it for a Hyundai. However if he has a Hyundai, he is just as likely to trade it for a new Hyundai as to trade it for a Maruti or a Ford. In 2014, he bought his first car which was a Hyundai. find the probability that he has

of 2016 Ford

by 2016 Hyundai

of 2016 Maruti

The state space of the system is { Maruti (M), Ford (F), Hyundai (H) }

The associated transition matrix is

$$P = \begin{matrix} & \begin{matrix} M & F & H \end{matrix} \\ \begin{matrix} M \\ F \\ H \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{matrix}$$

$$P^{(0)} = \begin{bmatrix} M & F & H \\ 0 & 0 & 1 \end{bmatrix} \quad \left\{ \begin{array}{l} \text{since in 2014 his first car was} \\ \text{Hyundai} \end{array} \right\}$$

With 2014 as the first year 2016 is 2 years after. $\therefore P^2 = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}$

$$P^{(2)} = P^{(0)} \cdot P^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}$$

a) 2016 Ford = Hyundai to Ford = $a_{32}^{(2)} = \frac{4}{9}$
 b) 2016 Hyundai = " to Hyundai = $a_{33}^{(2)} = \frac{4}{9}$
 c) 2016 Maruthi = " to Maruthi = $a_{31}^{(2)} = \frac{1}{9}$

imp
 A software engineer goes to his workplace everyday by motorbike or by car. He never goes by bike on 2 consecutive days but if he goes by car on a day he is equally likely to go by car or by bike on the next day. Find the transition matrix for the chain of the mode of transport he uses. If car is used on the first day of the week find the probability that
 a) bike is used
 b) car is used after 4 days (or on fifth day)

Here state space of the system is

$S = \{ \text{motor bike (B)}, \text{Car (C)} \}$

The associated transition matrix is

$$P = \begin{matrix} & \begin{matrix} B & C \end{matrix} \\ \begin{matrix} B \\ C \end{matrix} & \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{matrix}$$

Since the car is used on the first day, the prob that bike is used is zero and that car is used is 1. $\therefore p(0) = (0, 1)$

The probability distribution of the mode of transport after 4 days is given by

$$P^{(4)} = P^{(0)} P^4$$

$$P^4 = \begin{pmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{pmatrix}$$

$$\therefore P^{(4)} = (0, 1) \begin{pmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{pmatrix}$$

$$= \left(\frac{5}{16}, \frac{11}{16} \right)$$

$P^{(4)}$
Bike $P^{(4)}$
Car

After 4 days (or on fifth day) prob of using a bike = $5/16$ & prob of using a car = $11/16$

Jan 24

5. Three boys A, B, C are throwing ball to each other. A always throws the ball to B & B always throws the ball to C. C is just as likely to throw the ball to B as to A. If C was the first person to throw the ball find the probabilities that after three throws

a) A has the ball

b) B has the ball

c) C has the ball

$$P = \begin{matrix} & \begin{matrix} A & B & C \end{matrix} \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \end{matrix}$$

Initially if C has the ball, the associated initial probability vector is $p^{(0)} = (0, 0, 1)$

$$p^{(3)} = p^{(0)} P^3 = (0, 0, 1) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

$$= \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2} \right)$$

$\begin{matrix} p_A^{(3)} & p_B^{(3)} & p_C^{(3)} \end{matrix}$

a) A has the ball = $p_A^{(3)} = \frac{1}{4}$

b) B has the ball = $p_B^{(3)} = \frac{1}{4}$

c) C has the ball = $p_C^{(3)} = \frac{1}{2}$

If a gambler's luck follows a pattern. If he wins a game, the probability of winning the next game is 0.6. However if he loses a game, the probability of losing the next game is 0.7. There is an even chance of gambler winning the first game. If so

a) What is the prob of he winning the second game?

b) What is the prob of he winning the third game?

c) In the long run, how often he will win?

$$P = \begin{matrix} & \begin{matrix} W & L \end{matrix} \\ \begin{matrix} W \\ L \end{matrix} & \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix} \end{matrix}$$

Probability of winning first game is $\frac{1}{2}$

Initial prob vector $p^{(0)} = (\frac{1}{2}, \frac{1}{2})$

af $p^{(1)} = p^{(0)} \cdot P = (\frac{1}{2}, \frac{1}{2}) \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix} = (0.45, 0.55)$ $p_W^{(1)}$ $p_L^{(1)}$

$\therefore$ prob of winning second game 0.45

by $p^{(2)} = p^{(0)} P^2 = (0.435, 0.565)$ $p_W^{(2)}$ $p_L^{(2)}$

$\therefore$ prob of winning third game $= 0.435$

cf $v = (x, y)$, $x + y = 1$ 1

We have $vP = v$

$$[x \ y] \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix} = [x \ y]$$

$$0.6x + 0.3y = x$$

$$0.3y = 0.4x$$

$$y = \frac{0.4}{0.3} x$$

from eq ①

$$x + y = 1$$

$$x + \frac{0.4}{0.3} x = 1$$

$$0.7x = 0.3$$

$$x = \frac{3}{7}$$

$$\begin{aligned} y &= 1 - x \\ &= 1 - \frac{3}{7} \\ &= \frac{4}{7} \end{aligned}$$

$$v = [\frac{3}{7}, \frac{4}{7}] = [p^{(W)}, p^{(L)}]$$

$\therefore$ In the long run he wins $\frac{3}{7}$ of the time

imp Jan 24

Q. Each year, a man trades his car for a new car in 3 brands of the popular company. If he has a swift he trades it for Dzire. If he has a dzire he trades it for a Wagon R. If he has a Wagon R he is just as likely to trade it for a new Wagon R or for a Dzire or a swift one. In 2020 he bought his first car which was Wagon R. Find the probability that he has

a) 2022 Wagon R

b) 2022 swift

c) 2023 dzire

d) 2023 Wagon R

The state space of the system is $\{S, D, W\}$ where S is swift, D is Dzire, W is Wagon R.

The associated transition matrix is

$$P = \begin{matrix} & \begin{matrix} S & D & W \end{matrix} \\ \begin{matrix} S \\ D \\ W \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{matrix}$$

$P^{(0)} = [0, 0, 1]$ as his first car was Wagon R in 2020.

With 2020 as the first year 2022 is 2 years later.

$$P^{(2)} = P^{(0)} P^2 = [0, 0, 1] \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix} \\ = \begin{matrix} P_S^{(2)} & P_D^{(2)} & P_W^{(2)} \end{matrix} \begin{bmatrix} \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}$$

$$\text{of 2022 wagon R} = 4/9$$

$$\text{of 2022 swift} = 1/9$$

With 2020 as first year, 2023 is after 3 years

$$P^{(3)} = P^{(0)} P^3 = [0, 0, 1] \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/9 & 4/9 & 4/9 \\ 4/27 & 7/27 & 16/27 \end{bmatrix}$$

$$= \left[\overset{P_S^{(3)}}{4/27}, \overset{P_B^{(3)}}{7/27}, \overset{P_A^{(3)}}{16/27} \right]$$

$$\text{of 2023 derive} = 7/27$$

$$\text{of 2023 Wagon R} = 16/27$$

HQ

Three boys A, B, C are throwing ball to each other. A is just as likely to throw the ball to B as to C. B always throws the ball to A and C is just as likely to throw the ball to A as to B. find the probability that C has the ball after three throws if now A has the ball.

$$P = \begin{matrix} & \begin{matrix} A & B & C \end{matrix} \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{matrix}$$

Initially A has the ball. The associated initial probability vector is $P^{(0)} = [1, 0, 0]$

since C has the ball after 3 throws,

$$p^{(3)} = p^{(0)} p^3 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/4 & 3/8 & 3/8 \\ 3/4 & 1/4 & 0 \\ 3/8 & 3/8 & 1/4 \end{bmatrix}$$

$$= \begin{bmatrix} \overset{P_A^{(3)}}{1/4}, \overset{P_B^{(3)}}{3/8}, \overset{P_C^{(3)}}{3/8} \end{bmatrix}$$

$\therefore$ Probability that C has the ball after 3 throws is $\underline{\underline{p_C^{(3)} = 3/8}}$

10. The student's study habits are as follows. If he studies on one night he is 60% sure not to study on next night. On the other hand, if he does not study on to night he is 80% sure to study next night. Write the transition probability matrix for his chain of study. In the long run how often he studies on Monday night. What is the probability that he does not study on Friday night?

The state space of the system is $\{S = \text{student studies}, T = \text{student does not study}\}$

$$P = \begin{matrix} & \begin{matrix} S & T \end{matrix} \\ \begin{matrix} S \\ T \end{matrix} & \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix} \end{matrix}$$

Since he studies on Monday night is taken

$$p^{(0)} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Friday is four days after Monday

$$\begin{aligned}
 P^{(4)} &= P^{(0)} P^4 \\
 &= (1, 0) \begin{pmatrix} 0.5824 & 0.4176 \\ 0.5568 & 0.4432 \end{pmatrix} \\
 &= (0.5824, 0.4176)
 \end{aligned}$$

Probability that the student does not study on Friday night is 0.4176.

Let $v = (x, y)$, $x + y = 1$ — (1)

We have $VP = v$

$$(x, y) \begin{pmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{pmatrix} = (x, y)$$

In long run,

$$0.4x + 0.8y = x$$

$$0.6x + 0.2y = y$$

from (1) $y = 1 - x$

$$\therefore 0.4x + 0.8(1 - x) = x$$

$$0.4x + 0.8 - 0.8x = x$$

$$0.8 = 1.4x$$

$$x = 0.5714$$

$$y = 1 - 0.5714 = 0.4286$$

$\therefore$ In the long run, student studies 57.14% of the time.

MQP

11. A salesman territory consists of 3 cities A, B & C. He never sells in the same city on successive days. If he sells in city A then the next day he sells in city B. However if he sells in either B or C then the next day he is twice as likely to sell in city A as in other city. In long run, how often does he sell in each of the cities.

Let A, B, C be the 3 cities as given

$$P = \begin{matrix} & \begin{matrix} A & B & C \end{matrix} \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 2/3 & 0 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix} \end{matrix}$$

Let $v = (x, y, z)$, $x + y + z = 1$ — (1)

We have $vP = v$

$$(x, y, z) \begin{bmatrix} 0 & 1 & 0 \\ 2/3 & 0 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix} = (x, y, z)$$

$$\left[\frac{2}{3}y + \frac{2}{3}z, x + \frac{1}{3}z, \frac{2}{3}x \right] = [x, y, z]$$

$$\frac{2}{3}y + \frac{2}{3}z = x, \quad x + \frac{1}{3}z = y, \quad \frac{2}{3}x = z$$

$$y = 3z$$

$$\frac{2}{3}(3z) + \frac{2}{3}(z) = x$$

$$2z + \frac{2}{3}z = x$$

$$8z = 3x \quad \therefore x = \frac{8}{3}z$$

We have $x + y + z = 1$

$$\frac{8}{3}x + 3x + z = 1$$

$$20x = 3$$

$$x = 3/20$$

$$x = \frac{8}{3} z = \frac{8}{3} \left(\frac{3}{20} \right) = 8/20 = 2/5$$

$$y = 3x = 3 \left(\frac{3}{20} \right) = 9/20$$

In long run in city

$$A \rightarrow x \rightarrow \frac{2}{5} \times 100 = 40\%$$

$$B \rightarrow y \rightarrow \frac{9}{20} \times 100 = 45\%$$

$$C \rightarrow z \rightarrow \frac{3}{20} \times 100 = 15\%$$

Hence in long run, salesman sells 40%, 45% & 15% in cities A, B & C.

12. Three boys x, y, z are throwing a ball to each other. x always throws the ball to y & y always throws the ball to z . But z is ^{just} as likely to throw the ball to y or as to x . Write t.p.m if z is the first person to throw the ball find the prob that x has the ball after fourth

throw?

Given x, y, z be three boys.

$$P = \begin{matrix} & \begin{matrix} x & y & z \end{matrix} \\ \begin{matrix} x \\ y \\ z \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \end{matrix}$$

Initially, z throws the ball, the associated initial probability vector is $p^{(0)} = (0, 0, 1)$

since it is to be taken after 4 throws,

$$p^{(4)} = p^{(0)} P^4 = (0, 0, 1) \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix} = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4} \right)$$

$\begin{matrix} p_x^{(4)} & p_y^{(4)} & p_z^{(4)} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{matrix}$

Probability that x has the ball after fourth throw = $\frac{1}{4}$

HQP

Every year a man trades his car for a new car. If he has a Maruthi, he trades it for an ambassador. If he has an ambassador, he trades it for a Santro. However, if he has a Santro, he is just as likely to trade it for a new Santro as to trade it for a Maruthi or an Ambassador. In 2000, he bought his first car which was a Santro. find the probability that he has

of 2002 Santro

by 2002 Maruti

of 2003 Ambassador

by 2003 Santro

The state space of the system is $\{M, A, S\}$ where
M - maruti, A - ambassador, S - Santro

The associated transition matrix is

$$P = \begin{matrix} & \begin{matrix} M & A & S \end{matrix} \\ \begin{matrix} M \\ A \\ S \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{matrix}$$

Since his first car was Santro the initial probability is $p^{(0)} = [0 \ 0 \ 1]$

With 2000 as the first year, 2002 is 2 years later

$$\begin{aligned} p^{(2)} &= p^{(0)} \cdot P^2 = [0 \ 0 \ 1] \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix} \\ &= \begin{bmatrix} \overset{p_M^{(2)}}{\frac{1}{9}} & \overset{p_A^{(2)}}{\frac{4}{9}} & \overset{p_S^{(2)}}{\frac{4}{9}} \end{bmatrix} \end{aligned}$$

of Prob that he has 2002 Santro = $\frac{4}{9}$

by Prob that he has 2002 Maruti = $\frac{1}{9}$

With 2000 as the first year, 2003 is 3 years later

later

$$p^{(3)} = p^{(0)} \cdot P^3 = [0 \ 0 \ 1] \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{1}{27} & \frac{7}{27} & \frac{16}{27} \end{bmatrix}$$

cf Prob that he has 2003 ambassador = $7/27$

cf Prob that he has 2003 santro = $16/27$

==



Statistical Inference - I

⇒ Statistical hypothesis:- A statistical hypothesis is an assumption about the parameters of population distribution.

⇒ Null hypothesis:- It is a statistical hypothesis which is tested for possible rejection or acceptance under the assumption that it is true.

• It is denoted by H_0 .

⇒ Alternative hypothesis:- It is a statistical hypothesis which is complementary to the Null hypothesis.

• It is denoted by H_1 .

For eg:- To test whether the population has specified mean μ_0 then H_0 is : $\mu = \mu_0$

H_1 : $\mu \neq \mu_0$ (i.e., $\mu > \mu_0$ & $\mu < \mu_0$)

• This alternative hypothesis is known as a two-tailed alternative.

⇒ H_1 is $\mu > \mu_0$, this alternative hypothesis is known as Right tail alternative.

⇒ H_1 is $\mu < \mu_0$, this alternative hypothesis is known as Left tail alternative.

*Note:- The setting of alternative hypothesis is very important since it enables us to decide whether we have to use a right-tailed, left-tailed or a two-tail test.

1) A population parameter is a statistical measure or constant obtained from the population.

For eg:- Population mean denoted by μ
population variance denoted by σ^2

2) A sample statistic is a statistical measure computed from the sampled observations.

For eg:- Sample mean denoted by $\bar{x}$
Sample variance denoted by s^2

3) The process of drawing or obtaining samples is called sampling if $n \geq 30$. The sampling is known as large sampling and if $n < 30$, the sampling is known as small sampling where 'n' denotes sample size.

* Standard error :- The standard deviation of the sampling distn of a statistic is known as standard error of that statistic. The reciprocal of the std error is called precision.

S.No.	Statistic	Standard error	Std. Normal Variate.
1	Mean ($\bar{x}$)	$\frac{\sigma}{\sqrt{n}}$	$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
2	Proportions (P)	$\sqrt{\frac{pq}{n}}$	$z = \frac{p - P}{\sqrt{pq/n}}$
3	Difference of means ($\bar{x}_1 - \bar{x}_2$)	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
4	Difference of Proportions ($p_1 - p_2$)	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$	$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$

* Errors in Sampling:- The main objective in sampling theory is to draw valid inferences about the population parameters on the basis of the sample results.

In practice we decide to accept or reject the hypothesis after examining a sample from it and we are liable to commit two types of errors

1) Type - I error:- • Type-I error is rejecting H_0 when it is true i.e.,

$$P \{ \text{rejecting } H_0 / H_0 \text{ is true} \} = \alpha \quad \text{where,}$$

α is known as the size of Type-I error.

(or)

$$P \{ \text{rejecting } H_0 / H_0 \} = \alpha$$

2) Type - II error:- • Type-II error is accepting H_0 when H_0 is wrong i.e.,

$$P \{ \text{accepting } H_0 / H_0 \text{ is false} \} = \beta$$

(or)

$$P \{ \text{accepting } H_0 / H_1 \} = \beta$$

where, β is known as size of Type-II error.

Note:- Power of the test = $1 - \beta$

* critical region and level of significance:-

- A region in the sample space S which amounts to rejection of H_0 is known as critical region or region of rejection.

- If ω is the critical region and if 't' is the value of the statistic based on the random sample then the probability of $P(t \in \omega / H_0) = \alpha$.
- the probability α that a ~~any~~ of the statistic belongs to the critical region is known as the level of significance.
- In other words level of significance is the size of Type-I error.
- In general we use the level of significance α as 5%.
- The level of significance is always fixed in advance before collecting the sample information.

* Confidence limits :- • Confidence limits shows how accurate the estimation of mean is likely to be.

- Confidence limits are the lowest and the highest number at the end of the confidence interval.

* Confidence interval :- • It is a range around a measurement that conveys how precise the measurement is.

- Analysis often used confidence intervals that contain either 95% or 99% of expected observations.
- The ends of the confidence intervals are called Confidence limits.

* One-tailed and Two-tailed Test :-

- In any test, the critical region is represented by a portion of the area under the probability curve of the test statistic.
- A test of any statistical hypothesis where the alternative hypothesis, H_1 is one-tailed (right-tailed or left-tailed) is called a one-tailed test.

For eg:- consider a test for testing the mean of a population

Then, $H_0: \mu = \mu_0$

$H_1: \mu > \mu_0$ (Right-tailed)

CR

$$H_0 = \mu = \mu_0$$

$$H_1 = \mu < \mu_0 \text{ (left tailed)}$$

These are single tailed / one-tailed test.

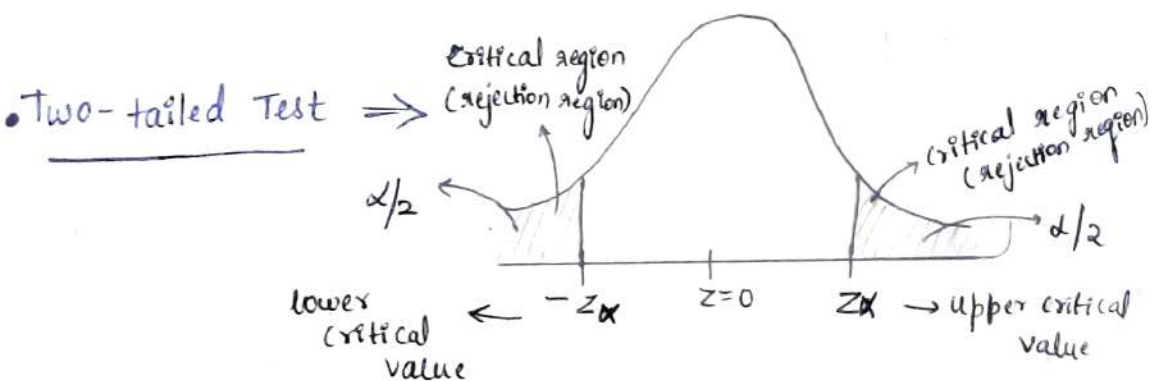
In a right tailed test, the critical region lies entirely in the right tail. while for the left tailed test, the critical region lies entirely in the left tail.

- A list of any statistical hypothesis where the alternative hypothesis, H_1 is two-tailed is called a two-tailed test.

For eg:- $H_0 : \mu \neq \mu_0$

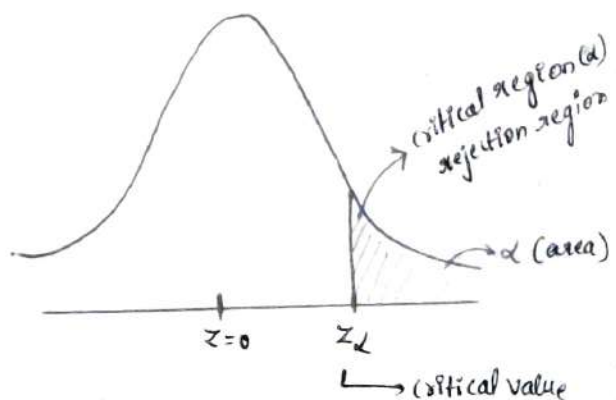
$$H_1 : \mu > \mu_0 \text{ and } \mu < \mu_0 \text{ (two-tailed)}$$

In a two-tailed test, the critical region is given by the portion of area lying in both tails of the test statistic.



$$P(Z > Z_{\alpha/2}) = \alpha/2$$

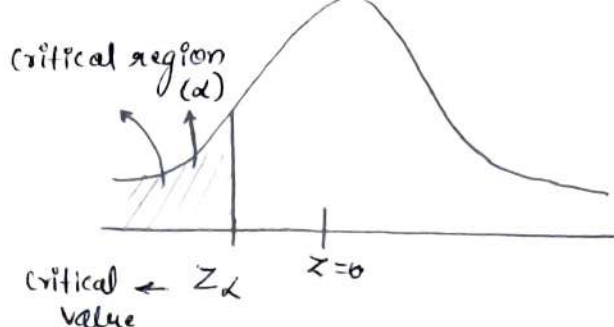
- Right-tailed Test $\Rightarrow$



$$P(Z > Z_{\alpha}) = \alpha$$

• Left-tailed test $\Rightarrow$

$$P(Z < Z_{\alpha}) = \alpha$$



Critical / Significant value : That value of the test statistic which separates the critical or rejection region and acceptance region is called critical / significant value.

> It depends on - i) The level of significance
 ii) Alternative hypothesis [which gives whether it is right tailed / left / two-tailed test]

> The standard normal variate corresponding to the statistic t is given by $z = \frac{t - E(t)}{S.E(t)}$ follows $N(0,1)$.

$S.E(t) \Rightarrow$ Standard error of t .

* ^{imp} The value of z under the NULL hypothesis H_0 is known as Test statistic

Note : The critical value of a test statistic at level of significance α is given by Z_{α} .

* Reference Table for Z_{α} :-

Critical Value (Z_{α})	Level of Significance (α)			
	1 %	2 %	5 %	10 %
Two-tailed Test	$ Z_{\alpha} = 2.58$	$ Z_{\alpha} = 2.33$	$ Z_{\alpha} = 1.96$	$ Z_{\alpha} = 1.645$
Right-tailed Test	2.33	1.96	1.645	1.28
Left-tailed Test	-2.33	-1.96	-1.645	-1.28

Note :- Confidence limit $\Rightarrow t \pm Z_{\alpha} [S.E(t)]$

* Procedure for Testing of Hypothesis :-

Step 1: Set up the null hypothesis

Step 2: Set up the Alternative hypothesis

[This will enable us to decide whether we have to perform a right-tailed, left-tailed or a two-tailed test]

Step 3: Fix the level of significance α i.e., a suitable level of significance is selected in advance.

Step 4: Test Statistic $\rightarrow$ compute the test statistic

$$Z = \frac{t - E(t)}{S.E(t)} \text{ under } H_0.$$

Step 5: Conclusion $\rightarrow$ compare Z and Z_α , Z_α is critical value obtained from the table at the given level of significance and the type of test.

- If $|Z| > Z_\alpha$, the null hypothesis H_0 is rejected.
- If $|Z| < Z_\alpha$, the null hypothesis H_0 is accepted.

* Test of Significance for Large Samples:-

- 1) Test of Significance for single proportion.
- 2) Test of significance for difference of proportion.
- 3) Test of significance for single mean.
- 4) Test of significance for difference of mean.

1) Test of Significance for single proportion:-

$\rightarrow$ Here, the test statistic is given by $Z = \frac{t - E(t)}{S.E(t)}$ i.e.,

$$Z = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1)$$

Here, $Q = 1 - P$ & n is the sample size.

Note: The confidence limit for the test of single proportion is given by:

$$p \pm z_{\alpha} (\sqrt{pq/n})$$

& Confidence interval is given by $(p - z_{\alpha} \sqrt{pq/n}, p + z_{\alpha} \sqrt{pq/n})$

→ The confidence coefficient is the confidence level stated as a proportion, rather than as a percentage.

If confidence level = 95%, confidence coefficient = $\frac{95}{100} = 0.95$

$$\Rightarrow 1 - \alpha = 95\% \text{ or } 0.95$$

$$\Rightarrow \alpha = 0.05$$

Q) A coin is tossed 10,000 times & it turns up head 5195 times, can the coin be regarded as an unbiased one? Test at 5% level of significance

solⁿ: Given, $n = 10,000$; $p = \frac{x}{n} = \frac{5195}{10,000} = 0.5195$

To test whether the coin is unbiased or not i.e., to test whether $P = \frac{1}{2}$ or not.

Step 1 $\Rightarrow H_0 : P = 0.5$

Step 2 $\Rightarrow H_1 : P \neq 0.5$ [$P > 0.5$ & $P < 0.5$] = (Two tailed test)

Step 3 $\Rightarrow \alpha = 0.05$ & for Two tailed test $\boxed{z_{\alpha} = 1.96}$

Step 4 $\Rightarrow$ compute test statistic:

[from the table]

$$z = \frac{p - P}{\sqrt{pq/n}} = \frac{0.5195 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{10,000}}}$$

$$p = 0.5$$

$$q = 1 - 0.5 = 0.5$$

$$\therefore |z| = 3.9$$

Step 5: $|Z| > Z_\alpha \Rightarrow$ we reject the H_0

$\therefore$ The coin is not an unbiased one.

Q2) A coin was tossed 400 times and head turned up 216 times. Test the hypothesis that the coin is unbiased at 5% level of significance

Soln:- Given, $n = 400$; $p = \frac{x}{n} = \frac{216}{400} = 0.54$

Step 1 To test whether the coin is an unbiased one

Let the coin be an unbiased coin.

Step 1 $\Rightarrow H_0 : P = 0.5$

Step 2 $\Rightarrow H_1 : P \neq 0.5$ [Two-tailed test]

Step 3 $\Rightarrow \alpha = 0.05$ & for Two-tailed test, $Z_\alpha = 1.96$

Step 4 $\Rightarrow$ Compute test statistic:

$$Z = \frac{\hat{p} - P}{\sqrt{PQ/n}} = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} = 1.6$$

$$\therefore |Z| = 1.6$$

Step 5: $|Z| < Z_\alpha \Rightarrow$ we accept the H_0

$\therefore$ The coin is an unbiased one.

Q3) In a random sample of 125 cool drinkers, 68 said they prefer thumbs up to pepsi. Test the null hypothesis $P = 0.5$ against the alternative hypothesis $P > 0.5$.

Soln:- Given, $n = 125$, $p = \frac{x}{n} = \frac{68}{125} = 0.544$

step 1 $\Rightarrow H_0 : P = 0.5$

step 2 $\Rightarrow H_1 : P > 0.5$ [Right tailed Test]

step 3 $\Rightarrow$ Let $\alpha = 0.05$ & for right tailed test,

$$Z_{\alpha} = 1.645$$

step 4 $\Rightarrow$ Compute Test Statistic,

$$Z = \frac{p - P}{\sqrt{PQ/n}} = \frac{0.544 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{125}}} = 0.98$$

$\therefore |Z| = 0.98$

step 5 $\Rightarrow |Z| < Z_{\alpha} \Rightarrow$ we accept the H_0

$\therefore P = 0.5 //$

Q4 A coin is tossed 10,000 times and heads turn up 540 times. Decide on the hypothesis that the coin is unbalanced at 1% level of significance.

Soln:- Given, $n = 10,000$, $p = \frac{x}{n} = \frac{540}{10,000} = 0.054$

step 1 $\Rightarrow H_0 : P = 0.5$

step 2 $\Rightarrow H_1 : P \neq 0.5$ [Two-tailed test]

step 3 $\Rightarrow \alpha = 0.01$ & for two-tailed test, $Z_{\alpha} = 2.58$

step 4 $\Rightarrow$ compute test statistic

$$Z = \frac{p - P}{\sqrt{PQ/n}} = \frac{0.054 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{10,000}}} = -89.2$$

$$\therefore |z| = 89.2$$

Step 5: $|z| \leq z_{\alpha} \Rightarrow$ we ^{reject} accept the H_0

$\therefore$ The coin is unbiased one

* imp
Q5) A die is thrown 9000 times and a throw of 3 or 4 is observed 3240 times. Show that the die cannot be regarded as an unbiased one and find the limits between which the probability of a throw of 3 or 4 lies.

Soln: Given, $n = 9000$, $p = \frac{x}{n} = \frac{3240}{9000} = 0.36$

To test whether die is unbiased or not. Suppose the die is unbiased then, $P(\text{getting 3 or 4}) = P(\text{getting 3}) + P(\text{getting 4})$
 $= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

Step 1 $\Rightarrow H_0: p = 0.333$

Step 2 $\Rightarrow H_1: p \neq 0.33$

Step 3 $\Rightarrow \alpha = 0.05$ & for two-tailed Test, $|z_{\alpha}| = 1.96$

Step 4 $\Rightarrow$ compute test statistic

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}} = \frac{0.36 - 0.33}{\sqrt{\frac{0.33 \times 0.67}{9000}}}$$

$$q = 1 - p \\ = 1 - 0.33 \\ = 0.67$$

$$\therefore |z| = 6.05$$

Step 5 $\Rightarrow |z| \leq z_{\alpha} \Rightarrow$ we ^{reject} accept the H_0

$\therefore p \neq 0.33 \Rightarrow$ The die is not an unbiased one

Confidence limit for single proportion is given by $p \pm Z_{\alpha} \left(\sqrt{\frac{pq}{n}} \right)$

$$\Rightarrow 0.36 \pm 1.96 \left(\sqrt{\frac{0.33 \times 0.67}{9000}} \right)$$

$$\Rightarrow 0.36 \pm 1.96 (5) \times 10^{-3}$$

$$\Rightarrow 0.36 \pm 0.01$$

$$\Rightarrow 0.35 \text{ to } 0.37$$

confidence interval is given by $\Rightarrow (0.35, 0.37)$

*imp.
Q6) A die was thrown 9000 times and a throw of 5 or 6 was observed 3240 times. On the assumption of random throwing, do the data indicate an unbiased die at 1% level of significance.

soln: Given, $n = 9000$, $p = \frac{x}{n} = \frac{3240}{9000} = 0.36$

To Test whether die is unbiased, the die is unbiased

$$\begin{aligned} \text{then, } P(\text{getting 5 or 6}) &= P(\text{getting 5}) + P(\text{getting 6}) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{2}{6} = \frac{1}{3} \\ &= 0.33 \end{aligned}$$

Step 1 $\Rightarrow H_0: P = 0.33$

Step 2 $\Rightarrow H_1: P \neq 0.33$

Step 3 $\Rightarrow \alpha = 0.01$ & for two tailed Test, $Z_{\alpha} = 2.58$

Step 4 $\Rightarrow$ Compute test statistic

$$Z = \frac{p - P}{\sqrt{Pq/n}} = \frac{0.36 - 0.33}{\sqrt{\frac{0.33 \times 0.67}{9000}}}$$

$$\therefore |z| = 6$$

Step 5: $|z| > z_{\alpha} \Rightarrow$ we reject the H_0

$\therefore P \neq 0.33 \Rightarrow$ The die is not the unbiased one.

Q7) In 324 throws of six faced die, an odd number turned up 181 times. Is it reasonable to think that die is unbiased, one at 5% level of significance.

solⁿ: Given, $n = 324$, $p = \frac{x}{n} = \frac{181}{324} = 0.558$

To test whether the die is unbiased, suppose that the die is unbiased then, probability of getting an odd number $\Rightarrow P(\text{getting an odd number}) = \frac{3}{6} = \frac{1}{2} = 0.5$
 $\therefore$ [the die is unbiased]

Step 1 $\Rightarrow H_0: P = 0.5$

Step 2 $\Rightarrow H_1: P \neq 0.5$ [TTT]

Step 3 $\Rightarrow \alpha = 0.05$ and for two-tailed test $z_{\alpha} = 1.96$

Step 4 $\Rightarrow$ compute test statistic

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.558 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{324}}} = 2.08$$

$$\therefore |z| = 2.08$$

Step 5 $\Rightarrow |z| > z_{\alpha} \Rightarrow$ we reject H_0

$\therefore P \neq 0.5 \Rightarrow$ the die is not an unbiased one

II. Test of significance for difference of proportions

- Let p_1 & p_2 be the sample proportions with sample sizes n_1 & n_2 respectively.
- Let P_1 & P_2 be the Population proportions.
- Let $E(p_1) = P_1$ & $E(p_2) = P_2$.
- To test whether there is significant difference between the proportions: $E(p_1 - p_2) = P_1 - P_2$

Suppose there is no significant difference between the proportions

$$\text{i.e., } P_1 - P_2 = 0 \quad \text{or} \quad P_1 = P_2$$

$$\text{i.e., } H_0: P_1 = P_2 = P \text{ (say)}$$

The test statistic is given by $Z = \frac{t - E(t)}{S.E(t)}$

$$\Rightarrow Z = \frac{(p_1 - p_2) - E(p_1 - p_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

$$\Rightarrow Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

$$\Rightarrow \frac{(p_1 - p_2) - 0}{\sqrt{\frac{PQ}{n_1} + \frac{PQ}{n_2}}}$$

under
 H_0 ,
 $P_1 = P_2$

$$\Rightarrow Z = \frac{(p_1 - p_2)}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where, $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ //

Q1) Before an increase in excise duty on, 800 persons out of a sample of 1000 persons were found to be tea drinkers. After an increase in duty, 800 people were tea drinkers in a sample of 1200 people. Using standard error of proportion, state whether there is a significant decrease in the consumption of tea after the increase in excise duty.

Soln:- Given, $n_1 = 1000$ & $n_2 = 1200$

$$p_1 = \frac{800}{1000} = 0.8 \quad \& \quad p_2 = \frac{x_2}{n_2} = \frac{800}{1200} = 0.67$$

To test whether there is significant decrease in the consumption of tea after an increase in excise duty. Suppose that there is no significant difference in the consumption of tea before and after an increase in excise duty.

Step 1 $\Rightarrow H_0: P_1 = P_2$

Step 2 $\Rightarrow H_1: P_1 > P_2$ [RTT]

Step 3 $\Rightarrow \alpha = 0.05\%$ & for a right tailed test, $\boxed{Z_{\alpha} = 1.645}$

Step 4 $\Rightarrow$ compute the test statistic:

$$z = \frac{p_1 - p_2}{\sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0.8 - 0.67}{\sqrt{0.73(0.27) \left(\frac{1}{1000} + \frac{1}{1200} \right)}}$$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{1000(0.8) + 1200(0.67)}{1000 + 1200}$$

$$= 0.73 //$$

$$q = 1 - p$$

$$= 1 - 0.73$$

$$= 0.27 //$$

$$\therefore Z = 6.83 //$$

$$\boxed{\therefore |Z| = 6.83}$$

step 5 $\Rightarrow |z| > z_{\alpha}$, we reject the H_0

$\Rightarrow P_1 > P_2$ or $P_2 < P_1$, therefore, there is a significant decrease in the consumption of tea after an increase in excise duty.

Q2) Random samples of 400 men & 600 women were asked whether they would like to have a flyover near their residence. 200 men & 325 women were in favour of the proposal. Test the hypothesis that the proportions of men & women in favour of the proposal are same against that they are not, at 5% level of significance.

Given, $n_1 = 400$ & $n_2 = 600$

$$p_1 = \frac{200}{400} = 0.5 \quad \& \quad p_2 = \frac{325}{600} = 0.542$$

To test whether the proportions are same:-

step 1 $\Rightarrow H_0 : P_1 = P_2$

step 2 $\Rightarrow H_1 : P_1 \neq P_2$ [TTT]

step 3 $\Rightarrow \alpha = 0.05\%$ & for a TTT, $z_{\alpha} = 1.96$

step 4 $\Rightarrow$ compute the test statistic

$$z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0.5 - 0.54}{\sqrt{0.524 \times 0.476 \left(\frac{1}{400} + \frac{1}{600} \right)}}$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{(400) 0.5 + 600(0.54)}{1000}$$

$$= 0.524 //$$

$$Q = 1 - P$$

$$= 0.476 //$$

$$= -1.24 //$$

$$\therefore |z| = 1.24$$

step 5: $|Z| < Z_{\alpha} \Rightarrow$ we accept H_0

$\Rightarrow P_1 = P_2 \Rightarrow$ the proportions of men and women in favour of the proposal are same.

imp.

Q3) In a city A 20% of a random sample of 900 school boys had a certain slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference b/w the proportions, significant at 5% significance level.

Soln: Given, $n_1 = 900$ & $n_2 = 1600$
 $x_1 = 180$ $x_2 = 1288$ 296
 $p_1 = \frac{180}{900} = 0.2$ $p_2 = \frac{x_2}{n_2} = \frac{296}{1600} = 0.185$

To Test whether there is significant difference b/w the proportions. Suppose there is no significant difference b/w the proportions (ie, $P_1 = P_2$)

step 1 $\Rightarrow H_0: P_1 = P_2$

step 2 $\Rightarrow H_1: P_1 \neq P_2$ [TTT]

step 3 $\Rightarrow \alpha = 5\%$ & for TTT $Z_{\alpha} = 1.96$

step 4 $\Rightarrow$ compute the test statistic

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.2 - 0.185}{\sqrt{0.8096 \times 0.1904 \left(\frac{1}{900} + \frac{1}{1600}\right)}} = 0.917 //$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$Q = 1 - P = 0.8096$$

$$= \frac{900(0.2) + 1600(0.185)}{2500}$$

$$= 0.1904$$

$$\therefore |Z| = 0.917$$

step 5: $|Z| < Z_{\alpha}$, we accept H_0

$P_1 = P_2 \Rightarrow$ There is no significant difference between the proportions.

imp.

Q4) In a sample of 600 men from a certain city, 450 are smokers. In another sample of 900 men in another city, 450 are smokers. Do they indicate that the cities are significantly different w.r.t the habit of smoking among men. Test at 5% significance level.

soln: given, $n_1 = 600$ & $n_2 = 900$

$$p_1 = \frac{x_1}{n_1} = \frac{450}{600} = 0.75 \quad \& \quad p_2 = \frac{x_2}{n_2} = \frac{450}{900} = 0.5$$

To test whether cities are significantly different w.r.t the habit of smoking. Suppose is that the difference is not significant (i.e. $P_1 = P_2$)

step 1 $\Rightarrow H_0 : P_1 = P_2$

step 2 $\Rightarrow H_1 : P_1 \neq P_2$ [TTT]

step 3 $\Rightarrow$ for $\alpha = 5\%$ & for TTT $\boxed{Z_{\alpha} = 1.96}$

step 4 $\Rightarrow$ compute the test statistic

$$Z = \frac{p_1 - p_2}{\sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0.75 - 0.5}{\sqrt{0.6 \times 0.4 \left(\frac{1}{600} + \frac{1}{900} \right)}} = \frac{0.25}{0.026} = 9.681$$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{600(0.75) + 900(0.5)}{1500}$$

$$= 0.6 //$$

$$Q = 1 - 0.6 = 0.4 //$$

$$\therefore \boxed{|Z| = 9.68}$$

Step 5 : $|Z| > Z_{\alpha}$, we reject H_0 .

$\therefore P_1 \neq P_2 \Rightarrow$ The cities are significantly different w.r.t habit of smoking.

Q5) One type of aircraft is found to develop engine trouble in 5 flights out of a total of 100 and another type in 7 flights out of a total of 200 flights. Is there a significant difference in the two types of aircrafts so far as engine defects are concerned?
Test at 5% significance level.

Soln Given, $n_1 = 100$ & $n_2 = 200$

$$p_1 = \frac{x_1}{n_1} = \frac{5}{100} = 0.05 \quad p_2 = \frac{x_2}{n_2} = \frac{7}{200} = 0.035$$

To Test whether there is a significant difference b/w two types of aircrafts. Suppose there is no significant difference b/w the aircrafts [$P_1 = P_2$]

Step 1 $H_0 : P_1 = P_2$

Step 2 $\Rightarrow H_1 : P_1 \neq P_2$ [TTT]

Step 3 $\Rightarrow \alpha = 5\% \Rightarrow \boxed{Z_{\alpha} = 1.96}$

Step 4 $\Rightarrow$ Test statistic $\Rightarrow Z = \frac{p_1 - p_2}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow \text{①}$

Where $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{12}{300} = 0.04 \rightarrow \text{②}$

$$\therefore Q = 1 - P = 1 - 0.04 = 0.96$$

substituting ② in ①,

$$Z = \frac{0.05 - 0.035}{\sqrt{0.04 \times 0.96 \left(\frac{1}{100} + \frac{1}{200} \right)}} = 0.625$$

$$\therefore |z| = 0.625$$

Step 5: $|z| < z_{\alpha} \Rightarrow$ we accept H_0

$$\text{Hence, } p_1 = p_2$$

$\therefore$ There is no significant difference in two types of air crafts.

Test of Significance for single Mean:-

Suppose we want to test whether the sample has been drawn from a population with mean μ and variance σ^2 i.e., to test there is no significant difference in the mean (To check whether $\mu = \mu_0$) then the test statistic is given by:-

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

which follows $\sim N(0, 1)$

Note:- ① If the population standard deviation σ is unknown then we use its estimate $\hat{\sigma} = s$ where s is the sample standard deviation then the test statistic is given by:-

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

Q1) A random sample of 400 students is found to have a mean height 171.38 cm. Can it be regarded as a sample from a large population with mean height 171.17 cm and standard deviation 3.30 cm. Test at 5% level of significance and also find the confidence limit.

Sol:- Given, $n = 400$, $\bar{x} = 171.38$ cm

$$\mu = 171.17 \text{ cm}$$

$$\sigma = 3.30 \text{ cm}$$

To test whether the sample is drawn from a population with mean $\mu = 171.17$.

step 1 $\Rightarrow H_0: \mu = 171.17$

step 2 $\Rightarrow H_1: \mu \neq 171.17$ [TTT]

step 3 $\Rightarrow$ For $\alpha = 5\%$ & for a TTT, $Z_{\alpha} = 1.96$

step 4 $\Rightarrow$ Compute Test Statistic:

$$Z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} = \frac{171.38 - 171.17}{(3.3/\sqrt{400})} = 1.27$$

step 5 $\Rightarrow$

$$\therefore |Z| = 1.27$$

$|Z| < Z_{\alpha}$, we accept H_0

i.e., $\mu = 171.17 \Rightarrow$ the sample is taken from a large population with mean 171.17

Confidence limit $\Rightarrow \bar{x} \pm Z_{\alpha} (S.E.(\bar{x}))$

$$= \bar{x} \pm 1.96 (\sigma/\sqrt{n})$$

$$= 171.38 \pm 1.96 \left(\frac{3.3}{\sqrt{400}} \right)$$

$$= 171.38 \pm 0.3234$$

$$= 171.7 \text{ \& } 171.06 // (171.06, 171.7)$$

Q2) A sample of 900 members has a mean 3.4 cm and standard deviation 2.61 cm. Is the sample taken from a large population of mean 3.25 cm and standard deviation 2.61 cm. If the population is normal & its mean is unknown find the 95% confidence limits for the mean and also test the hypothesis.

Q11 : Given, $n = 900$, $\bar{x} = 3.4$, $s = 2.61$

$$\mu = 3.25, \sigma = 2.61$$

To test whether the sample is drawn or sample is taken from a population with mean $\mu = 3.25$ cm.

step 1 $\Rightarrow H_0 : \mu = 3.25$

step 2 $\Rightarrow H_1 : \mu \neq 3.25$ [TTT]

step 3 $\Rightarrow$ For $\alpha = 95\%$ & for a TTT, $\boxed{z_\alpha = 1.96}$
confidence level

$$1 - \alpha = 95\%$$

$$\alpha = 5\%$$

step 4 $\Rightarrow$ Compute Test Statistic :

$$Z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} = \frac{3.4 - 3.25}{(2.61/\sqrt{900})} = 1.72$$

$$\therefore \boxed{|Z| = 1.72}$$

step 5 $\Rightarrow |Z| < z_\alpha$, we accept H_0

i.e., $\mu = 3.25 \Rightarrow$ the sample is drawn from the population of mean $\mu = 3.25$ cm.

Confidence limit $\Rightarrow t \pm z_\alpha (S.E(t))$

$$= \bar{x} \pm 1.96 (\sigma/\sqrt{n})$$

$$= 3.4 \pm 1.96 \left(\frac{2.61}{\sqrt{900}} \right)$$

$$= 3.4 \pm 0.171$$

$$= 3.57 \text{ \& } 3.23 //$$

$$(3.23, 3.57)$$

Q3) An oceanographer want to check whether the depth of the ocean in a certain region is 57.4 has previously recorded. What can we conclude at level of significance $\alpha = 0.05$, if the reading taken at 40 random locations in the given region yielded in a mean of 59.1 with standard deviation of 5.2.

Soln Given, $n = 40$, $\bar{x} = 59.1$, $s = 5.9$
 $\mu = 57.4$

Step 1 $\Rightarrow H_0: \mu = 57.4$

Step 2 $\Rightarrow H_1: \mu \neq 57.4$ [TTT]

Step 3 $\Rightarrow$ For $\alpha = 5\%$ & for a TTT, $z_{\alpha} = 1.96$

Step 4 $\Rightarrow$ Compute the test statistic

$$Z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} = \frac{59.1 - 57.4}{(5.2/\sqrt{40})}$$

$$= 2.07 //$$

$$\therefore |Z| = 2.07$$

Step 5 $\Rightarrow |Z| > z_{\alpha}$, we reject H_0

i.e., $\mu \neq 57.4 //$ ~~the~~

Q) In an exit poll enquiry it was revealed that 600 voters in one locality and 400 votes from another locality favoured 55% & 48% respectively a particular party to come to power. Test the hypothesis that there is a difference in the locality in respect of the opinion

Soln Given, $n_1 = 600$ & $n_2 = 400$

$$p_1 = 55\% = \frac{55}{100} = 0.55 \text{ \& } p_2 = \frac{48}{100} = 0.48$$

To test whether there is a difference in the locality w.r.t the opinion i.e. to test whether the difference in proportions are significantly different or not.

Suppose that there is no significant differences in the opinion proportions (i.e., $P_1 = P_2$)

$$\text{Step 1} \Rightarrow H_0: P_1 = P_2$$

$$\text{Step 2} \Rightarrow H_1: P_1 \neq P_2 \text{ [TTT]}$$

$$\text{Step 3} \Rightarrow \text{Let } \alpha = 5\% \text{ \& for a TTT, } \boxed{Z_\alpha = 1.96}$$

Step 4 $\Rightarrow$ compute test statistic

$$Z = \frac{p_1 - p_2}{\sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0.55 - 0.48}{\sqrt{0.522(0.478) \left(\frac{1}{600} + \frac{1}{400} \right)}}$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 2.17 //$$

$$= \frac{600(0.55) + 400(0.48)}{1000}$$

$$= 0.522$$

$$q = 1 - 0.522 = 0.478$$

$$\boxed{\therefore |Z| = 2.17}$$

Step 5: $|Z| > Z_\alpha$, we reject H_0 .

$P_1 \neq P_2 \Rightarrow$ there is a significant difference in the locality (opinion).

Q5) A random sample of 100 recorded deaths in past years showed an average life span of 71.8 years. Assuming population S.D of 8.9 yrs; does the data indicate the data indicate that the average span today is greater than 70 years, use 5% level of significance.

soln Given, $n = 100$, $\bar{x} = 71.8$, $\sigma = 8.9$, $\alpha = 5\%$.

Step 1 $\Rightarrow H_0: \mu = 70$

Step 2 $\Rightarrow H_1: \mu > 70$ [RTT]

Step 3 $\Rightarrow$ For $\alpha = 5\%$ & for RTT, $Z_\alpha = 1.645$

Step 4 $\Rightarrow$ Compute test statistic.

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{71.8 - 70}{8.9 / \sqrt{100}} = 2.02$$

$$\therefore |Z| = 2.02$$

Step 5 $\Rightarrow |Z| > Z_\alpha$, we reject H_0 .

i.e., $\mu > 70$

Q6) A sample of 200 tyres is taken from a lot. The mean life of tyres is found to be 40,000 kms with S.D of 3200 kms. Is it reasonable to assume the mean life of tyres in the lot of 41,000 kms? Also establish 95% confidence limits within which mean life of tyres in the lot is expected to lie.

[Hint: TTT, $|Z| =$

sdn Given, $n = 200$, $\bar{x} = 40000$, $\mu = 41000$, $\sigma = 3200$

Step 1: $H_0: \mu = 41000$

Step 2: $H_1: \mu \neq 41000$ (TTT)

Step 3: for 95% confidence level $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$ & for TTT $Z_{\alpha} = 1.96$

Step 4: Test statistic

$$Z = \frac{\bar{x} - \mu}{\left(\sigma / \sqrt{n} \right)} = \frac{40000 - 41000}{\left(3200 / \sqrt{200} \right)} = -4.4194$$

$$\therefore |Z| = 4.4194$$

Step 5: $|Z| > Z_{\alpha}$, we reject H_0
i.e., $\mu \neq 41000$

7) A stenographer claims that she can type at rate of 120 wpm. Can we reject her claim on the basis of 100 trials in which she demonstrates a mean of 116 wpm with SD of 15 words? Use 5% level of significance

Solⁿ: Given, $n = 100$, $\bar{x} = 116$, $\mu = 120$, $\sigma = 15$.

Step 1: $H_0: \mu = 120$

Step 2: $H_1: \mu \neq 120$ [TTT]

Step 3: for $\alpha = 5\%$ and for TTT $Z_{\alpha} = 1.96$

Step 4: compute test statistic.

$$Z = \frac{\bar{x} - \mu}{\left(\sigma / \sqrt{n} \right)} = \frac{116 - 120}{\left(15 / \sqrt{100} \right)} = -2.66 \quad \therefore |Z| = 2.66$$

Step 5: $|Z| > Z_{\alpha} \Rightarrow$ we reject H_0 i.e., $\mu \neq 120$.

(IV) Test of Significance for difference of means.

Let $\bar{x}_1$ be the mean of a sample of size n_1 from a population with mean μ_1 & variance σ_1^2 .

Let $\bar{x}_2$ be the mean of a sample size n_2 from a population with mean μ_2 & variance σ_2^2 .

To test whether there is significant difference between the mean.

Then, the test statistic is given by:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

if the population variances σ_1^2 & σ_2^2 are known.

$$(or) Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

if the population variances σ_1^2 & σ_2^2 are unknown.

If common variance σ^2 is given then the test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$(or) Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

if the common variance σ is known.

$$(or) \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

if the common standard deviation σ is unknown.

* Q1) The means of 2 single large sample of 1000 & 2000 members are 67.5 and 68 respectively. Can the sample be regarded as drawn from the same population of standard deviation 2.5.

Sol: Given, $n_1 = 1000$, $n_2 = 2000$

$$\bar{x}_1 = 67.5, \bar{x}_2 = 68, \sigma = 2.5$$

Test whether both the sample are drawn from the same population.

Suppose both the sample are drawn from same population. i.e., ($\mu_1 = \mu_2$)

Step 1: $H_0 = \mu_1 = \mu_2$

Step 2: $H_1: \mu_1 \neq \mu_2$ [TTT]

Step 3: Let $\alpha = 5\%$ for the TTT $Z_\alpha = 1.96$

Step 4: Compute test statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{67.5 - 68}{2.5 \sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$= -5.1$$

$$|z| = 5.1$$

Step 5: $|z| > z_{\alpha} \Rightarrow$ we reject H_0

$\therefore \mu_1 \neq \mu_0 \rightarrow \therefore$ Sample are not drawn from the same population.

Q2) A random sample 1200 house holders from one town gives the mean income as ₹ 500 per month with S.D of ₹ 70 and another random sample of 1000 house holders from another town gives the mean income of ₹ 600 per month with S.D of ₹ 90. Test whether the mean income of house holders from 2 town differ significantly (or) not. Test at 5% level of significant.

Soln: Given, μ_1 μ_2

$n_1 = 1200$	$n_2 = 1000$
$\bar{x}_1 = 500$	$\bar{x}_2 = 600$
$s_1 = 70$	$s_2 = 90$

Test whether the mean income of 2 house holders differ significantly or not ($\mu_1 = \mu_2$)

Step 1: $H_0 : \mu_1 = \mu_2$

Step 2: $H_1 : \mu_1 \neq \mu_2$ (TTT)

Step 3: $\alpha = 5\% = 0.05 \Rightarrow \boxed{z_{\alpha} = 1.96}$

Step 4: Compute test Statistics

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{500 - 600}{\sqrt{\frac{70^2}{1200} + \frac{90^2}{1000}}} = -28.64$$

$$\therefore |z| = 28.64$$

Step 5: $|z| > z_{\alpha} \Rightarrow$ we reject H_0

$\therefore \mu_1 \neq \mu_2 \Rightarrow$ the mean income of the two house holders differ significantly.

**** Q3)** A random sample of 1000 workers in a company has mean wage of ₹ 50 per day & S.D of ₹ 15. Another sample of 1500 workers from another company has mean wage of ₹ 45 per day & S.D of ₹ 20. Does the mean rate of wages vary b/w the two companies at 95% confidence limits? Also find the confidence limits.

Soln: Given, $n_1 = 1000$, $\bar{x}_1 = 50$ & $S_1 = 15$
 $n_2 = 1500$, $\bar{x}_2 = 45$ & $S_2 = 20$

Suppose the mean rate of wages b/w the companies do not differ significantly.

Step 1: $H_0: \mu_1 = \mu_2$

Step 2: $H_1: \mu_1 \neq \mu_2$

Step 3: for confidence limit = 95%, $1 - \alpha = 0.95$

$$\Rightarrow \boxed{\alpha = 0.05} \quad \& \text{ for TTT } \boxed{z_{\alpha} = 1.96}$$

Step 4: compute the test statistic.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{50 - 45}{\sqrt{\frac{15^2}{1000} + \frac{20^2}{1500}}} = 7.13$$

Step 5: $|z| > z_{\alpha} \Rightarrow$ we reject H_0

$\therefore$ mean rate of wages from two companies.

Q4) A sample of 100 bulbs produced by a company showed mean life of 1190 hr and S.D 90 hr also a sample of 75 bulbs produced by a company showed mean life of 1230 hr and S.D 120 hr is there a difference b/w the mean life of the bulbs produced by the two companies at:

(i) 5% level of significance (ii) 1% level of significance

→ Given, $n_1 = 100$, $n_2 = 75$, $\bar{x}_1 = 1190$, $\bar{x}_2 = 1230$, $s_1 = 90$, $s_2 = 120$

Step 1: $H_0: \mu_1 = \mu_2$

Step 2: $H_1: \mu_1 \neq \mu_2$ (TTT)

Step 3: case ① for $\alpha = 5\%$ for TTT $\boxed{z_\alpha = 1.96}$

case ② for $\alpha = 1\%$ for TTT $\boxed{z_\alpha = 2.58}$

Step 4: Compute the test statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1190 - 1230}{\sqrt{\frac{90^2}{100} + \frac{120^2}{75}}}$$

$$Z = \frac{-40}{16.52} = -2.42$$

$$\therefore |z| = 2.42$$

Step 5: $|z| > z_\alpha \Rightarrow$ we reject H_0 for 5%. $2.17 > 1.96$

$|z| < z_\alpha \Rightarrow$ we accept H_0 for 1%. $2.17 < 2.58$

At 5% level - there is a statistically significant difference b/w the mean lives at 5% level.

At 1% level the evidence is not strong enough to conclude a difference at 1% level.

Q3) To find confidence limits & interval.

Mean rate of wages from the two companies differ significantly

$$\text{confidence limit} \Rightarrow t \pm z_{\alpha} (S.E(t))$$

$$\Rightarrow (\bar{x}_1 - \bar{x}_2) \pm 1.96 \left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

$$\Rightarrow (50 - 45) \pm 1.96 \left(\sqrt{\frac{(15)^2}{1000} + \frac{(20)^2}{1506}} \right)$$

$$\Rightarrow 3.63 \text{ \& } 6.37$$

$$\text{confidence interval } (3.63, 6.37) //$$

Q5) The mean height of 50 male students who showed above average participation in dg athletics was 68.2 inches with S.D of 2.5 inches while 50 males student who showed no interest in such participation had a mean height of 67.5 inches with S.D of 2.8 inches. Test the hypothesis that the male student who participate in athletics are taller than other male students.

Soln Given, $n_1 = 50$ $n_2 = 50$
 $\bar{x}_1 = 68.2$ $\bar{x}_2 = 67.5$
 $s_1 = 2.5$ $s_2 = 2.8$

To test whether the average height of students who showed interest in athletics is greater than the average height of student who showed no interest in athletics.

suppose that their average height are same ($\mu_1 = \mu_2$)

step 1: $H_0: \mu_1 = \mu_2$

step 2: $H_1: \mu_1 > \mu_2$ (RTT)

step 3: $\alpha = 0.05$ for RTT $\boxed{z_{\alpha} = 1.645}$

step 4: compute test statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{68.2 - 67.5}{\sqrt{\frac{(2.5)^2}{50} + \frac{(2.8)^2}{50}}} = 1.32$$

$$\boxed{\therefore |z| = 1.32}$$

step 5 : $|z| < z_{\alpha} \rightarrow$ we accept H_0 i.e., $\mu_1 = \mu_2$

i.e., the height of students who participated in athletic is not taller than the students who not participated in athletic.

Q6) In a certain factory there are 2 independent processes manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 g with a S.D of 12. while the corresponding figures in a sample of 400 items from the other process are 124 g & 14.

① obtain the standard error of difference b/w the 2 sample means
is this difference significant?

② Also find the 99% confidence limits for the difference in the average weights of items produced by the two process respectively.

Soln:- Given, $n_1 = 250$, $n_2 = 400$, $\bar{x}_1 = 120$, $\bar{x}_2 = 124$, $s_1 = 12$, $s_2 = 14$

step 1 : $H_0 : \mu_1 = \mu_2$

step 2 : $H_1 : \mu_1 \neq \mu_2$ (TTT)

step 3 : for $\alpha = 5\%$ for TTT $\Rightarrow \boxed{z_{\alpha} = 1.96}$

step 4 : compute the test statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{120 - 124}{\sqrt{\frac{12^2}{250} + \frac{14^2}{400}}} = \frac{-4}{1.032} = -3.875$$

$$\boxed{\therefore |z| = 3.875}$$

steps : $|z| > z_{\alpha} \Rightarrow$ we reject H_0

$\Rightarrow \mu_1 \neq \mu_2$

$\therefore$ The difference of 4 b/w the sample mean is statistically significant.

Confidence limit for 99%.

$$\Rightarrow t \pm z_{\alpha} (S.E(t))$$

$$= (\bar{x}_1 - \bar{x}_2) \pm 2.58 \left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

$$= (120 - 124) \pm 2.58 \left(\sqrt{\frac{12^2}{250} + \frac{14^2}{400}} \right)$$

$$= -4 \pm 2.58 (1.032)$$

$$= -6.658, -1.33$$

$$= (1.33, 6.658)$$

STATISTICAL INFERENCE - II

Central limit theorem:

If $\bar{x}$ is the mean of a random sample of size n taken from the population with mean μ and variance σ^2 , then the limiting form of the distribution of

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

as $n \rightarrow \infty$ is the standard normal distribution

$$Z(0.1)$$

- ① An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with mean = 800 hrs with standard deviation of 40 hrs. Find the probability that the random sample of 60 bulbs will have average sample life of 775 hrs.

Sol: Given:

$$\mu = 800 \quad \text{and} \quad \sigma = 40 \quad n = 60$$

To find $P(\bar{x} < 775)$

$$P(\bar{x} < 775) = P\left(\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{775 - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$P\left(Z < \frac{775 - 800}{\frac{40}{\sqrt{60}}}\right)$$

$$= P(Z < -2.5)$$

$$= 0.00621$$

$$0.5 - P(0.5)$$

$$0.5 - 0.49379$$

mode 3 (STAT)

1 → AC

shift → 1 → 5 → 1

② Travelling between two campuses of University in a city via shuttle bus takes an average 28 min with a standard deviation of 5 min. In a given week, A bus transported passengers 40 times. What is the probability that the average transport time was more than 30 mins.

Sol: Given:

$$\mu = 28 \quad \sigma = 5 \quad n = 40$$

To find $P(\bar{x} > 30)$

$$P(\bar{x} > 30) = P\left[\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{30 - \mu}{\sigma/\sqrt{n}}\right] \quad [\text{By CLT}]$$

$$= P(Z > 2.53)$$

$$= 1 - P(Z \leq 2.53)$$

$$= 0.006$$

③ An unknown distribution has a mean of 90 and a standard deviation of 15. Samples of size $n = 25$ are drawn randomly from the population find the probability that the sample mean is between 85 & 92.

Sol: Given:

$$\mu = 90$$

$$\sigma = 15$$

$$n = 25$$

To find $P(85 < \bar{x} < 92)$

$$\begin{aligned}
 P[85 < \bar{x} < 92] &= P\left[\frac{85-\mu}{\sigma/\sqrt{n}} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{92-\mu}{\sigma/\sqrt{n}}\right] \text{ by CLT} \\
 &= P\left[\frac{85-90}{15/\sqrt{25}} < Z < \frac{92-90}{15/\sqrt{25}}\right] \\
 &= P[-1.67 < Z < 0.67] \\
 &\quad P(\text{Right}) - P(\text{Left}) \\
 &= P(Z < 0.67) - P(Z < -1.67) \\
 &= 0.7
 \end{aligned}$$

Imp *

* ④ Evaluate $P(50 < \bar{x} < 56)$ where $\bar{x}$ represents the mean of a random sample of size 100 from an infinite population with mean $\mu = 53$ & $\sigma^2 = 400$ [variance]

Sol: Given:

$$\mu = 53, \quad \sigma = 20, \quad n = 100$$

$$\begin{aligned}
 P(50 < \bar{x} < 56) &= P\left(\frac{50-\mu}{\sigma/\sqrt{n}} < Z < \frac{56-\mu}{\sigma/\sqrt{n}}\right) \\
 &= P\left(\frac{50-53}{20/\sqrt{100}} < Z < \frac{56-53}{20/\sqrt{100}}\right) \\
 &= P(-1.5 < Z < 1.5) \\
 &= P(Z < 1.5) - P(Z < -1.5) \\
 &\quad 0.93317 - 0.066807 \\
 &= 0.87
 \end{aligned}$$

- ⑤ In a recent study reported on Flurry Blog the mean age of tab users is 34 years. Suppose that the S.D is 15 years take a sample of size $n=100$ and using central limit theorem the sample mean age is more than 30 years.

Sol: Given:

$$\mu = 34 \quad \sigma = 15 \quad n = 100$$

To find $P(\bar{x} > 30)$

$$P(\bar{x} > 30) = P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{30 - \mu}{\sigma/\sqrt{n}}\right) \quad [\text{by CLT}]$$

$$= P\left[Z > \frac{30 - 34}{15/\sqrt{100}}\right]$$

$$= P(Z > -2.67)$$

$$= 1 - P(Z \leq -2.67)$$

$$= 0.996$$

- ⑥ The weight of workers in a large factory are normally distributed with mean 68 kgs and SD of 3kgs. If 80 samples consisting 35 workers each are chosen, how many of 80 samples will have the mean between 67 and 68.25 kgs

Sol: Given $\mu = 68 \quad \sigma = 3$

Sample size $n = 35$

$$P(67 < \bar{x} < 68.25) = P\left(\frac{67 - \mu}{\sigma/\sqrt{n}} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{68.25 - \mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(\frac{67 - 68}{3/\sqrt{35}} < Z < \frac{68.25 - 68}{3/\sqrt{35}}\right)$$

$$= P(-1.97 < Z < 0.493)$$

$$= P(Z < 0.493) - P(Z < -1.97)$$

$$= 0.66$$

$\therefore$ out of 80 samples, no. of samples having mean between 67 & 68.25

is $80 \times P(67 < \bar{x} < 68.25)$

$$80 \times 0.66$$

$$52.8$$

≈ 53 samples

(7) A random sample size 64 is taken from an infinite population having mean 112 and variance 144. Using Central limit theorem, find the probability of getting the sample mean $\bar{x}$ greater than 114.5.

Sol: $n = 64$;

$$\mu = 112 \quad \sigma^2 = 144$$

$$\sigma = \sqrt{144}$$

$$\sigma = 12$$

To find $P(\bar{x} > 114.5)$ where $n = 64$

$$P(\bar{x} > 114.5) = P(\bar{x} > 114.5) \text{ where } n = 64$$

$$P(\bar{x} > 114.5) = P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{114.5 - \mu}{\sigma/\sqrt{n}}\right) \text{ [By CLT]}$$

$$= P\left[Z > \frac{114.5 - 112}{12/\sqrt{64}}\right]$$

$$\begin{aligned}
 &= P[Z > 1.67] \\
 &= 1 - P[Z \leq 1.67] \\
 &= 1 - 0.95 \\
 &= 0.05
 \end{aligned}$$

For 90% confidence

$$\mu = \bar{x} \pm 1.645 \left(\frac{\sigma}{\sqrt{n}} \right)$$

For 95% C.I

$$\mu = \bar{x} \pm 1.96 \left(\frac{\sigma}{\sqrt{n}} \right)$$

For 99% C.I

$$\mu = \bar{x} \pm 2.58 \left(\frac{\sigma}{\sqrt{n}} \right)$$

Confidence limits or Fiducial limits for μ

I] For Small Samples

If $t_v(0.05)$ is the tabulated value of t for df $v = n - 1$ at 5% level of significance then 95% confidence limits for μ is given by $\left(\bar{x} \pm t_v(0.05) \frac{s}{\sqrt{n}} \right)$

$$\Rightarrow \mu \in \left(\bar{x} - t_v(0.05) \frac{s}{\sqrt{n}}, \bar{x} + t_v(0.05) \frac{s}{\sqrt{n}} \right)$$

Similarly 99% confidence limits for μ is $\left(\bar{x} \pm t_v(0.01) \frac{s}{\sqrt{n}} \right)$

$$\Rightarrow \mu \in \left(\bar{x} - t_v(0.01) \frac{s}{\sqrt{n}}, \bar{x} + t_v(0.01) \frac{s}{\sqrt{n}} \right)$$

$$1 - \alpha = 99$$

$$\alpha = 1\%$$

$$\alpha = 0.01$$

$$1 - \alpha = 95\%$$

$$1 - \alpha = 0.95$$

$$\alpha = 0.05$$

II] For Large Samples:

At $\alpha = 5\%$ level of significance, 95% confidence limits for μ is given by $\left(\bar{x} \pm Z_\alpha(0.05) \frac{\sigma}{\sqrt{n}} \right)$

$$\Rightarrow \mu \in \left(\bar{x} - Z_\alpha(0.05) \left(\frac{\sigma}{\sqrt{n}} \right); \bar{x} + Z_\alpha(0.05) \left(\frac{\sigma}{\sqrt{n}} \right) \right)$$

Similarly 99% confidence limits for μ is

$$\left(\bar{x} \pm Z_\alpha(0.01) \frac{\sigma}{\sqrt{n}} \right)$$

$$\mu \in \left(\bar{x} - z_{\alpha/2} (0.01) \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} (0.01) \frac{\sigma}{\sqrt{n}} \right)$$

Note:

* When the sample size is small say $n < 30$, the sampling distribution in many cases may not be normal. Here the test statistics will change.

Test of Significance for small samples:
Some important test for small samples are

- 1) t -test or t -distribution test or student's t -test
- 2) F -test or F -distribution test
- 3) χ^2 -test or χ^2 -distribution test

Degree of Freedom [denoted by ν]:

It is the number of independent variates which makeup the statistics.

In other words it is a number which indicates how many of the values of the variable maybe independently or freely chosen.

test for $\rightarrow$

I] t-test for a Specific Mean:

Let $x_1, x_2, x_3, \dots, x_n$ be the random sample of size n drawn from the normal population with mean μ and variance σ^2 then the test statistic is given by

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \Rightarrow t_{df} \text{-distribution where } \nu \text{ is the degree of freedom and } \underline{\underline{\nu = n-1}}$$

Also $S^2 = \frac{1}{(n-1)} \sum (x_i - \bar{x})^2$, S^2 is an unbiased estimate of the population variance σ^2 .

Note:

* We have the sample variance s^2 as $s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$

$$\Rightarrow ns^2 = \sum (x_i - \bar{x})^2 \rightarrow (1)$$

$$\text{Also } S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \Rightarrow (n-1)S^2 = \sum (x_i - \bar{x})^2 \rightarrow (2)$$

(1) * (2)

$$ns^2 = (n-1)S^2$$

$$\frac{s^2}{n-1} = \frac{S^2}{n} \Rightarrow \frac{s}{\sqrt{n-1}} = \frac{S}{\sqrt{n}}$$

$\therefore (*) \Rightarrow t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$ if the sample variance s^2 is known.

① A machinist is making an engine parts with axle with 0.7 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a standard deviation of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specifications.

Sol: Given:

$$\mu = 0.7$$

$$n = 10$$

$$\bar{x} = 0.742$$

$$s = 0.040$$

To test whether the work is meeting the specification [that is $\mu = 0.7$]

Step 1: $H_0: \mu = 0.7$

Step 2: $H_1: \mu \neq 0.7$ [t+t]

Step 3: For $\alpha = 5\%$ & d.o.f $\Rightarrow r = n - 1 = 10 - 1 = 9$

we have $t_{\alpha/2} = t_{9(0.05)} = 2.262$ Handbook
[from the
table]

Step 4: Compute the test statistic [S is bigger
& is smaller]

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$$t = \frac{0.742 - 0.7}{0.040/\sqrt{9}}$$

$$t = 3.15$$

$$|t| = 3.15$$

Step 5: $|t| > t_r(\alpha)$

$\Rightarrow$ we reject the H_0

$\Rightarrow \mu \neq 0.7$

$\therefore$ the work doesn't meet the specification

② The mean weekly sales of soap bars in departmental store was 146.3 bars per store. After an advertising campaign the mean weekly sales 22 stores for a typical week increased to 153.7 and showed SD of 17.2 was the advertising campaign successful?

Sol:

$$\mu = 146.3, n = 22, \bar{x} = 153.7, s = 17.2$$

To test whether the advertising campaign is successful or not that is to test whether $\mu > 146.3$ or not

Step 1: $\mu = 146.3$

Step 2: $\mu > 146.3$ [one tailed test]

Step 3: Let $\alpha = 5\%$ for d.f $\Rightarrow v = n - 1 = 21$

we have

$$t_r(\alpha) = t_{21}(0.05) = 1.721$$

from
table
t-test tab

Step 4: Compute test statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{153.7 - 146.3}{17.2/\sqrt{21}} = 1.97$$

$$|t| = 1.97$$

$\Rightarrow$ we reject H_0

$\therefore$ the advertising campaign is successful

③ A random sample of 10 boys had the following IQ's:

70, 120, 101, 88, 83, 95, 98, 107, 100, 110

Do these data support the assumption of a population mean IQ of 100?

Find the reasonable range in which most of the mean IQ values of sample of 10 boys lie.

Sol:

To test whether the data support the mean IQ of 100
[$\mu = 100$]

Step 1: $H_0: \mu = 100$

Step 2: $H_1: \mu \neq 100$ [TTT]

Step 3: Let $\alpha = 5\%$ * for d.f $\Rightarrow v = n - 1 \Rightarrow 10 - 1 = 9$
we have $t_v(\alpha) = [t_{(9)} = (0.05) = 2.262]$ from table

Step 4: Compute test statistic

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \rightarrow (*)$$

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

$$\bar{x} = \frac{70 + 120 + 101 + 88 + 83 + 95 + 98 + 107 + 100 + 110}{10}$$

$$\bar{x} = 97.2$$

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

$$= \frac{1}{10-1} [70-97.2)^2 + (120-97.2)^2 + (101-97.2)^2 + \dots]$$

$$S^2 = \frac{1}{9} [739.84 + 519.84 + 14.44 + 201.64 + 84.64 + 4.84 + 0.64 + 96.04 + 7.84 + 163.84]$$

$$S^2 = 203.73$$

$$S = 14.27$$

$$t = \frac{97.2 - 100}{\sqrt{\frac{203.73}{10}}}$$

$$t = -0.62$$

$$|t| = 0.62$$

$$|t| < t_{\alpha/2}$$

We accept H_0 $\mu = 100$

95% confidence limits for the mean is given by

$$\bar{x} \pm t_{\alpha/2} \left(\frac{S}{\sqrt{n}} \right)$$

$$\begin{aligned} \text{or confidence interval} &\Rightarrow \left(\bar{x} - t_{\alpha/2} \left(\frac{S}{\sqrt{n}} \right), \bar{x} + t_{\alpha/2} \left(\frac{S}{\sqrt{n}} \right) \right) \\ &= \left(97.2 - 2.26 \left(\sqrt{\frac{S^2}{n}} \right), 97.2 + 2.26 \sqrt{\frac{S^2}{n}} \right) \\ &= \left(97.2 - 2.26 \left(\sqrt{\frac{203.73}{10}} \right), 97.2 + 2.26 \sqrt{\frac{203.73}{10}} \right) \\ &= (86.99, 107.4) \end{aligned}$$

4) The Heights of ^{10 males} given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66. Is it reasonable to believe that the average height is greater than 64 inches. Test at 5% level of significance assuming that the degree of freedom is 9

Sol: To test whether the average height is greater than 64
Suppose that average height is 64

Step 1: $H_0: \mu = 64$

Step 2: $H_1: \mu > 64$ [One tailed test]

Step 3: $\alpha = 5\%$ * for d.f $\Rightarrow r = n - 1 = 10 - 1 = 9$
 $t_{\alpha}(r) = t_{0.05}(9) = 1.833$

Step 4: Compute test Statistics

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \longrightarrow (*)$$

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

$$\bar{x} = \frac{70 + 67 + 62 + 68 + 61 + 68 + 70 + 64 + 64 + 66}{10}$$

$$\bar{x} = 66$$

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

$$s^2 = \frac{1}{9} \left[(70-66)^2 + (67-66)^2 + (62-66)^2 + (68-66)^2 + (61-66)^2 + (68-66)^2 + (70-66)^2 + (64-66)^2 + (64-66)^2 + (66-66)^2 \right]$$

$$S^2 = \frac{90}{9}$$

$$S^2 = 10$$

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$t = \frac{66 - 64}{\sqrt{\frac{10}{10}}}$$

$$t = 2$$

$$|t| = 2$$

Step 5: $|t| > t_{\alpha/2}$
 $\Rightarrow$ we reject H_0
 $\therefore \mu > 64$

$\therefore$ The average height is greater than 64 inches

⑤ A random sample of 16 values from a normal population showed a mean of 41.5 inches and the sum of squares of deviations from this mean equal to 135.59 inches. Show that the assumption of the mean of 43.5 inches. For the population is not reasonable. Obtain the 95% and 99% fiducial limits for the same.

Sol: $n = 16$ $\bar{x} = 41.5$ } Given
 $\sum (x - \bar{x})^2 = 135$

To test whether $\mu = 43.5$

Step 1: $H_0 : \mu = 43.5$

Step 2: $H_1 : \mu \neq 43.5$ [TTT]

Step 3: $\alpha = 5\%$ for d.f $\Rightarrow v = n - 1 = 16 - 1 = 15$
 $t_{v}(\alpha) = t_{15}(0.05) = 2.131$ from the table

Step 4: To compute test statistic.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \rightarrow (*)$$

Where $\bar{x} = 41.5$ and $\sum (x_i - \bar{x})^2 = 135$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1 = 16 - 1}$$

$$s^2 = \frac{135}{15}$$

$$s^2 = 9$$

$$s = 3$$

$$(*) \Rightarrow t = \frac{41.5 - 43.5}{3/\sqrt{16}}$$

$$t = -2.667$$

$$|t| = 2.667$$

Step 5: $|t| > t_{v}(\alpha) \Rightarrow$ we reject H_0
 $\mu \neq 43.5$

95% fiducial limits for mean

$$41.5 \pm 2.131 \left(\frac{3}{\sqrt{16}} \right)$$

$$(43.09, 39.90)$$

$$(39.90, 43.09)$$

lower bound upper bound

99% fiducial limits for mean

$$\bar{x} \pm t_{15}(0.01) \left(\frac{s}{\sqrt{n}} \right)$$

$$41.5 \pm 2.947 \left(\frac{3}{\sqrt{16}} \right)$$

$$43.71, 39.28$$

$$(39.28, 43.71)$$

II] t-test for difference of means:

Let x_i where $(i = 1, 2, 3, \dots, n_1)$ and y_j where $(j = 1, 2, \dots, n_2)$ be 2 independent sample of sizes n_1 and n_2 respectively drawn from 2 populations with mean μ_x and μ_y respectively under the assumption that the population

variances are same [i.e., $\sigma_x^2 = \sigma_y^2 = \sigma^2$]

Then the test statistic is given by $t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ or

$$\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Where

$$\bar{x} = \frac{\sum x_i}{n_1}$$

$$\bar{y} = \frac{\sum y_j}{n_2}$$

$$S^2 = \frac{\sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2}{n_1 + n_2 - 2}$$

Here d.f $\Rightarrow v = (n_1 + n_2 - 2)$

if sample variances s_1^2 & s_2^2 are known then,

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

to test whether there is significant difference between mean we have the test statistic as

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{under } H_0$$

① While measuring specimens of nylon yarn taken from two specimen machine, it was found that 8 specimens the first machine had a mean denier of 9.67 with SD of 1.81, while 10 specimens from the second machine had a mean denier of 7.43 with a SD of 1.48.

Assuming that the population are normal, test the null hypothesis $\mu_x - \mu_y = 1.5$ against the alternative hypothesis $\mu_x - \mu_y > 1.5$ at 0.05 level of significance.

$$n_1 = 8$$

$$\bar{x} = 9.67$$

$$s_1 = 1.81$$

Machine 2

$$n_2 = 10$$

$$\bar{y} = 7.43$$

$$s_2 = 1.48$$

Step 1: $H_0: \mu_x - \mu_y = 1.5$

Step 2: $H_1: \mu_x - \mu_y > 1.5$

Step 3: For $\alpha = 0.05$ and d.f. $\Rightarrow v = (n_1 + n_2 - 2)$

$$t_{v(\alpha)} = t_{16}^{(0.05)} = 1.746$$

$$\frac{8+10-2}{16}$$

Step 4: Compute test statistics

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow *$$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$s^2 = \frac{8(1.81)^2 + 10(1.48)^2}{16}$$

$$s^2 = 3$$

$$* \Rightarrow t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(9.67 - 7.43) - 1.5}{\sqrt{3 \left(\frac{1}{8} + \frac{1}{10} \right)}}$$

$$= 0.901$$

$$\therefore |t| = 0.901$$

Step 5: $|t| < t_{v(\alpha)} \Rightarrow$ we accept $H_0 \Rightarrow \mu_x - \mu_y = 1.5$

② A gain in weights (in kgs) of pigs fed on diet A and diet B

Diet A: 25, 32, 30, 34, 24, 140, 32, 24, 30, 31, 35, 25

Diet B: 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 180, 21, 25, 31, 29, 22

Test if two diets differ significantly

Sol: Suppose there is no significance difference between two diets ($\mu_x - \mu_y$)

Step 1: $H_0: \mu_x - \mu_y$

Step 2: $H_1: \mu_x \neq \mu_y$ [TTT]

Step 3: Let $\alpha = 5\% = 0.05$ * d.f $\nu = (n_1 + n_2 - 2)$
 $\nu = 12 + 15 - 2$

$$t_{\alpha} = t_{25}(0.05) = 2.06 \quad \nu = 25$$

Step 4: Compute test statistic

$$t = \frac{(\bar{x} - \bar{y})}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow (*)$$

$$s^2 = n_1 s_1^2$$

where,

$$\bar{x} = \frac{\sum x_i}{n_1}$$

$$\bar{y} = \frac{\sum y_j}{n_2}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2}{n_1 + n_2 - 2}$$

$$\frac{25+32+30+34+24+14+32+24+30+31+35+20}{12}$$

$$= 28$$

$$\bar{y} = \frac{44+34+22+10+47+31+40+30+32+35+18+21+35+29+22}{15}$$

$$\bar{y} = 30$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2}$$

$$\sum (x_i - \bar{x})^2 = (x_i - 28)^2 = 9+16+4+36+16+16+4+9+49+9$$

$$= 380$$

$$\sum (y_i - \bar{y})^2 = (y_i - 30)^2 = 196+16+64+400+289+1+100+0+4$$

$$25+144+81+25+1+64$$

$$= 1410$$

$$S^2 = \frac{380+1410}{25}$$

$$S^2 = 71.6$$

$$t = \frac{28-30}{\sqrt{71.6 \left(\frac{1}{12} + \frac{1}{15} \right)}} = -0.61 \quad |t| = 0.61$$

Step 5: $|t| < t_{\alpha/2}$ $\Rightarrow$ we accept H_0
 $\mu_x = \mu_y$

there is no significant difference between two diet

③ Samples of two types of electric bulbs were tested for length of life for the following data were observed

Type I	Type I	Type II
Sample no	$n_1 = 8$	$n_2 = 7$
Sample means	$\bar{x}_1 = 1234 \text{ hrs}$	$\bar{x}_2 = 1036 \text{ hrs}$
Sample SD's	$s_1 = 36 \text{ hrs}$	$s_2 = 40 \text{ hrs}$

Is the difference in the means sufficient to warrant that Type I is superior to Type II regarding length of life

Sol:

To test whether Type I is superior to Type II

Suppose that there is no difference between 2 types

$$[i.e. = \mu_x = \mu_y]$$

Step 1: $H_0: \mu_x = \mu_y$

Step 2: $H_1: \mu_x > \mu_y$ [one tailed test]

Step 3: Let $\alpha = 5\%$ & for d.f $\rightarrow v = n_1 + n_2 - 2$

$$= 8 + 7 - 2$$

$$= 13$$

$$t_{v(\alpha)} = t_{13}(0.05) = 1.771$$

Step 4: Compute test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow *$$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{8(36)^2 + 7(40)^2}{13} = 1659.08$$

$$\sqrt{1659.08 \left(\frac{1}{8} + \frac{1}{7} \right)}$$

$$t = 9.39$$

$$|t| = 9.39$$

Step 5: $|t| > t_{\alpha} \Rightarrow$ We reject H_0

$$\Rightarrow \mu_x > \mu_y$$

$\therefore$ type I is superior to type II

Q The heights of six randomly chosen sailors are 63, 65, 68, 69, 71, 72 (inches). Those of 10 randomly chosen soldiers are 61, 62, 65, 66, 69, 69, 70, 71, 72, 73. Discuss the light that these data thrown on the suggestion that sailors are on the average taller than the soldiers.

(one tail, $t = 0.792$)

Sol: To test whether the average height of sailor is greater than the average height of soldiers $[\mu_x > \mu_y]$
Suppose that the average height of sailors & soldiers are seen i.e. $[\mu_x = \mu_y]$

Step 1: $H_0: \mu_x = \mu_y$

Step 2: $H_1: \mu_x > \mu_y$ [Single Tail Test]

Step 3: For $\alpha = 5\%$ d.f. $\Rightarrow v = n_1 + n_2 - 2 = 6 + 10 - 2 = 14$

$$t_{\alpha} = t_{14}(0.05) = 1.761$$

Step 4: Compute test statistic.

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow \text{②}$$

Where;

$$\bar{x} = \frac{\sum x_i}{n_1}$$

$$\bar{y} = \frac{\sum y_j}{n_2}$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2}{n_1 + n_2 - 2}$$

$$\bar{x} = \frac{408}{6} = 68$$

$$\bar{y} = \frac{678}{10} = 67.8$$

$$\sum (x_i - 68)^2 = 25 + 9 + 0 + 1 + 9 + 16 = 60$$

$$\begin{aligned} \sum (y_j - 67.8)^2 &= 46 \cdot 24 + 33 \cdot 64 + 784 + 3 \cdot 24 + 1 \cdot 44 + 1 \cdot 44 + \\ &\quad 4 \cdot 84 + 10 \cdot 24 + 17 \cdot 64 + 27 \cdot 04 \\ &= 153.6 \end{aligned}$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2}{n_1 + n_2 - 2}$$

$$S^2 = \frac{60 + 153.6}{14}$$

$$S^2 = 15.257$$

$$(*) t = \frac{68 - 67.8}{\sqrt{15.257 \left(\frac{1}{6} + \frac{1}{10} \right)}} = 0.099 \quad |t| = 0.099$$

Step 5: $|t| < t_{\alpha/2} \Rightarrow$ we accept H_0

i.e. $\mu_x = \mu_y$

$\therefore$ The sailors are not taller than the

5) To list the claim that the resistance of electric wire can be reduced by at least 0.05 ohm by reduced by at least 0.05 ohm by alloying, 25 values obtained for each alloyed wire and std wire produced the following results

	mean	Standard deviation
Alloyed wire	0.083 ohm	0.003 ohm
Standard wire	0.136 ohm	0.002 ohm

Test at 5% level whether or not the claim is substantiated.

Sol: $H_0: \mu_x - \mu_y \geq 0.05$

$H_1: \mu_x - \mu_y < 0.05$ [LTT]

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow \textcircled{*}$$

where $S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{25(0.003)^2 + 25(0.002)^2}{25 + 25 - 2}$

$$S^2 = 0.0000067$$

$\textcircled{*} \Rightarrow$

$$t = \frac{(0.083 - 0.136) - (0.05)}{\sqrt{0.0000067 \left(\frac{1}{25} + \frac{1}{25} \right)}} = \frac{-0.103}{0.00071}$$

$$t = -145.07$$

$$|t| = 145.07 \rightarrow \textcircled{1}$$

$$\nu = n_1 + n_2 - 2 = 48$$

For $\alpha = 5\%$ & for LTT, $t_r(\alpha) = -1.645 \rightarrow \textcircled{2}$

$\textcircled{1} \times \textcircled{2} \Rightarrow |t| > t_r(\alpha) \Rightarrow$ we reject H_0

- ⑥ A group of boys and girls were given an intelligence test. The mean score, SDs and number in each group are as follows

	Boys	Girls
Mean	124	121
SD	12	10
n	18	14

Is the mean score of boys significantly different from that of girls?

Sol:

$$H_0: \mu_x = \mu_y$$

$$H_1: \mu_x \neq \mu_y \text{ [TTT]}$$

$$v = n_1 + n_2 - 2 = 18 + 14 - 2 = 30$$

$$\alpha = 5\%$$

$$t_{\alpha/2} = 2.042 \rightarrow \textcircled{1}$$

TTT

$$\alpha/2 = 0.025$$

$$t = \frac{\bar{x} - \bar{y}}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{124 - 121}{\sqrt{133.1 \left(\frac{1}{18} + \frac{1}{14} \right)}}$$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{18(12)^2 + 14(100)}{30}$$

$$= 133.1$$

$$\textcircled{*} \quad t = \frac{3}{11.54 \times 0.36} = 0.722$$

$$|t| = 0.722 \rightarrow \textcircled{2}$$

① * ② $\Rightarrow |t| < t_{\alpha/2} \Rightarrow$ we accept H_0

$\therefore$ mean score of boys does not differ significantly from

7) Two horses A and B were tested according to the time (in sec) to run a particular race with the following results

Horse A: 28 30 32 33 33 29 24

Horse B: 29 30 30 24 27 29

1: Step 1: $H_0: \mu_x = \mu_y$

Step 2: $H_1: \mu_x \neq \mu_y$

Step 3: For $\alpha = 5\%$ and for TTT $y = n_1 + n_2 - 2 \Rightarrow 7 + 6 - 2 = 11$

$$T_{13}(0.05) = 2.201 \rightarrow \textcircled{1}$$

Step 4: Compute the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

For Given, $\bar{x} = 31.29$, $\bar{y} = 28.17$

$$S^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2}{n_1 + n_2 - 2} \rightarrow \textcircled{*}$$

$$\begin{aligned} \sum (x_i - \bar{x})^2 &= 10.82 + 1.664 + 0.5041 + 2.9241 + 2.9241 + 5.2441 \\ &\quad + 7.3441 \\ &= 31.4256 \end{aligned}$$

$$\begin{aligned} \sum (y_j - \bar{y})^2 &= 0.6889 + 3.3489 + 3.3489 + 17.3889 + 1.3689 \\ &\quad + 0.6889 \\ &= 26.83 \end{aligned}$$

From $\textcircled{*}$

$$S^2 = \frac{31.425 + 26.833}{11} = \frac{4.5916}{11} = 5.296$$

$$S^2 = 5.296$$

$$t = \frac{31.29 - 28.17}{\sqrt{5.296 \left(\frac{1}{7} + \frac{1}{6} \right)}} \quad t = 2.437$$

$$|t| = 2.437 \rightarrow ③$$

Step 5: $|t| > t_{\alpha/2}$, we reject H_0
we accept H_1

⑧ A group of 10 boys fed on a diet A and another group of 8 boys fed on a different diet B for a period of 6 months recorded the following increase in weight

Diet A: 5 6 8 1 12 4 3 9 6 10

Diet B: 2 3 6 8 10 1 2 8

Test whether diets A and B differ significantly w.r.t their effect on increase in weight

Sol: Given,

$$n_1 = 10$$

$$n_2 = 8$$

$$\bar{x} = 6.4$$

$$\bar{y} = 5$$

Step 1: $H_0: \mu_x = \mu_y$

Step 2: $H_1: \mu_x \neq \mu_y$ [TTT]

Step 3: For $\alpha = 5\%$, for TTT $v = n_1 + n_2 - 2 = 16$

Step 4: Compute the test statistic

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \rightarrow (*)$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2}{n_1 + n_2 - 2}$$

$$\begin{aligned} \sum (x_i - 6.4)^2 &= 1.96 + 0.16 + 2.56 + 29.16 + 31.36 + 5.76 + 11.56 + 6.76 + 0.16 + \\ &12.96 \\ &= 102.4 \end{aligned}$$

$$\begin{aligned} \sum (y_j - 5)^2 &= 9 + 4 + 1 + 9 + 25 + 16 + 9 + 9 \\ &= 82 \end{aligned}$$

$$S^2 = \frac{102.4 + 82}{16} = 11.525$$

$$\text{From } (*) \Rightarrow t = \frac{6.4 - 5}{\sqrt{11.525 \left(\frac{1}{10} + \frac{1}{8} \right)}} = 0.87$$

$$|t| = 0.87$$

Step 5: $|t| < t_{\alpha/2}$, we accept H_0

i.e., $\mu_x = \mu_y$

diet A and B does not differ significantly

III] t-test for difference of means

Here, we consider the case when

- (i) the sample sizes are equal i.e., $n_1 = n_2 = n$ (say)
- (ii) the two samples are not independent but the sample observations are paired together i.e., the pair of observation (x_i, y_i) :
($i = 1, 2, \dots, n$)

* Corresponds to same i th sample means unit
 * The problem is to test if the sample means differ significantly or not

Define $[d_i = x_i - y_i]$; ($i = 1, 2, \dots, n$) [the increments]
 then, the test statistic is given by

$$t = \frac{\bar{d}}{S/\sqrt{n}} \rightarrow *$$

where $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$ * $S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (d_i - \bar{d})^2$

Here, the df is $[v = n-1]$

For eg Suppose we want to test the efficiency of a particular drug for inducing sleep.

Let $x_1, x_2, \dots, x_n$ and $y_1, y_2, \dots, y_n$ be the readings, in hours of sleeps, on the i th individual, before and after the drug is given.
 Here, instead of applying the difference of means test, we apply the paired ~~to~~ test or t-test for paired data

1) A certain stimulus administered to each of the 12 patients resulted in the following increases of blood pressures 5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4, 6. Can it be concluded that the stimulus will, in general be accompanied by increase in blood pressure

Sol: Here the increments in blood pressure are given i.e.

$d_i = (x_i - y_i)$ is given

To test whether the stimulus will be accompanied by an increase in bp,

i.e. $[u_y > u_x \text{ or } u_x > u_y]$

Suppose that there is no significant difference before and after administering the drug [i.e. $u_x = u_y$]

Step 1: $H_0: u_x = u_y$

Step 2: $H_1: u_x < u_y$ (Single tailed test)

Step 3: For $\alpha = 5\%$ & d.f $\Rightarrow v(n-1) = 12-1 = 11$

$$t_{v(\alpha)} = \boxed{t_{11}(0.05) = 1.796} \quad [\text{from table}]$$

Step 4: $t = \frac{\bar{d}}{\sqrt{s^2/n}} \rightarrow \otimes$

where $\bar{d} = \frac{\sum d_i}{n}$

$$s^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}$$

$$d_i = 5 + 2 + 8 - 1 + 3 + 0 - 2 + 1 + 5 + 0 + 4 + 6 = 31$$

$$\sum d_i = 31 \quad - 9.58$$

$$\begin{aligned}
 (d_i - 2.58)^2 &= 5.856 + 0.336 + 29.38 + 6.66 + 6.66 + 2.016 \\
 &\quad + 20.98 + 2.496 + 5.86 + 6.66 + 6.66 + 2.016 \\
 &\quad + 11.696 \\
 &= 104.94
 \end{aligned}$$

$$s^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = \frac{104.94}{11} = 9.54$$

$$* \Rightarrow t = \frac{2.58}{\sqrt{9.54/10}} = 2.9$$

$$|t| = 2.9$$

Step 5: $|t| > t_{\alpha/2} \Rightarrow$ we reject H_0

$$\mu_x < \mu_y$$

$\therefore$ The stimulus is accompanied by an increase in BP

② 10 soldiers participated in a shooting competition in the first week. After intensive training they participated in the competition in the second week. Their scores before & after training are.

Score before: 67, 24, 57, 55, 63, 54, 56, 68, 33, 43

Score after: 70, 38, 58, 58, 56, 67, 68, 75, 42, 38

Do the data indicate that the soldiers have been benefited by the training.

Sol: To test whether the soldiers have been benefited by training. It is enough to test average score after the training is greater than the average score before the

Suppose that there is no significance difference in the average score before and after the training.

$$[i.e.; \mu_x = \mu_y]$$

Step 1: $H_0: \mu_x = \mu_y$

Step 2: $H_1: \mu_x < \mu_y$ [Single tail test]

Step 3: For $\alpha = 5\%$ & d.f $\Rightarrow v = n - 1 = 10 - 1 = 9$

$$\therefore t_v(\alpha) = [t_{9, 0.05}] = 1.833$$

Step 4: Compute test statistic.

$$t = \frac{\bar{d}}{\sqrt{\frac{s^2}{n}}} \rightarrow *$$

Where, $\bar{d} = \frac{\sum d_i}{n}$

$$s^2 = \frac{\sum (d_i - \bar{d})^2}{n - 1}$$

$$x_i = 67 + 24 + 57 + 55 + 63 + 54 + 56 + 68 + 33 + 43 = 520$$

$$y_j = 70 + 38 + 58 + 58 + 56 + 67 + 68 + 75 + 42 + 38 = 570$$

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-50}{10} = -5 \quad d_i = (x_i - y_j) = -3, -14, -1, -3, 7, -13, -12, 7, -9, 5.$$

$$d_i = x_i - y_j = 520 - 570 = -50$$

$$\sum (d_i - (-5))^2 = 16 + 81 + 16 + 4 + 144 + 64 + 4 + 49 + 4 + 16 + 100 = 482$$

$$s^2 = \frac{\sum (d_i - \bar{d})^2}{n - 1} = \frac{482}{9} = 53.56$$

$$(*) \Rightarrow t = \frac{-5}{\sqrt{53-56/10}}$$

$$t = -2.16$$

$$|t| = 2.16$$

Step 5: $|t| > t_{\alpha}$ we reject H_0

$\mu_x \neq \mu_y$ i.e. The soldiers have been benefited by the training

Imp*

③ The following data relate to the marks obtained by 11 students in 2 tests. Second test is after intense coaching do the data indicate that the student has been benefited by the coaching

Test 1: 19, 23, 16, 24, 17, 18, 20, 18, 21, 19, 20

Test 2: 17, 24, 20, 24, 20, 22, 20, 20, 18, 22, 19

Sol: To test whether the students are benefited by the training, and it is enough to test the average marks after the training is greater than the average marks before the training (i.e. $\mu_y > \mu_x$ or $\mu_x < \mu_y$)

Suppose there is no significant difference in the average marks before and after training (i.e. $\mu_x = \mu_y$)

Step 1: $H_0: \mu_x = \mu_y$

Step 2: $H_1: \mu_x < \mu_y$ (single tailed test)

Step 3: For $\alpha = 5\%$ & d.f $\Rightarrow v = n - 1 = 11 - 1 = 10$

$$t_r(\alpha) = t_{10}(0.05) = 1.812$$

Step 4: Compute the test statistic

$$t = \frac{\bar{d}}{\sqrt{s^2/n}} \rightarrow (*) \quad [d_i = x_i - y_i]$$

Where,

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-11}{10} = -1.1$$

$$d_i = 2, -1, -4, 0, -3, -4, 0, -2, 3, -3, 1$$

$$s^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} \rightarrow (1)$$

$$\begin{aligned} \sum (d_i + 1.1)^2 &= 9.61 + 0.01 + 8.41 + 1.21 + 3.61 + 8.41 + 1.21 + 0.81 + \\ &\quad 16.81 + 3.61 + 4.41 \\ &= 58.11 \end{aligned}$$

$$s^2 = \frac{58.11}{11-1} = \frac{58.11}{10} = 5.811$$

$$\text{from } (*) \Rightarrow t = \frac{-1.1}{\sqrt{\frac{5.811}{11}}} = -1.513 = |t| = 1.513$$

Step 5: $|t| < t_r(\alpha)$, we accept H_0

i.e. $\mu_x = \mu_y$

$\therefore$ Students have not been benefited by training

Step 3: For $\alpha = 5\%$ & dof $\Rightarrow v = n_1 + n_2 - 2$

$$\therefore t_v(\alpha) = t_{14}(0.05) = 1.761$$

Step 4: $t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$

where,

$$\bar{x} = \frac{\sum x_i}{n_1} =$$

$$\bar{y} = \frac{\sum y_j}{n_2} =$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_j - \bar{y})^2}{n_1 + n_2 - 2}$$

$$\sum (x_i - 50.88)^2 = 30.902$$

$$\sum (y_j - 52.88)^2 = 16.898$$

$$S^2 = \frac{30.902 + 16.898}{14} = 3.414$$

$$\Rightarrow t = \frac{50.875 - 52.875}{\sqrt{3.414 \left(\frac{1}{8} + \frac{1}{8} \right)}} = -2.164$$

$$|t| = 2.164$$

Step 5: $|t| > t_v(\alpha)$

We reject H_0

(11)

Here, same set of animals were used
Hence, we apply 3rd t-test or t-test for

Step 1: $H_0: \mu_x = \mu_y$

Step 2: $H_1: \mu_x < \mu_y$

Step 3: for $\alpha = 5\%$ & for d.f $r = n - 1 = 8 - 1 = 7$

$\therefore t_r(\alpha) = t_7(0.05) = 1.895$

Step 4: Compute test statistic

$$t = \frac{\bar{d}}{\sqrt{s^2/n}} \rightarrow (*)$$

$$d_i = x_i - y_i = 16$$

$$(d_i - (-2))^2 = 1 + 0 + 1 + 1 + 1 + 4 + 0 + 4 = 12$$

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-16}{8} = -2$$

$$s^2 = \frac{\sum (d_i - \bar{d})^2}{n-1} = \frac{12}{7} = 1.714$$

$$(*) \Rightarrow t = \frac{-2}{\sqrt{\frac{1.714}{8}}} = 4.321$$

$$\boxed{|t| = 4.321}$$

Step 5:

$|t| > t_r(\alpha) \Rightarrow$ we reject H_0

$$\mu_x < \mu_y$$

Food B is better than Food A.

χ^2 -test (Chi-square distribution test) or
Test for goodness of fit.

A very powerful test for testing the significance of difference between theory and experiment was given by Prof. Karl Pearson and is known as χ^2 -test of goodness of fit. It enables us to find if the deviation of the experiment from theory is just by chance or is it really due to the inadequacy of the theory to fit the observed data.

If O_i ($i=1, 2, \dots, n$) is a set of observed or experimental frequencies and L_i ($i=1, 2, \dots, n$) is the corresponding set of expected or theoretical frequencies, then,

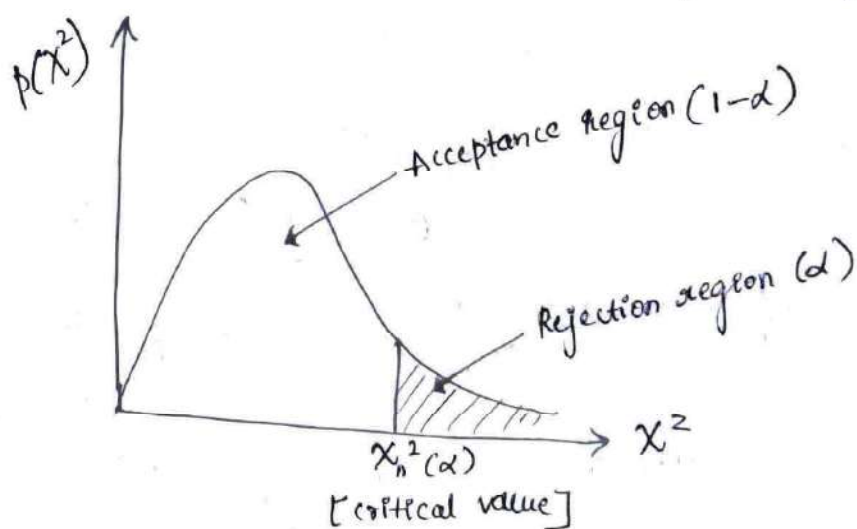
$$\chi^2 = \sum_{i=1}^n \left[\frac{(O_i - L_i)^2}{L_i} \right] \sim \chi^2_{(n-1)}$$

where d.f. = $n-1$ and $\sum_{i=1}^n O_i = \sum_{i=1}^n L_i = N$ (total frequency)

Note:

- * It is an approximate test for large values of n
- * Conditions for the validity of χ^2 -test of goodness of fit:
 - * (i) the sample observations should be independent
 - (ii) $\sum O_i = \sum L_i$
 - (iii) The total frequency N , should be reasonably large. i.e., $N > 50$.
 - (iv) No theoretical frequency should be less than or equal to 5.

- * χ^2 -test depends only on the observed and expected frequencies and on degree of freedom (d.f).
- * Since χ^2 definition does not involve any population parameters, it is termed as a statistic and the test is known as Non-parametric test or distribution-free test.
- * χ^2 is not symmetrical. It lies entirely in the 1st quadrant and hence not a normal (bell curve) curve. i.e., χ^2 varies from 0 to ∞ .
- * $P[\chi^2 > \chi_n^2(\alpha)] = \alpha \Rightarrow \chi_n^2(\alpha)$ is the critical value for n-degree of freedom.



- * From the χ^2 table, we see that the critical values of χ^2 increase as the degree of freedom increases and the level of significance decreases.
- * As per the condition of the application of χ^2 -test the theoretical frequencies should not be less than or equal to 5. In case any of the cell frequencies are small (less than 5) in two or more cells, they should be combined into a single cell (pooling method); so that the pooled frequencies is greater than 5.

Note:

Degree of freedom for χ^2 -test

(i) If the data is given in a series of n -numbers then,

$$d.f = n - 1$$

(ii) In case of Binomial distribution, $d.f \Rightarrow n - 1$

In case of Poisson distribution, $d.f \Rightarrow n - 2$

In case of normal distribution, $d.f \Rightarrow n - 3$

Q1) The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

Digits :-	0	1	2	3	4	5	6	7	8	9	Total
Frequency :-	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test whether the digits may be taken to occur equally frequently in the directory.

Solⁿ Step 1 H_0 : The digits occur equally frequently (uniformly in the directory).

Step 2: H_1 : The digits do not occur equally frequently (uniformly in the directory).

To find theoretical or expected frequency e_i .

The digits occur equally frequently implies $e_i = \frac{\sum O_i}{10}$

$$\frac{N}{10} = \frac{10,000}{10} \Rightarrow 1000$$

x_i	O_i	e_i	$(O_i - e_i) / e_i$
0	1026	1000	0.676
1	1107	1000	11.449
2	997	1000	0.009
3	966	1000	1.156
4	1075	1000	5.625
5	933	1000	4.489
6	1107	1000	11.449
7	972	1000	0.784
8	964	1000	1.216
9	853	1000	21.609

$$\chi^2_{(cal)} = \sum_{i=0}^9 \left[\frac{(O_i - e_i)^2}{e_i} \right] = 58.542 //$$

$$d.f = v = (n-1) = 10-1 = 9$$

$$\therefore \boxed{\chi^2_v(\alpha) = \chi^2_{(9)}(0.05) = 16.919} \quad (\text{from the table})$$

$$\therefore \chi^2_{(cal)} > \chi^2(\alpha) \Rightarrow \text{we reject } H_0.$$

The digits do not occur equally frequently (uniformly in digits)

Q2) A pair of dice are thrown 360 times and frequency of each sum is indicated below.

Sum:	2	3	4	5	6	7	8	9	10	11	12
frequency:	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of χ^2 -test at 0.05 level of significance [given $\chi^2_{(10)}(0.05) = 18.307$]

Soln:-

Step 1: H_0 : Dice are fair (unbiased)

Step 2: H_1 : Dice are not fair.

to find theoretical frequencies (e_i) :-

x_i	$p(x_i)$	$e_i = \frac{p(x_i)}{1/36} \times 360 = 10 \times x_i$	O_i	e_i	$\frac{(O_i - e_i)^2}{e_i}$	
1	1/36	10	2	8	10	0.4
2	2/36	20	3	24	20	0.8
3	3/36	30	4	35	30	0.833
4	4/36	40	5	37	40	0.225
5	5/36	50	6	44	50	0.72
6	6/36	60	7	65	60	0.416
7	6/36	50	8	51	50	0.02
8	5/36	40	9	42	40	0.1
9	4/36	30	10	26	30	0.533
10	3/36	20	11	14	20	1.8
11	2/36	10	12	14	10	1.6
12	1/36					<u>7.444</u>

$$\chi^2_{(cal)} = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 7.444 //$$

$$d.f = v = (n-1) = 11-1 = 10$$

$$\therefore \chi^2_v(\alpha) = \chi^2_{(10)}(0.05) = 18.307$$

$$\therefore \chi^2_{(cal)} < \chi^2_v(\alpha) \rightarrow \text{we accept } H_0$$

Dies are fair

Q3) ^{*imp} The number of automobile accidents per week in a certain community were as follows :-

12, 8, 20, 2, 14, 10, 15, 6, 9, 4

Are these frequencies in agreement with the belief that the accident conditions were same during these 10 week period.

Soln: H_0 : The accident conditions are same during the 10 week period

H_1 : Accident conditions are not same during the 10 week period.

* To find theoretical frequency (e_i)

As the accident conditions are same during the 10 week period we have, $e_i = \frac{N}{10} = \frac{\sum O_i}{10} = \frac{100}{10} = 10$

x_i	O_i	e_i	$(O_i - e_i)^2 / e_i$
1	12	10	0.4
2	8	10	0.4
3	20	10	10
4	22	10	0.8
5	14	10	0
6	10	10	2.5
7	15	10	1.6
8	6	10	2.45
9	9	10	
10	4	10	
			<u>18.15</u>

* Here, we have two cell values as 2 & 4, which is less than 5 hence we apply pooling method, i.e., we combine it with either preceding or succeeding cell value to obtain a new frequency which is greater than 5. The same procedure is done for e_i also]

$$\chi^2_{(cal)} = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 18.15$$

$$\boxed{\chi^2_{(cal)} = 18.15}$$

$$d.f \Rightarrow (n-1) - 2 = 7$$

$$\boxed{\chi^2_7(0.05) = 14.07}$$

$\therefore \chi^2_{(cal)} > \chi^2_{(2)}$, we reject H_0

the Accident conditions are not same during 10 week period

Q4) 4 coins are tossed 160 times and the following results were obtained

No. of heads : 0 1 2 3 4

observed frequency : 17 52 54 31 6

under the assumption that coins are balanced, find the expected frequencies of 0, 1, 2, 3, 4 and test the goodness of fit at 5% level of significance.

H_0 : Coins are balanced [unbiased]

H_1 : Coins are not balanced.

To find then,

Here, the experiment of tossing 4 coins follows a binomial distribution with random variable X has obtaining the no. of heads. i.e., $X \sim B(n, p)$ where

$p \Rightarrow$ probability of success $\Rightarrow$ prob. of getting a head $= 1/2$

$q \Rightarrow$ probability of failure $\Rightarrow$ prob. of getting a tail $= 1/2$

$n \Rightarrow$ no. of trials

$$p.m.f \Rightarrow p(x) = {}^n C_x p^x q^{n-x} ; x \rightarrow 0, 1, 2, \dots, n$$

by $\textcircled{*}$, $n=4$

$$p(x_i) : \begin{matrix} 1/16 & 4/16 & 6/16 & 4/16 & 1/16 \end{matrix} \quad \begin{matrix} p(0) = {}^4 C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 \\ = 1/16 \end{matrix}$$

$$e_i : \frac{1}{16} \times 160 \quad \frac{4}{16} \times 160 \quad \frac{6}{16} \times 160 \quad \frac{4}{16} \times 160 \quad \frac{1}{16} \times 160 \quad \begin{matrix} p(1) = 4/16 \end{matrix}$$

$$e_i : 10 \quad 40 \quad 60 \quad 40 \quad 10 \quad \begin{matrix} p(2) = 3/8 \text{ or } 6/16 \\ p(3) = 1/4 \text{ or } 4/16 \end{matrix}$$

$$x_i \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad \begin{matrix} p(4) = 1/16 \end{matrix}$$

$$0 \quad 17 \quad 10 \quad 4.9$$

$$1 \quad 52 \quad 40 \quad 3.6$$

$$2 \quad 54 \quad 60 \quad 0.6$$

$$3 \quad 31 \quad 40 \quad 2.025$$

$$4 \quad 6 \quad 10 \quad 1.6$$

$$\underline{12.725}$$

$$\therefore \chi^2_{(cat)} = \sum \left[\frac{(o_i - e_i)^2}{e_i} \right] = 12.725 //$$

In Binomial distⁿ $\Rightarrow d.f = n-1 = 5-1 = 4$

$$\therefore \chi^2_{(4)}(0.05) = 9.49 //$$

$\therefore \chi^2_{(cat)} > \chi^2_{(2)}$, we reject H_0

coins are not balanced.

Q5) A Sample Analysis of examination results of 200 MBAs was made it was found that 46 students had failed, 68 secured a 3rd division, 62 secured a 2nd division & rest were placed in 1st division. Are these figures commensurate with the general examination result which is in the ratio 4:3:2:1 for various categories respectively.

Soln: H_0 : The data commensurate with general examination result.

H_1 : The data do not commensurate with the general examination result.

To find theoretical freq (e_i):

The ratio = 4:3:2:1 wrt the categories

Fail, III division, II division, I division.

x_i : Failed III div II div I div

$$e_i: \frac{4}{10} \times 200 = 80 \quad \frac{3}{10} \times 200 = 60 \quad \frac{2}{10} \times 200 = 40 \quad \frac{1}{10} \times 200 = 20$$

	O_i	e_i	$(O_i - e_i)^2 / e_i$
x_i	46	80	14.45
F	68	60	1.0667
III	62	40	12.10
I	24	20	0.8
			<u>28.417</u>

$$\therefore \chi^2_{(cal)} = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 28.417 //$$

$$d.f \Rightarrow n-1 = 4-1 = 3$$

$$\therefore \chi^2_{(3)}(0.05) = 7.815$$

$$\therefore \chi^2_{(cal)} > \chi^2_{(2)}, \text{ we reject } H_0$$

the data do not commensurate with the general examination result.

Q) A survey of 800 families with 4 children each revealed the following distribution.

	0	1	2	3	4
No. of boys	4	3	2	1	0
No. of girls	4	3	2	1	0
No. of families	32	178	290	236	64

Is this result consistent with the hypothesis that male & female births are equally probable.

Soln:- H_0 : Male & female births are equally probable

H_1 : Male & female births are not equally probable.

To find theoretical frequencies,

here, the experiment follows Binomial distⁿ with n & p

X are obtaining no. of boys. i.e., $X \sim B(n, p)$

$p \Rightarrow$ prob. of success $\Rightarrow$ prob. of getting a boy = $1/2$

$q \Rightarrow$ prob. of failure $\Rightarrow$ prob. of getting a girl = $1/2$

$$n \Rightarrow 4$$

$$p(x) = {}^n C_x p^x q^{n-x} ; x=0,1,2,3,4$$

$\searrow$ *

$$x_i : 0 \quad 1 \quad 2 \quad 3 \quad 4$$

$$p(x_i) : \frac{1}{16} \quad \frac{4}{16} \quad \frac{6}{16} \quad \frac{4}{16} \quad \frac{1}{16}$$

$$e_i : \frac{1 \times 800}{16} \quad \frac{4 \times 800}{16} \quad \frac{6 \times 800}{16} \quad \frac{4 \times 800}{16} \quad \frac{1 \times 800}{16}$$
$$= 50 \quad = 200 \quad = 300 \quad = 200 \quad = 50$$

$$x_i \quad o_i \quad e_i \quad (o_i - e_i)^2 / e_i$$

$$0 \quad 32 \quad 50 \quad 6.48$$

$$1 \quad 178 \quad 200 \quad 2.42$$

$$2 \quad 290 \quad 300 \quad 0.333$$

$$3 \quad 236 \quad 200 \quad 6.48$$

$$4 \quad 64 \quad 50 \quad 3.92$$

$$\underline{\underline{19.633}}$$

$$\therefore \chi^2_{(cal)} = \sum \left[\frac{(o_i - e_i)^2}{e_i} \right] = 19.63$$

$$\text{In binomial distn} \Rightarrow d.f = n-1 = 5-1 = 4$$

$$\therefore \chi^2_{(4)} (0.05) = 9.49$$

$$\therefore \chi^2_{(cal)} > \chi^2_{(2)} , \text{ we reject } H_0$$

Male & female births are not equally probable.

57) When the first book of 392 pages of book of 1200 are read, the distribution of printing mistakes were found to be as follows:-

No. of mistakes in a page (x) :	0	1	2	3	4	5	6
No. of pages :	275	72	30	7	5	2	1

fit a poisson distribution to the above data & test the goodness of fit.

- H_0 : Poisson distribution is a good fit.
 H_1 : Poisson distribution is not a good fit.

To find the theoretical freq (e_i):
 As poisson distribution is a good fit we have the means as

$$\lambda = \bar{x} = \frac{\sum x_i f_i}{N} = \frac{0(275) + 1(72) + 2(30) + 3(7) + 4(5) + 5(2) + 6(1)}{392}$$

$$\Rightarrow \frac{189}{392} = 0.48211$$

$$\therefore \boxed{\lambda = 0.482}$$

To find the p.m.f of Poisson distn is given by:

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, 3, 4, 5, 6$$

x_i :	0	1	2	3	4	5	6
$p(x_i)$:	0.6175	0.298	0.072	0.012	0.0014	0.00013	0.0000108
e_i :	242.1	116.82	28.224	4.704	0.55	0.051	0.004
$p(x_i) \times 392$							

$$p(0) = e^{-0.482} = 0.6175$$

$$p(1) = e^{-0.482} \times (0.482) = 0.298$$

$$p(2) = 0.072$$

$$p(3) = 0.012$$

$$p(4) = 0.0014$$

$$p(5) = 0.00013$$

$$p(6) = 0.0000108$$

x_i	O_i	e_i	O_i	O_i	O_i	e_i	$(O_i - e_i)^2 / e_i$
0	275	242.1	275	242.1	275	242.1	4.471
1	72	116.8	72	116.8	72	116.8	17.1
2	30	28.2	30	28.2	30	28.2	0.115
3	7	4.7	7	4.7	15	5.3	17.75
4	5	0.5	8	0.6			<u>39.44</u>
5	2	0.1					
6	1	0					

$$\chi^2_{(cal)} = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 39.44$$

Poisson distribution $\Rightarrow$ d.f $\Rightarrow (n-1) - 3$

$$\Rightarrow n - 5$$

$$(or) n - 2$$

$$= 7 - 5$$

$$= 4 - 2$$

$$\Rightarrow 2$$

$$= 2$$

$$\therefore \chi^2_{(2), 0.05} = 5.99 \quad \text{from table.}$$

$$\therefore \chi^2_{(cal)} > \chi^2_{(2)} \Rightarrow \text{we reject } H_0$$

$\Rightarrow$ Poisson distribution is not a good fit.

Q) Fit a Poisson distn to the following data and test for goodness of fit @ 5% Los.

$x:$	0	1	2	3	4
$f:$	419	352	154	56	19

$$[\chi^2_{(0.05)} = 7.81 \text{ for } 3 \text{ d.f}]$$

547
 H_0 : Poisson distⁿ is a good fit.
 H_1 : Poisson distⁿ is not a good fit for the data.
 To find theoretical frequency (e_i).

Poisson distⁿ is a good fit $\Rightarrow \lambda = \bar{x} = \frac{\sum f_i x_i}{N} = \frac{904}{1000}$

$\Rightarrow \lambda = 0.904$

As $X \sim P(\lambda)$, $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$; $x = 0, 1, 2, 3, 4$

$p(0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-0.904} = 0.405$

$p(1) = \frac{\lambda^1 e^{-\lambda}}{1!} = \frac{(0.904) e^{-0.904}}{1} = 0.37$

$p(2) = \frac{\lambda^2 e^{-\lambda}}{2!} = \frac{(0.904)^2 e^{-0.904}}{2} = 0.17$

$p(3) = \frac{\lambda^3 e^{-\lambda}}{3!} = \frac{(0.904)^3 e^{-0.904}}{3!} = 0.05$

$p(4) = \frac{\lambda^4 e^{-\lambda}}{4!} = \frac{(0.904)^4 e^{-0.904}}{4!} = 0.01$

$x_i: 0 \quad 1 \quad 2 \quad 3 \quad 4$

$p(x_i): 0.405 \quad 0.37 \quad 0.17 \quad 0.05 \quad 0.01$

$e_i: 0.405 \quad 0.37 \quad 0.17 \times 1000 \quad 0.05 \quad 0.01$
 $\times 1000 \quad \times 1000 \quad = 170 \quad \times 1000 \quad \times 1000$
 $= 405 \quad 370 \quad = 50 \quad = 10$

$\chi^2_{(cal)} = 11.686$

$x_i \quad O_i \quad e_i \quad (O_i - e_i)^2 / e_i$

0 419 405 0.484

1 352 370 0.876

2 154 170 1.506

3 56 50 0.72

4 19 10 8.1
 11.686

Poisson distribution d.f $\Rightarrow n - 2 = 5 - 2 = 3$

$\chi^2_{(3)}(0.05) = 7.81$

$\chi^2_{(cal)} > \chi^2_{(2)} \Rightarrow$ we reject H_0

$\therefore$ Poisson distribution is a good fit.

Q) The theory predicts the proportions of beans in the four groups - G_1, G_2, G_3 & G_4 should be in ratio $9:3:3:1$. In experiment with 1600 beans the numbers in the groups were 882, 313, 287 & 118.

Does the experiment support the theory?

$$[\chi^2_{(3)}(0.05) = 7.815]$$

Soln:- H_0 : The experiment support the theory
 H_1 : The experiment do not support the theory.

Given,

x_i	G_1	G_2	G_3	G_4
O_i	882	313	287	118

To find the expected or theoretical freq (e_i)

Ratio $\Rightarrow 9:3:3:1$ w.r.t G_1, G_2, G_3 & G_4

x_i	G_1	G_2	G_3	G_4
e_i	$\frac{9}{16} \times 1600$ $= 900$	$\frac{3}{16} \times 1600$ $= 300$	$\frac{3}{16} \times 1600$ $= 300$	$\frac{1}{16} \times 1600$ $= 100$

x_i	O_i	e_i	$(O_i - e_i)^2 / e_i$
G_1	882	900	0.36
G_2	313	300	0.56
G_3	287	300	0.56
G_4	118	100	3.24
			<u>4.72</u>

$$\chi^2_{(cal)} = \sum \left(\frac{(O_i - e_i)^2}{e_i} \right)$$

$$= 4.72$$

$$d.f \Rightarrow n-1 = 4-1 = 3$$

$$\chi^2_{(3)}(0.05) = 7.815$$

$\therefore \chi^2_{(cal)} < \chi^2_{(2)}$, we accept H_0 .

The experiment support the theory.

Q10) Four coins are tossed 100 times and the following results were obtained. Fit a Binomial distn for the data and calculate the theoretical frequencies.

no. of heads :	0	1	2	3	4
Frequencies :	5	29	36	25	5

[Given $\chi^2_{(0.05)} = 9.49$ for 4 d.f.]

Sol:- H_0 : Binomial distn is a good fit.

H_1 : Binomial distn is not a good fit

$X \Rightarrow$ r.v of getting a head.

$X \sim B(n, p)$; $p \Rightarrow$ prob of getting head $\Rightarrow 1/2$
 $q \Rightarrow$ prob of getting tail $\Rightarrow 1/2$

$p(x) = {}^nC_x p^x q^{n-x} \rightarrow (*) \quad x=0, 1, 2, 3, 4$

to find theoretical or expected frequencies (e_i)

x_i :	0	1	2	3	4
$p(x_i)$:	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$
$e_i =$	$1/16 \times 100$	$4/16 \times 100$	$6/16 \times 100$	$4/16 \times 100$	$1/16 \times 100$
	$= 6.25 \approx 6$	$= 25$	$= 38$	$= 25$	$= 6$

$(*) \Rightarrow$
 $p(0) = {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 = \frac{1}{16}$
 $p(1) = {}^4C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3 = \frac{4}{16}$
 $p(2) = {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{6}{16}$
 $p(3) = {}^4C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right) = \frac{4}{16}$
 $p(4) = {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0 = \frac{1}{16}$

x_i	O_i	e_i	O_i	e_i	$(O_i - e_i)^2 / e_i$
0	5	6	34	31	0.29
1	29	25	36	38	0.11
2	36	38	30	31	0.03
3	25	25			
4	5	6			
					<u>0.43</u>

$\chi^2_{(cal)} = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 0.43$

$\chi^2_{(2)}(\alpha) = 5.99$

$\therefore \chi^2_{(cal)} < \chi^2(\alpha)$, we accept H_0 .

$\Rightarrow$ Binomial distribution is a good fit.

911) The following data shows the no. of aircraft accidents that occurred during the various days of a week. Test whether the accidents are uniformly distributed over the week $[\chi^2_{5\%} (\nu=5) = 11.07]$

Day :	Mon	Tue	Wed	Thurs	Fri	Sat
No. of accidents:	15	19	13	12	16	15

Soln:- H_0 : Accidents are uniformly distributed over the week
 H_1 : Accidents are not uniformly distributed over the week.

To find theoretical or expected frequencies (e_i)

Accidents are uniformly distributed over the week.

$$\Rightarrow e_i = \frac{N}{5} = \frac{75}{5} = 15$$

$$N = \sum O_i = 15 + 19 + 13 + 12 + 16 + 15 = 75$$

x_i	O_i	e_i	$(O_i - e_i)^2 / e_i$
1 Mon	15	15	0
2 Tue	19	15	1.07
3 Wed	13	15	0.27
4 Thurs	12	15	0.6
5 Fri	16	15	0.07
6 Sat	15	15	0
			<u>2.01</u>

$$\chi^2_{(cal)} = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 2.01$$

$$d.f \Rightarrow n-1 = 6-1 = 5$$

$$\chi^2_{5\%} (\nu=5) = 11.07$$

$$\chi^2_{(cal)} < \chi^2_{(\alpha)} \Rightarrow \text{we accept } H_0$$

$\therefore$ accidents are uniformly distributed over the week.

Q12) In an experiment on pea breeding, the following frequencies of seeds were obtained.

Round & Yellow	wrinkled & Yellow	Round & Green	wrinkled & Green	Total
315	101	108	32	556

theory predicts that the frequencies should be in proportions 9:3:3:1. Examine the correspondence between theory and experiment. [$\chi^2_{(3)}(0.05) = 7.81$]

H_0 : Theory fitted to the experiment (discrepancy is negligible)

H_1 : Theory do not fit the experiment

To find theoretical or expected freq (e_i)

Given ratio $\Rightarrow 4:3:3:1$

x_i :	R & Y	w & Y	R & G	w & G
e_i :	$\frac{9}{16} \times 556$ $= 312.75$ ~ 313	$\frac{3}{16} \times 556$ $= 104.25$ ~ 104	$\frac{3}{16} \times 556$ $= 104.25$ ~ 104	$\frac{1}{16} \times 556$ $= 34.75$ ~ 35

x_i	O_i	e_i	$(O_i - e_i)^2$	$(O_i - e_i)^2 / e_i$
R & Y	315	313	4	0.013
w & Y	101	104	9	0.087
R & G	108	104	16	0.154
w & G	32	35	9	0.257
				<u>0.511</u>

$$\therefore \chi^2 = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 0.511$$

$$d.f \Rightarrow \nu = n - 1 = 4 - 1 = 3$$

$$\therefore \chi^2_{\nu}(\alpha) = \chi^2_{(3)}(0.05) = 7.81$$

$\therefore \chi^2 < \chi^2_{\nu}(\alpha) \Rightarrow$ we accept $H_0 \Rightarrow$ Theory is fitted to experiment

Q13) A die was thrown 60 times and the following frequency distn was observed

Faces : 1 2 3 4 5 6

frequency : 15 6 4 7 11 17

Test whether the die is unbiased at 5% level of significance.

Soln: H_0 : The die is unbiased
 H_1 : The die is not unbiased

To find expected or theoretical freq (e_i)

As the die is unbiased, $p(x_i) = \frac{1}{6}$

x_i : 1 2 3 4 5 6

$p(x_i)$: $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$ $\frac{1}{6}$

e_i : $60 \times \frac{1}{6}$ $60 \times \frac{1}{6}$ $60 \times \frac{1}{6}$ $60 \times \frac{1}{6}$ $60 \times \frac{1}{6}$ $60 \times \frac{1}{6}$
 $= 10$ $= 10$ $= 10$ $= 10$ $= 10$ $= 10$

x_i	O_i	e_i	O_i	e_i	$(O_i - e_i)^2 / e_i$
1	15	10	15	10	2.5
2	6	10	10	20	5
3	4	10	7	10	0.9
4	7	10	11	10	0.1
5	11	10	17	10	4.9
6	17	10			<u>13.4</u>

$$\therefore \chi^2 = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 13.4$$

Here, d.f $\Rightarrow \nu = (n-1) - 1 \Rightarrow 5-1=4$

$$\therefore \chi^2_{\nu}(\alpha) = \chi^2_{(4)}(0.05) = 9.49$$

$\therefore \chi^2 > \chi^2_{\nu}(\alpha) \Rightarrow$ we reject H_0
 $\Rightarrow$ the die is not unbiased

814) A survey of 320 families with 5 children each revealed the following distribution

	5	4	3	2	1	0
No. of boys:						
No. of girls:	0	1	2	3	4	5
No. families:	14	56	110	88	40	12

Is the result consistent with the hypothesis that male and female births are equally probable at 5% level of significance.

Soln:-
 H_0 : Male and female births are equally probable
 H_1 : Male and female births are not equally probable.

To find theoretical frequency (e_i)

Male & female births are equally probable

$\Rightarrow$ probability of male birth $\Rightarrow p = 1/2$

$\Rightarrow$ probability of female birth $\Rightarrow q = 1/2$

Let X be the no. of male births in a family of 5 children $\Rightarrow X \sim B(n, p)$

$$\therefore p(x) = {}^nC_x p^x q^{n-x}; x=0, 1, 2, 3, 4, 5$$

①

x_i : 0 1 2 3 4 5

$p(x_i)$: $\frac{1}{32}$ $\frac{5}{32}$ $\frac{10}{32}$ $\frac{10}{32}$ $\frac{5}{32}$ $\frac{1}{32}$

e_i : $\frac{320 \times 1}{32}$ $\frac{320 \times 5}{32}$ 100 100 50 10
 $= 10$ $= 50$

by ①

$$p(0) = {}^5C_0 p^0 q^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$$p(1) = {}^5C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = \frac{5}{32}$$

$$p(2) = {}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = \frac{10}{32}$$

$$p(3) = {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

$$p(4) = {}^5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right) = \frac{5}{32}$$

$$p(5) = {}^5C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = \frac{1}{32}$$

x_i	O_i	e_i	$(O_i - e_i)^2$	$(O_i - e_i)^2 / e_i$
5	14	10	16	1.6
4	56	50	36	0.72
3	110	100	100	1
2	88	100	144	2
1	40	50	100	0.4
0	12	10	4	<u>7.16</u>

$$\therefore \chi^2 = \sum \left[\frac{(O_i - e_i)^2}{e_i} \right] = 7.16$$

$$d.f \Rightarrow \gamma = n - 1 = 6 - 1 = 5 //$$

$$\chi^2_{\gamma}(\alpha) = \chi^2_{(5)}(0.05) = 11.07$$

$$\therefore \chi^2 < \chi^2_{\gamma}(\alpha) \Rightarrow \text{we accept } H_0$$

$\Rightarrow$ Male & female births are equally probable.