SEMESTER : V
MAJOR BASED ELECTIVE : I

Inst Hour	**: 5**
Credit	**: 5**
Code	**: 18K5MELMIS**

PROBABILITY AND STATISTICS

UNIT 1 :

Theory of Probability : Different definitions of Probability – Sample space – Probability of an event – Independence of events – Theorems of Probability – Conditional Probability – Baye's Theorem.
(Chapter 4 : Sections 4.5 – 4.9)

UNIT 2 :

Random variables – Distribution functions – Discrete & Continuous random variables – Probability mass & density functions – Joint probability distribution functions.
(Chapter 5 : Sections 5.1 – 5.5.5)

UNIT 3 :

Expectation – Variance – Covariance – Moment generating functions – Theorems on Moment generating functions – Moments – Various measures.
(Chapter 6: Sections 6.1 to 6.10.3 & Chapter 3 : Section 3.9)

UNIT 4 :

Correlation & Regression **:** Properties of Correlation & Regression coefficients – Numerical Problems for finding the correlation & regression coefficients.
(Chapter 10 : Sections 10.1 to 10.7.4)

UNIT 5 :

Binomial, Poisson, Normal distributions – Moment generating functions of these distributions- additive properties of these distributions – Recurrence relations for the moments about origin and mean for the Binomial, Poisson and Normal distributions – Properties of normal distributions.
(Chapter 7 :Sections 7.2 to 7.2.7, 7.2.10, 7.3 to 7.3.5, 7.3.8 and Chapter 8 :Sections 8.2, 8.2.2)

Text Book :

[1]. Fundamental of Mathematical Statistics by Gupta. S.C & Kapoor, V.K. Published by Sultan Chand & Sons, New Delhi – 2000 Edition.

Book For Reference :-

1]. Practical Statistics – Thambidurai . P – Rainbow Publishers – CBE (1991)
2]. Probability and Statistics – A. Singaravelu – A.R. Publications -2002

Question Pattern

Section A : 10 x 2 = 20 Marks, 2 Questions from each Unit.
Section B : 5 x 5 = 25 Marks, EITHER OR (a or b) Pattern, One question from each unit.
Section C : 3 x 10 = 30 Marks, 3 out of 5, One Question from each Unit.

Unit - 3

Mathematical Expectations.

let X be a continuous random variable with probability density fn $f(x)$

The mathematical expectation of X is denoted by $E(X)$ and is given by

$$E(X) = \int_{-\infty}^{\infty} x f(x) \quad \text{[for continuous random variable]}$$

$$= \sum_{x} x f(x) \quad \text{[for discrete random variable]}.$$

rth moment (about origin).

For the probability distribution $f(x)$ the rth moment (about origin) is defined as

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x)\,dx$$

$$= E(x^r)$$

Thus $\mu_1' = E(x)$

$\mu_2' = E(x^2)$

$\therefore$ Mean $= \bar{x} = \mu_1' = E(x)$

and variance $= \mu_2 = \mu_2' - \mu_1'^2$

$= E(x^2) - [E(x)]^2$.

Note: The above result gives the variance in terms of expectation.

Now $E\{x - E(x)\}^r = \int_{-\infty}^{\infty} \{x - E(x)\}^r f(x)\,dx$.

$$= \int_{-\infty}^{\infty} \{x - \bar{x}\}^r f(x)\,dx.$$

This gives the rth moment about mean and it is denoted by μ_r.

Thus $\mu_r = \int_{-\infty}^{\infty} (x-\bar{x})^r f(x)\,dx$.

put $r=1$ we get

$$\mu_1 = \int_{-\infty}^{\infty} (x-\bar{x}) f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x f(x)\,dx - \int_{-\infty}^{\infty} \bar{x} f(x)\,dx$$

$$= \bar{x} - \bar{x} \int_{-\infty}^{\infty} f(x)dx.$$

$$= \bar{x} - \bar{x} \qquad \left[\because \int_{-\infty}^{\infty} f(x)dx = 1\right]$$

$$= 0.$$

put $r = 2$ we get

Variance $= \mu_2 = E\left[\{x - E(x)\}^2\right]$

$= \int_{-\infty}^{\infty} (x-\bar{x})^2 f(x)dx$. which gives the variance in terms of expectations.

Note: let $g(x) = k$ (constant), then

$$E[g(x)] = E(k) = \int_{-\infty}^{\infty} k f(x) dx$$

$$= k \int_{-\infty}^{\infty} f(x) dx.$$

$$= k.1 \qquad \left[\because \int_{-\infty}^{\infty} f(x) dx = 1\right]$$

$$= k$$

For discrete random vareables 'x'

$$E(x^r) = \sum_x x^r f(x)$$

$$\mu_r' = E(x^r) = \sum_x x^r f(x)$$

put $r = 1$ we get

Mean $= \mu_1' = \sum x f(x)$

Variance $= \mu_2 = \mu_2' - \mu_1'^2 = E(x^2) - \{E(x)\}^2$

The r^{th} moment about mean

$$\mu_r = E[\{x - E(x)\}^r]$$

$$= \sum_x (x - \bar{x})^r f(x), \quad E(x) = \bar{x}$$

put $r = 2$ we get

variance $= \mu_2 = \sum (x - \bar{x})^2 f(x)$.

Addition thm Expectation :

1) If x and y are random variables then

$$E(x+y) = E(x) + E(y).$$

<u>proof:</u> let x and y be continuous random variables with marginal p.d.f $f_x(x)$ and $f_y(y)$ and whose joint p.d.f $f_{xy} = (x, y)$.

Then $E(x) = \int_{-\infty}^{\infty} x f_x(x)\, dx$ — ①

$E(y) = \int_{-\infty}^{\infty} y f_y(y)\, dy$ — ②

$$E(x+y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{xy}(x,y)\,dx\,dy.$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{xy}(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{xy}(x,y)\,dx\,dy.$$

$$= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f_{xy}(x,y)\,dy\right] dx + \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f_{xy}(x,y)\,dx\right] dy$$

$$E(x+y) = \int_{-\infty}^{\infty} x f_x(x)\,dx + \int_{-\infty}^{\infty} y f_y(y)\,dy$$

$$= E(x) + E(y)$$

[Using m.d.f of Y $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$.

m.d.f of X, $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$.

Multiplication thm Expectations :-

If X and Y are independent variables, then.

$$E(XY) = E(X)\cdot E(Y).$$

Proof!

let X and Y be continuous random variables with joint p.d.f $f_{XY}(x,y)$ and marginal p.d.fs $f_X(x)$ and $f_Y(y)$ respectively.

WKT, $E(X) = \int_{-\infty}^{\infty} x\, f(x)\,dx$.

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\,dy.$$

$$\text{Now } E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\,dx\,dy.$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x)\cdot f_Y(y)\,dx\,dy.$$

[since X and Y are independent]

$$= \int_{-\infty}^{\infty} x\, f_X(x)\,dx \int_{-\infty}^{\infty} y\, f_Y(y)\,dy.$$

$$E(XY) = E(X)\cdot E(Y).$$

Note: If $X_1, X_2, X_3 \ldots X_n$ are independent random variables then $E[X_1, X_2 \ldots X_n] = E(X_1)\,E(X_2)\cdots E(X_n)$

Thm 3: If 'x' is a random variable and 'a' is a constant, then

(i) $E[a\,G(x)] = a\,E[G(x)]$.

(ii) $E[G(x)+a] = E[G(x)]+a$ where $G(x)$ is a fn of 'x' which is also a random variable :-

Proof: (i) $E[a\,G(x)] = \int_{-\infty}^{\infty} a\,G(x)\cdot f(x)\,dx$.

$$= a\int_{-\infty}^{\infty} G(x)\,f(x)\,dx$$

$$= a\,E[G(x)].$$

(ii) $E[G(x)+a) = \int_{-\infty}^{\infty} [G(x)+a]\,f(x)\,dx$

$$= \int_{-\infty}^{\infty} G(x)\,f(x)\,dx + a\int_{-\infty}^{\infty} f(x)\,dx$$

$$= E[G(x)] + a \quad \left[\because \int_{-\infty}^{\infty} f(x)\,dx = 1\right]$$

Thm If x is a random variable and 'a' and 'b' are constants Then

$$E(ax+b] = a\,E(x)+b.$$

Proof: $E[ax+b] = \int_{-\infty}^{\infty} (ax+b)\, f(x)\, dx$

$$= a\int_{-\infty}^{\infty} x f(x)\, dx + b\int_{-\infty}^{\infty} f(x)\, dx$$

$$= a\,E(x) + b.$$

cor: 1 $\Rightarrow$ If $b=0$ then we get.

$$E(ax] = a\,E[x].$$

cor: 2 $\Rightarrow$ If we take $a=1$ and $b=-E(x)=-\bar{x}$

then we get

$$E(x-\bar{x}) = E(x) - E(x) = 0.$$

cor: 3 $\Rightarrow$ let $g(x) = ax+b$.

$$\therefore g[E(x)] = aE(x)+b. \quad \text{--- (a)}$$

But $E[ax+b] = aE(x)+b$.

$$E[g(x)] = aE(x)+b. \quad \text{--- (b)}$$

From (a) and (b) we get.

$$E[g(x)] = g\,E(x)$$

Note:

$$E(1/x) \neq \frac{1}{E(x)}$$

$$E[\log(x)] \neq \log E(x).$$

$E(x^2) \neq [E(x)]^2$.

Expectation of a linear combination of Random variables.

let $x_1, x_2 \ldots x_n$ be any 'n' random variab[les] and if $c_1, c_2 \ldots c_n$ are constants, then.

$$E[a_1x_1 + a_2x_2 + \cdots + a_nx_n].$$

$$= a_1E(x_1) + a_2E(x_2) + \cdots + a_nE(x_n)$$

m: let x and y be two random variables such that $y \leq x$, then. $E(y) \leq E(x)$.

Proof: Given $y \leq x$.

$$\Rightarrow \quad 0 \leq x - y.$$

$$x - y \geq 0.$$

$$E(x-y) \geq 0.$$

$$E(x) - E(y) \geq 0$$

$$E(x) \geq E(y).$$

$$E(y) \leq E(x)$$

Note: $|E(x)| \leq E|x|$.

Result: If x is a random variable, then

$V(ax+b) = a^2V(x)$ where 'a' and 'b' are constants.

proof:

let $y = ax + b$ —(1)

Then $E(y) = aE(x) + b$ —(2)

(1) − (2) $y - E(y) = a[x - E(x)]$

$$\{y - E(y)\}^2 = a^2 \{x - E(x)\}^2$$

$$E\{y - E(y)\}^2 = a^2 E\{x - E(x)\}^2$$

$\Rightarrow$ $V(y) = a^2 V(x)$ {using $V(x) = E[x - E(x)]^2$

$[\because y = ax + b]$.

$\Rightarrow$ $V(ax + b) = a^2 V(x)$

Hence $V(ax + b) = a^2 V(x)$.

where 'a' and 'b' are constants.

covariance:

If x and y are random variables, then covariance b/w them is defn as

$$\operatorname{cov}(x, y) = E\{[x - E(x)][y - E(y)]\}$$

$$= E\{xy - xE(y) - E(x)y + E(x)E(y)\}$$

$$= E(xy) - E(x)E(y) - E(x)E(y) + E(x)E(y)$$

$$\operatorname{cov}(xy) = E(xy) - E(x) \cdot E(y) \quad \text{—(A)}$$

If x and y are independent, then

$$E(xy) = E(x)E(y) \quad \text{—(B)}$$

sub (B) in (A) we get

$$\operatorname{cov}(x, y) = 0.$$

$\therefore$ If x and y are independent then

$cov(x, y) = 0$

Note : 1

$cov(ax, by) = ab\, cov(x, y)$

2. $cov(x + a, y + b) = cov(x, y)$

3. $cov(ax + b, cy + d) = ac\, cov(x, y)$

4. $V(x_1 + x_2) = V(x_1) + V(x_2) + cov(x_1, x_2)$

5. $V(x_1 - x_2) = V(x_1) + V(x_2) - 2\,cov(x_1, x_2)$.

If x_1 and x_2 are independent.

$V(x_1 \pm x_2) = V(x_1) + V(x_2) \pm 2\,cov(x_1, x_2)$.

Moment Generating function :

The moment generating function (m.g.f) of a random variable 'x' (about origin) whose probability fn $f(x)$ is given by

$$M_x(t) = E(e^{tx}).$$

$$= \begin{cases} \int_{-\infty}^{\infty} e^{tx} f(x)\,dx & \text{for continuous probability fn.} \\ \sum_{-\infty}^{\infty} e^{tx} f(x) & \text{for discrete probability fn.} \end{cases}$$

To find the rth moment about origin.

WKT,

$$M_x(t) = E(e^{tx})$$

$$= E\left[1 + tx + (tx)^2 + (tx)^3 + \dots + (tx)^r + \dots\right]$$

$= 1 + E(tx) + E(t^2x^2) + \cdots + E(t^r x^r) + \cdots$

$= 1 + tE(x) + \frac{t^2}{2!} E(x^2) + \cdots + \frac{t^r}{r!} E(x^r) + \cdots$

$M_x(t) = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots$ (A)

$M_x(t) = \sum_{r=0}^{\infty} t^r/r!\, \mu_r'$ [using $\mu_r' = E(x^r)$]

This gives the m.g.f in terms of moments. Thus

the co-eff of $t^r/r!$ in $M_x(t)$ gives the r^{th} moment

about origin (μ_r').

since $M_x(t)$ generates moments it is known as

moment generating fn.

Note: Diff (A) wrt t we get,

$M_x'(t) = \mu_1' + \frac{2t}{2!}\mu_2' + \frac{3t^2}{3!}\mu_3' + \cdots$ (B)

put $t = 0$ in (B) we get.

$\mu_1' = M_x'(0)$.

The 1^{st} moment about origin is given by

$\mu_1' = M_x'(0)$.

Diff (B) wrt 't' we get.

$M_x''(t) = \mu_2' + t\mu_3' + \cdots$ (C)

put $t = 0$ in (D) we get

$M_x''(0) = \mu_2'$.

Hence the 2nd moment about origin is given by $\mu_2' = M_X''(0)$.

Note: The moment generating fn of x about the pt $x = a$ is given by

$$M_X(t) = E[e^{t(x-a)}]$$

$$= E\left[1 + t(x-a) + \frac{t^2}{2!}(x-a)^2 + \cdots + \frac{t^r}{r!}(x-a)^r + \cdots\right]$$

$$= 1 + E(t(x-a)) + E\left[\frac{t^2}{2!}(x-a)^2\right] + \cdots + E\left[\frac{t^r}{r!}(x-a)^r\right] \cdots$$

$$= 1 + tE(x-a) + \frac{t^2}{2!}E(x-a)^2 + \cdots + \frac{t^r}{r!}E(x-a)^r + \cdots$$

$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots$$

Thus

$$M_X(t)_{x=a} = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots$$

where $\mu_r' = E[x-a]^r$ which gives the r^{th} moment about the point $x = a$.

Result - 1 : $M_{cX}(t) = E(e^{tcx})$ — ①

$M_X(ct) = E(e^{ctx})$ — ②.

From ① & ② we get,

$$M_{cX}(t) = M_X(ct).$$

Thm : If $x_1, x_2 \cdots x_n$ are independent random variables then

$$M_{x_1+x_2+\cdots+x_n}(t) = M_{x_1}(t) M_{x_2}(t) \cdots M_{x_n}(t).$$

<u>proof</u>: By defn.

$$M_{x_1+x_2+\cdots+x_n}(t) = E\left[e^{t(x_1+x_2+\cdots x_n)}\right]$$

$$= E\left[e^{tx_1} \cdot e^{tx_2} \cdots e^{tx_n}\right]$$

$$= E(e^{tx_1}) \cdot E(e^{tx_2}) \cdots E[e^{tx_n}]$$

[$x_1, x_2 \cdots x_n$ are independent]

$$= M_{x_1}(t) \cdot M_{x_2}(t) \cdots M_{x_n}(t)$$

Thm: If $U = \frac{x-a}{h}$, then

$M_U(t) = e^{-\frac{at}{h}} \cdot M_x(t/h)$ a, h are constants

<u>Proof</u>: By defn.

$$M_u(t) = E[e^{tU}]$$

$$= E\left[e^{t\left(\frac{x-a}{h}\right)}\right]$$

$$= E\left[e^{\frac{tx}{h}} \cdot e^{\frac{-at}{h}}\right].$$

$$= e^{\frac{-at}{h}} \cdot E\left[e^{(1/h)x}\right]$$

$$= e^{-at/h} \cdot M_x(t/h) \text{ (By defn)}$$

$$\therefore M_U(t) = e^{-at/h} \cdot M_x(t/h) \text{ where } U = \frac{x-a}{h}.$$

Ex:1) Find the m.g.f of the random variable with the probability law.

$$P(X=x) = q^{x-1} p, \quad x = 1, 2, 3 \cdots$$

Soln: WKT,

$$M_X(t) = E(e^{tx})$$

$$= \sum_{x=1}^{\infty} e^{tx} \cdot p(x) \quad [\text{By defn}].$$

$$= \sum_{x=1}^{\infty} e^{tx} \cdot q^{x-1} p \cdot [\because p(x) = q^{x-1} p]$$

$$= \sum_{x=1}^{\infty} e^{tx} \cdot q^{x} \cdot q^{-1} p$$

$$= \sum_{x=1}^{\infty} (e^t q)^x \, p/q$$

$$= \frac{p}{q} \cdot q e^t \sum_{x=1}^{\infty} (qe^t)^{x-1}$$

$$= pe^t \left[1 + qe^t + (qe^t)^2 + \cdots\right]$$

$$= pe^t (1-qe^t)^{-1} \quad [\because (1-x)^{-1} = 1 + x + x^2 + \cdots]$$

$$M_X(t) = \frac{pe^t}{1-qe^t} \quad \text{--- (1)}.$$

Diff (1) w.r.t 't' we get.

$$\frac{d}{dt}\{M_X(t)\} = \frac{(1-qe^t)pe^t - pe^t(-qe^t)}{(1-qe^t)^2}$$

$$= \frac{pe^t - pqe^{2t} + pqe^{2t}}{(1-qe^t)^2}$$

$$M_X'(t) = pe^t / (1-qe^t)^2 \quad \text{--- (2)}$$

$\therefore \mu_1'(\text{about origin}) = M_x'(0)$.

$$= \frac{p}{(1-q)^2} \quad [\text{put } x=0 \text{ in } 2]$$

$$= \frac{p}{p^2}$$

$$= 1/p.$$

Diff 2 w.r.t 't' we get.

$$M_x''(t) = \frac{(1-qe^t)^2 pe^t - pe^t 2(1-qe^t)(-qe^t)}{(1-qe^t)^4}$$

$$= \frac{(1-qe^t)\{(1-qe^t)pe^t + 2pqe^{2t}\}}{(1-qe^t)^4}$$

$$= \frac{pe^t + pqe^{2t}}{(1-qe^{2t})^3}$$

$$M_x''(t) = \frac{pe^t(1+qe^t)}{(1-qe^t)^3} \quad —③$$

$\therefore \mu_2'(\text{about origin}) = M_x''(0)$

$$= \frac{p(1+q)}{(1-q)^3} \quad [\text{put } x=0 \text{ in } 3]$$

$$\text{Mean} = \mu_1' = 1/p$$

$$\text{Variance} = \mu_2' - \mu_1'^2$$

$$= \frac{1+q}{p^2} - 1/p^2$$

variance $= q/p^2$.

Ex 2 Find the m.g.f of the random variable whose moments are $\mu_r' = (r+1)!\,2^r$.

soln: WKT the m.g.f in terms of moments is given by

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r' \quad [\because \mu_r' = (r+1)!\,2^r]$$

$$= \sum_{r=0}^{\infty} (r+1)\,2^r \cdot \frac{t^r}{r!}$$

$$= \sum_{r=0}^{\infty} \frac{(2t)^r (r+1)!}{r!}$$

$$M_X(t) = 1 + 2 \cdot (2t) + 3(2t)^2 + \cdots$$

$$= (1-2t)^{-2}$$

[using $(1-x)^{-2} = 1 + 2x + 3x^2 + \cdots$]

Ex 3 If the moments of a random variable 'X' are defn $E(X^r) = 0.6$, $r = 1, 2, 3, \ldots$

S.T $P(X=0) = 0.4$, $P(X=1) = 0.6$, $P(X \geq 2) = 0$.

Soln: WKT,

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r' \quad [\text{Given } E(X^r) = \mu_r' = 0.6]$$

$$= 1 + \sum_{r=1}^{\infty} \frac{t^r}{r!} (0.6)$$

$$= 1 - 0.6 + 0.6 + \sum_{r=1}^{\infty} \frac{t^r}{r!} (0.6)$$

$$= 0.4 + 0.6 \left[1 + \sum_{r=1}^{\infty} t^r/r! \right]$$

$$M_X(t) = 0.4 + 0.6e^t \quad \text{--- (1)}$$

But $M_X(t) = E(e^{tx})$

$$= \sum_{x=0}^{\infty} e^{tx} \cdot p(x)$$

$$M_X(t) = p(0) + e^t p(1) + \sum_{x=2}^{\infty} e^{tx} p(x) \quad \text{--- (2)}$$

From (1) & (2) we get,

$$p(0) = 0.4$$

$$p(1) = 0.6$$

$$\sum_{x=2}^{\infty} e^{tx} p(x) = 0 \Rightarrow P(x) = 0, \ x \geq 2.$$

EX:4 Find the m.g.f of a random variable 'x' whose probability fn is $P(x) = \frac{1}{2^x}$ $x = 1, 2, 3 \ldots$ Hence find its mean.

Soln: WKT,

$$M_X(t) = \sum_{x=1}^{\infty} e^{tx} p(x)$$

$$= \sum_{x=1}^{\infty} e^{tx} \frac{1}{2^x}$$

$$= \sum_{x=1}^{\infty} \left(\frac{e^t}{2}\right)^x$$

$$= \frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + \left(\frac{e^t}{2}\right)^3 + \cdots$$

$$= \frac{e^t}{2}\left[1 + \frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + \cdots\right]$$

$$= \frac{e^t}{2}\left[1 - \frac{e^t}{2}\right]^{-1}$$

[Using $(1-x)^{-1} = 1+x+x^2+\cdots$]

$$= \frac{e^t}{2} \frac{(2-e^t)^{-1}}{2^{-1}}$$

$$M_X(t) = e^t (2-e^t)^{-1}$$

Now $M_X'(t) = -e^t (2-e^t)^{-2} + (2-e^t)^{-1} e^t$.

$$\mu_1' = M_X'(0)$$

Hence mean $= \mu_1' = 0$.

Ex:5 Find the m.g.f of the random variable 'x' having p.d.f $f(x) = \begin{cases} x & \text{for } 0 \le x < 1 \\ 2-x & \text{for } 1 \le x < 2 \\ 0 & \text{otherwise} \end{cases}$.

soln: WKT,

$$M_X(t) = \int_0^\infty e^{tx} f(x)\,dx \quad [\text{Here 'x' is a continuous variable}]$$

$$= \int_0^1 e^{tx} f(x)\,dx + \int_1^2 e^{tx} f(x)\,dx.$$

$$= \int_0^1 e^{tx}.x\,dx + \int_1^2 e^{tx}(2-x)\,dx.$$

$$= \int_0^1 e^{tx}.x\,dx + \int_1^2 e^{tx}(2-x)\,dx.$$

$$= \left\{ x\left(\frac{e^{tx}}{1}\right) - \left(\frac{e^{tx}}{t^2}\right)\right\}_0^1 + \left\{(2-x)\left(\frac{e^{tx}}{t}\right) - (-1)\left(\frac{e^{tx}}{t^2}\right)\right\}_1^2$$

[using integration by parts]

$$= \frac{e^t}{t} - \frac{e^t}{2t^2} + \frac{1}{t^2} + \frac{e^{2t}}{t^2} - \frac{e^t}{t} - \frac{e^t}{t^2}$$

$$= \frac{e^{2t}}{t^2} + \frac{1}{t^2} - \frac{2e^t}{t^2}.$$

$$M_X(t) = \frac{(e^t - 1)^2}{t^2}.$$

Ex:6 Find the m.g.f of a random variable 'y' having the p.d.f $f(x) = \frac{1}{3}, \; -1 < x < 2.$

$= 0$ otherwise.

Soln: WKT, the m.g.f for a continuous random variable X is

$$M_X(t) = \int_{\infty}^{-\infty} e^{tx} f(x)\,dx.$$

$$= \int_{-1}^{2} e^{tx} \cdot \tfrac{1}{3}\,dx$$

$$= \tfrac{1}{3}\left[\frac{e^{tx}}{t}\right]_{-1}^{2}$$

$$M_X(t) = \tfrac{1}{3}\left[\frac{e^{2t} - e^{t}}{t}\right]$$

chebychev's inequality :

If X is a random variable with mean μ and variance σ^2, then for any +ve no k, we have.

$$P\{|X-\mu| \geq k\sigma\} \leq 1/k^2$$

(or)

$$P\{|x-\mu| < k\sigma\} \geq 1-\frac{1}{k^2}.$$

soln: let X be a continuous random variable

By defn of variance we have.

$$\sigma^2 = \sigma_X^2 = E[X-E(X)]^2.$$

$$= E[X-\mu]^2, \quad \mu = E(X).$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \quad ; f(x) \text{ is p.d.f of } X.$$

[By defn of expectation for continuous random variable].

$$= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)dx.$$

$$+ \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx.$$

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx \quad \text{—(A)}$$

In the 1st integral x takes values less than or equal to $\mu - k\sigma$.

(ie) $x \le \mu - k\sigma \Rightarrow \mu - x \le k\sigma$.

In the 2nd integral x takes values greater than or equal to $\mu + k\sigma$.

$$x \ge \mu + k\sigma \Rightarrow x - \mu \ge k\sigma.$$

$$\mu - x \le -k\sigma \quad -(2).$$

sub (1) and (2) in (A) we get.

$$\sigma^2 \ge \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)\,dx.$$

$$= k^2\sigma^2 \left[\int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx \right]$$

$$= k^2\sigma^2 \{ P(x \le \mu - k\sigma) + P(x \ge \mu + k\sigma) \}$$

$$= k^2\sigma^2 \{ P(x - \mu \le -k\sigma) + P(x - \mu \ge k\sigma) \}$$

$$= k^2\sigma^2 P \{ |x - \mu| \ge k\sigma \}$$

$$\sigma^2 \ge k^2\sigma^2 P \{ |x - \mu| \ge k\sigma \}$$

$$P \{ |x - \mu| \ge k\sigma \} \le \frac{1}{k^2}$$

since total probability is 1 we have .

$$P\{|x-\mu| \geq k\sigma\} + P\{|x-\mu| < k\sigma\} = 1.$$

$$P\{|x-\mu| < k\sigma\} = 1 - P\{|x-\mu| \geq k\sigma\}$$

$$\geq 1 - 1/k^2 .$$

Ex:1 A random variable x has a mean $\mu = 12$ and a variance $\sigma^2 = 9$ and an unknown probability distribution Find $P(6 < x < 18)$.

soln: Given $\mu = 12$, $\sigma^2 = 9$ $\sigma = 3$

By chebychev's inequality we have .

$$P\{|x-\mu| < k\sigma\} \geq 1 - 1/k^2$$

$$P\{-k\sigma < x-\mu < k\sigma\} \geq 1 - 1/k^2$$

$$P\{\mu - k\sigma < x < \mu + k\sigma\} \geq 1 - 1/k^2 \quad \text{--- (1)}$$

$$\text{Now } P\{6 < x < 18\} = 1 - 1/k^2 \quad \text{--- (2)}$$

$$\left.\begin{array}{l} \mu - k\sigma = 6 \\ \mu + k\sigma = 18 \end{array}\right\} \Rightarrow \left.\begin{array}{l} 12 - 3k = 6 \\ 12 + 3k = 18 \end{array}\right\} \quad \text{(A)}$$

solving (A) we get $\boxed{k=2}$ sub $\boxed{k=2}$ in (2)

we get,

$$\therefore P\{6 < x < 18\} = 1 - 1/2^2 = 1 - 1/4 = 0.75$$

Try yourself:

A random Variable 'x' has a mean $\mu = 8$ variance $\sigma^2 = 9$ and an unknown probability distribution. Find $P\{|x-8| \geq 6\}$. $\left[\text{Ans}: \frac{15}{16}\right]$

Ex: 2 A random Variable 'x' has an mean 10 and a Variance 4 and an unknown probability distribution. Find the value of 'c' such that $P\{|x-10| \geq c\} \leq 0.04$.

soln: Given $\mu = 10$, $\sigma^2 = 4$ $\sigma = 2$.

By chebychev's inequality we have

$$P\{|X-\mu| \geq k\sigma\} \leq 1/k^2 \quad \text{— (1)}$$

Given $P\{|x-10| \geq c\} \leq 0.04$

(ie) $P\{|x-10| \geq c\} \leq 0.04$ — (2)

comparing (1) & (2) we get

$$\frac{1}{k^2} = 0.04 \text{ and } k\sigma = c.$$

$$\Rightarrow \frac{1}{k^2} = \frac{4}{100} \Rightarrow k^2 = \frac{100}{4}$$

$$\Rightarrow k = 10/2 \quad \text{— (3)}$$

and $k\sigma = c$.

$\therefore c = 2k$. $[\because \sigma = 2]$

$= 2 \times \frac{10}{2}$ [From (3)]

$\boxed{\therefore c = 10}$

Ex:3 The no. of planes landing at an airport in a 30 mins interval obey's the poisson law with mean 25, Use chebychev's inequality to find the least chance that the number of planes landing within a given 30 mins interval will be b/w 15 and 35.

Soln: Given $\mu = 25$.

$\therefore$ Variance $\sigma^2 = 25$ [For poisson distribution mean = variance chebychev's inequality is.

$$P\{|x-\mu| < k\sigma\} \geq 1 - 1/k^2.$$

(ie) $P\{-k\sigma < x-\mu < k\sigma\} \geq 1 - 1/k^2$

(ie) $P\{\mu - k\sigma < x < \mu + k\sigma\} \geq 1 - 1/k^2$.

$$P\{25 - k(5) < x < 25 < k(5)\} \geq 1 - 1/k^2 \quad \text{—(A)}$$

Now $P\{15 < x < 35\} = 1 - 1/k^2$.

(ii) From ① we get $25-5k=15$ and $25+5k=$

sub $\boxed{k=2}$ in (A) we get.

$$P\{15<x<35\}=1-1/4$$
$$=3/4.$$

Ex:4 If x is the no. on a die when it is thrown, P.T $P\{|x-\mu|>2.5\}<0.47$ where μ is the mean.

Soln: let 'x' be a random variable denotes the number on a die the probability function is given below.

x	1	2	3	4	5	6
$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6

Now $E(x)=\Sigma x\,p(x)=\frac{1}{6}(1+2+3+4+5+6)=\frac{7}{2}$

$E(x^2)=\Sigma x^2 p(x)=1/6\,[1^2+2^2+3^2+4^2+5^2+6^2]$

$$=91/6$$

Var (x) $=E(x^2)-[E(x)]^2$.

$$=\frac{91}{6}-\frac{49}{4}$$

$$\sigma^2=2.9167$$

$$\sigma=1.707.$$

By chebychev's inequality, we have

$$P\{|X-\mu| > k\sigma\} < 1/k^2 \quad \text{——} \ (1)$$

(ie) $$P\{|X-\mu| > 2.5\} < 1/k^2 \quad \text{—} \ (2)$$

comparing (1) and (2)

$$k\sigma = 2.5$$

$$k = \frac{2.5}{1.707} = 1.46 \qquad [\because \sigma = 1.707]$$

$$\therefore P\{|X-\mu| > 2.5\} = \cancel{1/4.6}\ \frac{1}{(1.46)^2}$$

$$< 0.47.$$

Try yourself:

Two unbiased dice are thrown. The random variable X represents the sum of the numbers showing up. Find E(x) and var (x)

Also prove that $P\{|X-7| \geq 2\} < \frac{35}{24}$

Property: 6 If $X \geq 0$ then $E(X) \geq 0$.

Proof: If X is a continuous random variables s,t $X \geq 0$ then.

$$E(X) = \int_{-\infty}^{\infty} x \cdot p(x)\,dx = \int_{0}^{\infty} x \cdot p(x)\,dx > 0$$

$[\because$ If $X \geq 0$, $P(x) = 0$ for $x < 0]$

Provided the expectation exists.

Property : 7 If X and Y are two random variables such that $Y \le X$ then $E(Y) \le E(X)$ provided all the expectations exist.

Proof : Since $Y \le X$, we have there r.v $Y - X \le 0$

$X - Y \ge 0$

Hence $E(X-Y) \ge 0 \Rightarrow E(X) - E(Y) \ge 0$

$E(X) \ge E(Y)$ $E(Y) \le E(X)$ as desired

Property : 8 $|E(X)| \le E|X|$, provided the expectation exist.

Proof : Since $X \le |X|$ we have property 7,

$E(X) \le E|X|$ — ①

Again since $-X \le |X|$ we have by property 7

$E(-X) \le E|X|$ — ②

$-E(X) \le E|X|$.

From ① and ② we get the desired result

$|E(X)| \le E|X|$.

Property : 9 If μ'_r exists for all $1 \le s \le r$ mathematically if $E(X^r)$ exists, then $E(X^s)$ exist for all $1 \le s \le r$ (ie).

$E(X^r) < \infty \Rightarrow E(X^s) < \infty \quad \forall\ 1 \le s \le r$.

Proof: $\int_{-\infty}^{\infty} |x|^s dF(x) = \int_{-1}^{1} |x|^s dF(x) + \int_{|x|>1} |x|^s dF(x)$

If $s < r$, then $|x|^s < |x|^r$ for $|x| > 1$.

$$\therefore \int_{-\infty}^{\infty} |x|^s dF(x) \le \int_{-1}^{1} |x|^s dF(x) + \int_{|x|>1} |x|^r dF(x).$$

$$\le \int_{-1}^{1} dF(x) + \int_{|x|>1} |x|^r dF(x),$$

Since for $-1 < x < 1$, $|x|^s < 1$:

$$\therefore \int_{-\infty}^{\infty} |x|^s dF(x) \le 1 + E|X|^r < \infty$$

[$\because E(X^r)$ exists]

$\Rightarrow E(X^s)$ exists $\forall\ 1 \le s \le r$.

Ex:6.3 In four tosses of a coin, let X be the number of heads. Tabulate the 16 possible outcomes with the corresponding values of X. By simple counting derive the probability distribution of X and hence calculate the expected value of X.

Soln: Let H represent a head, T a tail and X the random variable denoting the numbers of heads.

S.NO	outcomes	No of heads (x)	S.NO	outcomes	No of heads (x)
1	HHHH	4	9	HTHT	2
2	HHHT	3	10	THTH	2
3	HHTH	3	11	THHT	2
4	HTHH	3	12	HTTT	1
5	THHH	3	13	THTT	1
6	HHTT	2	14	TTHT	1
7	HTTH	2	15	TTTH	1
8	TTHH	2	16	TTTT	0

The random variable X takes the values 0, 1, 2, 3 and 4 . since from the above table, we find that the number of cases favourable to the coming of 0, 1, 2, 3 and 4 heads are 1, 4, 6, 4 and 1 respectively, we have

$P(X=0) = 1/16$, $P(X=1) = 4/16 = 1/4$, $P(X=2) = \frac{6}{16} = 3/8$, $P(X=3) = 4/16 = 1/4$, $P(X=4) = 1/16$.

The probability distribution of x can be summarized as follows.

x :	0	1	2	3	4
$P(x)$:	1/16	1/4	3/8	1/4	1/16

$$E(X) = \sum_{x=0}^{4} x\,p(x) = 1\cdot 1/4 + 2\cdot 3/8 + 3\cdot\frac{1}{4} + 4\cdot 1/16$$

$$= \frac{1}{4} + \frac{3}{4} + 3/4 + 1/4 = 2 .$$

Ex:6.4 An urn contains 7 white and 3 red balls. Two balls are drawn together at random, from this urn. compute the probability that neither of them is white. find also the probability of getting one white and one red ball. Hence compute the expected no. of white balls drawn.

Soln: let X denote the number of white balls drawn. The probability distribution of X is obtained as follows.

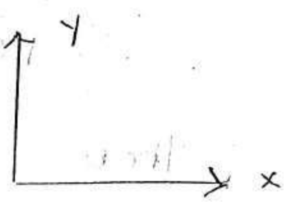
UNIT - 4 :-

CORRELATION & REGRESSION :-

1. Correlation:-

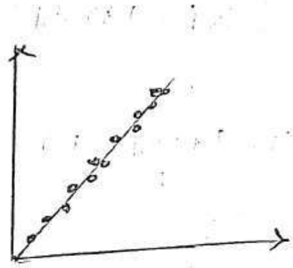
Assumptions:-

* Data is Continuous
* Two Variables are measured on each respondent
* y - independent Variable
* x - dependent Variable
* n - Sample



2. possible correlation outcomes:-

large +ve Correlation

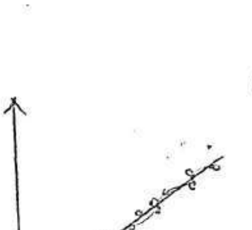


* All points form a Straight line
* Value y increases, x Value also increases.
* perfect positive correlation between 2 Variables
* Value of Correlation co-eff (r) = +1.00

3. Medium Positive Correlation:-

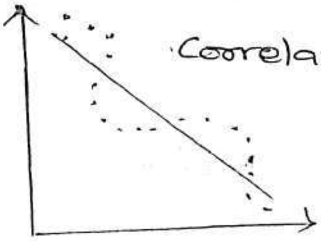
* points do not form a Stg line
* Clustered close together
* Value of y wont increase with x Value.
* Relationship is +ve
* Larger the Value, stronger the Correlation
* r=0 and 1.

medium Positive Correlation.



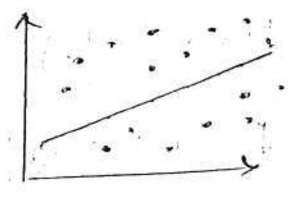
Medium negative Correlation

3. Small -ve Correlation



* points form a -ve slope
* points present on either side of the line
* Value of y decreases with an increase in x Value
* $r=0$ and -1
* More -ve the Value stronger the -ve Correlation

No Correlation



* points are scattered
* No definite line can be formed
* No relationship between x and y
* $r=0$

Summing up

* Correlation
* In bi-Variate distribution, having Value x and y
* If one Variable affects other - Correlated
* +ve Correlation
* -ve Correlation.

Karl- pearson co-efficient of correlation.

Correlation co-eff between two random Variables X and Y, usually denoted by $r(X,Y)$ is a numerical measure of linear relationship between them and is defined as,

$$r(x,y) = \frac{\text{cov}(x,y)}{\sigma_x , \sigma_y}$$

where $\text{cov}(x,y) = \frac{1}{n}\sum xy - \bar{x}\bar{y}$

$$\sigma_x = \sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2}$$

(n is the number of items in the given data)

NOTE!- Correlation co-eff cannot exceed unity numerically (ie) $-1 \le r_{xy} \le 1$

Example 1!-

Calculate the correlation co-eff for the following heights [in inches] of fathers x and their sons y.

x	65	66	67	67	68	69	70	72
y	67	68	65	68	72	72	69	71

Soln!- Method 1 :-

x	y	xy	x^2	y^2
65	67	4355	4225	4489
66	68	4488	4356	4624
67	65	4355	4489	4225
67	68	4556	4489	4624
68	72	4896	4624	5184

69	72	4968	4761	5184
70	69	4830	4900	4761
72	71	5112	5184	5041
544	552	37560	37028	38132

Now $\bar{x} = \dfrac{544}{8} = 68$

$\bar{y} = \dfrac{552}{8} = 69$

$\sigma_x = \sqrt{\dfrac{1}{n}\sum x^2 - \bar{x}^2} = \sqrt{\dfrac{37028}{8} - 4624}$

$= 2.121$

$\sigma_y = \sqrt{\dfrac{1}{n}\sum y^2 - \bar{y}^2} = \sqrt{\dfrac{38132}{8} - 4762}$

$= 2.345$

$\therefore r(x,y) = \dfrac{\frac{1}{n}\sum xy - \bar{x}\,\bar{y}}{\sigma_x \cdot \sigma_y} = \dfrac{\frac{1}{8} \times 37560 - 4692}{2.121 \times 2.345}$

NOTE:- Correlation co-eff is independent of change of origin and scale.

(ie) $r(x,y) = r(u,v)$

where $u = \dfrac{x-a}{h}$; $v = \dfrac{y-b}{k}$

where a and b are some arbitrary constants usually the mid values of the given data x and y respectively.

Soln:-

x_i	y_i	z_i	$d_1 = x_i - y_i$	$d_2 = y_i - z_i$	$d_3 = x_i -$	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	1	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
						200	214	60

- The rank correlation between x and y is

$$r_1(x,y) = 1 - \frac{6\sum d_1^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 200}{10(100-1)} = -0.212$$

The rank correlation between x and z is

$$r_3(x,z) = 1 - \frac{6\sum d_3^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 60}{10(10^2-1)} = 0.636$$

Since the rank correlation between x and z is maximum and also positive, we conclude that the pair of judges x and z has the nearest approach to common likings in music.

Repeated ranks:-

If any two or more individuals are equal in any classification with respect to characteristic A or B, or if there is more than one item with the same value in the series then Spearman's formula for calculating the rank correlation co-efficients break down. In this case common ranks are given to the repeated ranks. This common rank is the average of the ranks which these items would have assumed if they are slightly different from each other and the next item will get the rank next to the ranks already assumed. As a result of this, following adjustment or correlation is made in the correlation formula

The correlation formula, we add the factor $\frac{m(m^2-1)}{12}$ to Σd^2 where m is the number of items an item is repeated. This correlation factor is to be added for each repeated value.

Example 1:- Obtain the rank correlation Co-eff for the following data.

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

Soln:-

X	Y	Rank X (x_i)	Rank Y (y_i)	$d_i = x_i - y_i$	d^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
					72

In X Series 75 repeated twice which are in the positions 2nd and 3rd ranks. Therefore Common ranks 2.5 [which is the average of 2 & 3] is to be given for each 75. Also in X Series 64 is repeated thrice which are in the position 5th, 6th and 7th ranks.

Therefore the common ranks 6 [which is the average of 5, 6 and 7] is to be given for each 64.

Similarly in Y series is repeated twice which are in the positions - 3rd and 4th ranks. Therefore common ranks 3.5 [which is the average of 3 and 4] is to be given for each 68.

Correction factors:-

In X series 75 repeated twice

$$\therefore C.F. = \frac{2(2^2-1)}{12} = \frac{1}{2}$$

In X series 64 repeated thrice

$$\therefore C.F = \frac{3(3^2-1)}{12} = \frac{24}{12} = 2$$

$$\therefore \text{Rank Correlation } r = 1 - \frac{6\left(\Sigma d^2 + \frac{1}{2} + 2 + \frac{1}{2}\right)}{10(10^2-1)}$$

$$= 1 - \frac{6[72+0.5+2+0.5]}{10 \times 99}$$

$$= 1 - \frac{450}{990}$$

$$= 0.5454.$$

Example 2:- A sample of 12 fathers and their eldest sons have the following data about their heights in inches.

Fathers	65	63	67	64	68	62	70	66	68	67	69	71
Sons.	68	66	68	65	69	66	68	65	71	67	68	70

Calculate the rank correlation Co-eff.

Soln:-

Fathers x_i	Sons y_i	Rank of x_i	Rank of y_i	$d_i = x_i - y_i$	d_i^2
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	-1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	-3.5	12.25
66	65.	8	11.5	-3.5	12.25
68	71	4.5	1	+3.5	12.25
67	67	6.5	8	-1.5	2.25
69	68	3	5.5	-2.5	6.25
71	70	1	2	-1	1
					72.5

Correlation factors:-

In X Series 68 is repeated twice

$$\therefore \ C.F. = \frac{2(2^2-1)}{12} = \frac{1}{2}$$

In X series 67 is repeated twice

$$\therefore \ C.F. = \frac{2(2^2-1)}{12} = \frac{1}{2}$$

In Y Series 68 is repeated four times

$\therefore$ C.F. $= \dfrac{4(4^2-1)}{12} = 5$

In Y Series 66 is repeated twice

$\therefore$ C.F. $= \dfrac{2(2^2-1)}{12} = \dfrac{1}{2}$.

In Y Series 65 is repeated twice

$\therefore$ C.F. $= \dfrac{2(2^2-1)}{12} = \dfrac{1}{2}$

$\therefore$ Rank Correlation co-eff

$$r(x, y) = 1 - \frac{6[72.5 + 0.5 + 0.5 + 5 + 0.5 + 0.5]}{12(144-1)}$$

$= 1 - 0.277$

$= 0.722$

Regression:-

Regression is the mathematical measure of the average relationship between two or more ~~related~~ variables in terms of the original limits of the data.

Lines of Regression:-

- If the variable in a bivariate distributio are related we will find that the point in the scattered diagram will cluster around some curve. Curve of regression.

- If curve is straight line : Line of regression
- Curvilinear.

The line of regression of y on x is given by

$$y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

where r is the correlation co-efficient σ_y and σ_x are standard deviation.

The line of regression of x on y is given by

$$x - \bar{x} = r \cdot \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

Note:- Both line of regression passes through $(\bar{x}, \bar{y})$

Angle Between two lines of Regression.

If the equations of lines of regression of y on x and x on y are

$$y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

and

$$x - \bar{x} = r \cdot \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

The angle θ between the two lines of regression is given by

$$\tan\theta = \frac{1-r^2}{r}\left(\frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\right)$$

Note 1:- If $r=0$, we get $\tan\theta = \infty$

$$\Rightarrow \theta = \pi/2$$

$\therefore$ when $r=0$ the lines of regression are $\perp^r$ other

Note 2:- If $r=\pm 1$, then

$$\tan\theta = 0$$
$$\Rightarrow \theta = 0 \text{ or } \pi$$

when $r=\pm 1$, the two regression lines are $\parallel$ to each other (or) coincide.

Note 3:- when $r=0$, the two variables x and y are uncorrelated.

Note 4:- When $r=\pm 1$, the correlation between x and y is said to be perfect.

Regression Co-efficients:-

Regression co-eff of y on x

$$r\frac{\sigma_y}{\sigma_x} = b_{yx} \quad \rightarrow (1)$$

Regression Co-eff of x on y

$$r\frac{\sigma_x}{\sigma_y} = b_{xy} \quad \rightarrow (2)$$

from ① and ② we get

$$r\frac{\sigma_y}{\sigma_x} \; r\frac{\sigma_x}{\sigma_y} = b_{yx} \times b_{xy}$$

$$r^2 = b_{xy} \times b_{yx}$$

$$\therefore \quad r = \pm\sqrt{b_{xy} \times b_{yx}}$$

Correlation coefficient $r = \pm\sqrt{b_{xy} \times b_{yx}}$

The regression co-eff b_{yx} and b_{xy} can be easily obtained by using the following formula.

NOTE 1:- $$b_{yx} = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}$$

$$b_{xy} = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (y-\bar{y})^2}$$

NOTE 2:- Regression co-efficients are independent of change of origin but not of scale.

$$b_{yx} = b_{vu} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2}$$

$$b_{xy} = b_{uv} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma v^2 - (\Sigma v)^2}$$

where $u = x-a$, $v = y-b$,

Example 1:- From the following data, find.

(i) The two regression equations

(ii) The Co-eff of Correlation between the marks in Economics and Statistics.

(iii) The most likely marks in Statistics when marks in Economics are 30.

Marks in Economics.	25	28	35	32	31	36	29	38	34	32
Marks in Statistics.	43	46	49	41	36	32	31	30	33	39

Soln:-

x	y	$x-\bar{x}$ $= x-32$	$y-\bar{y}$ $= y-38$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
320	380	0	0	140	398	-93

Here $\bar{x} = \frac{\Sigma x}{n}$ and $\bar{y} = \frac{\Sigma y}{n}$

$= \frac{320}{10} = 32$ $\quad = \frac{380}{10} = 38$

Co-eff of regression of y on x is

$$b_{yx} = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (x-\bar{x})^2}$$

$$= \frac{-93}{140}$$

$$= -0.6643$$

Co-eff of regression of x on y is

$$b_{xy} = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\Sigma (y-\bar{y})^2}$$

$$= \frac{-93}{398}$$

$$= -0.2337$$

Equ of the line of regression of x on y is

$$x - \bar{x} = b_{xy} (y-\bar{y})$$

$$x - 32 = -0.2337 (y-38)$$

$$= -0.2337y + 0.2337 \times 38$$

$$x = -0.2337y + 40.8806$$

Equ of the line of regression of y on x is

$$y-\bar{y} = b_{yx} (x-\bar{x})$$

$$y - 3.8 = -0.6643(x - 32)$$

$$y = -0.6643x + 3.8 + 0.6643 \times 32$$

$$= -0.6643x + 59.2576.$$

Co-eff of correlation.

$$r^2 = b_{yx} \times b_{xy}$$

$$= -0.6643 \times (-0.2337)$$

$$= 0.1552$$

$$r = \pm\sqrt{0.1552}$$

$$= \pm 0.394$$

Now we have to find the most likely marks in Statistics (y) when marks in Economics (x) are 30. we use the line of regression of y on x.

$$y = -0.6643x + 59.2575$$

Put $x = 30$, we get $y = -0.6643 \times 30 + 59.2536$

$$= 39.3236$$

$$= 39$$

<u>Example 2:</u>- Height of fathers and sons are given in centimeters,

x: Height of father	150	152	155	157	160	161	164	166
y: Height of son	154	156	158	159	160	162	161	164

UNIT - 5

The important discrete distributions of a random variable 'x' are

1. Binomial distribution
2. Poisson distribution.

1. Binomial distribution

Let us consider 'n' independent trails. If the Successes (S) and failure (F) are recorded successively as the trails are repeated we get a result of the type

SSFFS..... FS

Let 'x' be number of success and hence we have $(n-x)$ number of failures.

$$P(SSFFS.....FS) = P(S)\,P(S)\,P(F)\,P(F)\,P(S)\,....\,P(F)\,P(S)$$
$$= ppqqp....qp$$
$$= \underbrace{ppp.......p}_{x\text{-factors}}\ \underbrace{qqq.......q}_{(n-x)\text{ factors}}$$
$$= p^x \cdot q^{n-x}$$

But 'x' successes in 'n' trails can occur in nC_x ways

∴ The probability of 'x' successes in 'n' trails is given by $nC_x \cdot p^x q^{n-x}$

(i.e) $p(x \text{ successes}) = nC_x \cdot p^x \cdot q^{n-x}$

$$P(x) = nC_x \cdot p^x\, q^{n-x}$$

NOTE: 1. $P(x) = nC_x \cdot p^x \cdot q^{n-x}$

Here $nC_x\, p^x\, q^{n-x}$ is the $(x+1)^{th}$ term in the expansion of $(q+p)^n$

$[\because (q+p)^n = q^n + nC_1 q^{n-1} p^1 + ... + nC_x q^{n-x} p^x + ...$ (A) which is a binomial Series and hence the distribution is called binomial distribution.]

2. $P(0 \text{ Success}) = nC_0\, p^0 q^{n-0} = q^n$

$P(1 \text{ Success}) = nC_1\, p^1 q^{n-1}$

$P(2 \text{ Success}) = nC_2\, p^2 q^{n-2}$ and so on.

These terms are the Successive terms in the above expansion (A)

3. Let an experiment constitutes n trails. If the experiment repeated N times, the frequency function of the binomial distribution is given by

$$f(x) = N\, p(x)$$
$$= N \cdot nC_x\, p^x q^{n-x}\ ; \quad x = 0, 1, 2, \dots n$$

4. In the problems we study in this section, we Shall always make the following assumption.

(i) There are only two possible outcomes for each trail [Success or trail]

(ii) The probability of a Success is the same for each trail.

(iii) There are 'n' trails, where n is a constant.

(iv) The 'n' trails are independent.

Mean and Variance of the Binomial Distribution

We know that for discrete probability distribution mean is ~~a~~ given by

$$\mu_1 = E(X) \qquad [\because E(X) = x]$$
$$= \sum_{x=0}^{n} x\, p(x) \qquad [p(x) \text{ if p.d.f.}]$$

$$= \sum_{x=0}^{n} x \; nC_x \, p^x q^{n-x}$$

$$= 0q^n + 1 \cdot nC_1 \, q^{n-1} p + 2 \, nC_2 \, q^{n-2} p^2 + \dots + np^n$$

$$= np \left[q^{n-1} + {}^{n-1}C_1 \, q^{n-2} p + (n-1)C_2 \, q^{n-3} p^2 + \dots + p^{n-1} \right]$$

$$= np(q+p)^{n-1}$$

$$= np \qquad [\because p+q=1]$$

Hence the mean of the binomial distribution is $\boxed{\bar{x} = np}$

Now
$$\mu_2' = \sum_{x=0}^{n} x^2 p(x) = \sum_{x=0}^{n} x^2 \, nC_x \, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} (x + x(x-1)) \, nC_x \, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x \, nC_x \, p^x q^{n-x} + \sum_{x=0}^{n} x(x-1) \, nC_x \, p^x q^{n-x}$$

$$= np + \sum_{x=0}^{n} x(x-1) \frac{n(n-1)}{x(x-1)} (n-2)C_{x-2} \, p^x q^{n-x}$$

$$= np + \sum_{x=0}^{n} n(n-1) \, {}^{n-2}C_{x-2} \, p^2 \cdot p^{x-2} q^{n-x}$$

$$= np + (n-1)n \cdot p^2 \cdot \sum_{x=0}^{n} (n-2)C_{x-2} \, p^{x-2} q^{n-x}$$

$$= np + n(n-1)p^2 \cdot (q+p)^{n-2}$$

$$= np[1 + (n-1)p] = np[1 - p + np] \qquad [\because p+q=1]$$

$$= np[q + np] \qquad [\because 1-p=q]$$

$\therefore$ Variance $(\mu_2) = \mu_2' - \mu_1'^2$

$$= np(q+np) - n^2 q^2$$

$$= npq$$

Standard deviation $= \sqrt{npq}$

Moment generating function (m.g.f.) of a binomial distribution about origin.

We know that the moment generating function of a random variable x about origin whose probability function f(x) is given by

$$M_x(t) = \sum_{x=0}^{n} e^{tx} f(x) \qquad [f(x) - \text{probability fn}]$$

let 'x' be a random variable which follows binomial distribution.

Then its m.g.f. about origin is given by

$$M_x(t) = E[e^{tx}] = \sum_{x=0}^{n} e^{tx} \cdot f(x)$$

$$= \sum_{x=0}^{n} e^{tx}\, nC_x\, p^x q^{n-x} \qquad [\because f(x) = nC_x\, p^x q^{n-x}]$$

$$= \sum_{x=0}^{n} (e^t)^x p^x\, nC_x\, q^{n-x} = \sum_{x=0}^{n} (pe^t)^x\, nC_x\, q^{n-x}$$

$$= \sum_{x=0}^{n} nC_x\, (pe^t)^x q^{n-x}$$

$$= q^n + nC_1\, q^{n-1} (pe^t)' + nC_2\, q^{n-2} (pe^t)^2 + \dots$$

$$= (q + pe^t)^n$$

Moment generating function (m.g.f.) of binomial distribution about mean (np)

W.K.T. the m.g.f. of random variable x about any point 'a' is $M_x(t)$ (about $x = a$) $= E[e^{t(x-a)}]$

Here a is the mean of the binomial distribution

(i.e) $M_x(t)$ (about $x = np$)

$$= E[e^{t(x-np)}] = E[e^{tx} e^{-tnp}]$$

$$= e^{-tnp}\, E[e^{tx}]$$

$$= e^{-tnp}.\ \text{m.g.f. of } x \text{ about origin}$$

$$= e^{-tnp}(q+pe^{t})^{n}$$

$$= (e^{-tp})^{n}(q+pe^{t})^{n}$$

$$= \{(q+pe^{t})^{n}(e^{-tp})^{n}\}$$

$$= \{qe^{-tp}+p.e^{-tp}e^{t}\}^{n}$$

$$= \{qe^{-tp}+pe^{tp}\}^{n}$$

$$= \left\{q\left(1-pt+\frac{p^2t^2}{2!}+-\frac{p^3t^3}{3!}+\dots\right)+p\left(1+qt+\frac{q^2t^2}{2!}+\frac{q^3t^3}{3!}+\dots\right)\right\}$$

$$\left[\because e^{x}=1+\frac{x}{1!}+\frac{x^2}{2!}+\dots\right]$$

$$= \left\{\left(q-qpt+\frac{qp^2t^2}{2!}-\frac{qp^3t^3}{3!}+\dots\right)+\left(p+pqt+\frac{pq^2t^2}{2!}+\dots\right)\right\}^{n}$$

$$= \left\{p+q-qpt+pqt+\frac{pqt^2(p+q)}{2!}+\frac{pq(q^2-p^2)t^3}{3!}+\dots\right\}^{n}$$

$$= \left\{1+\frac{t^2}{2!}pq+\frac{t^3}{3!}pq(q-p)+\dots\right\}^{n}$$

$$= \left\{1+\left[\frac{t^2}{2!}pq+\frac{t^3}{3!}pq(q-p)+\dots\right]\right\}^{n}$$

$$= 1+n\left\{\frac{t^2}{2!}pq+\frac{t^3}{3!}pq(q-p)+\dots\right\}$$

$$+\frac{n(n-1)}{2!}\left\{\frac{t^2}{2!}pq+\frac{t^3}{3!}pq(q-p)+\dots\right\}^2+\dots$$

$$\left[\text{using } (1+x)^n=1+nx+\frac{n(n-1)}{2!}x^2+\dots\right]$$

Now μ_2 = coefficient of $\frac{t^2}{2!}$ $= npq$

(i.e) Variance $= npq$

μ_3 = coefficient of $\frac{t^3}{3!}$ $= npq(q-p)$ and so on

Recurrence Relation for the moments of Binomial distribution

We know that

$$\mu_r = E\left[X - E(X)\right]^r$$

$$= \sum_{x=0}^{n} (x-np)^r \binom{n}{x} p^x q^{n-x}$$

Differentiating w.r.t. p,

$$\frac{d\mu_r}{dp} = \sum_{x=0}^{n} \binom{n}{x} \Big[p^x q^{n-x}\, r\, (x-np)^{r-1}(-n)$$

$$+ (x-np)^r q^{n-x}\, x\, p^{x-1} + (x-p)^r p^x (n-x) q^{n-x-1}(-1) \Big]$$

$$= \sum_{x=0}^{n} \binom{n}{x} \Big[-nr(x-np)^{r-1} p^x q^{n-x} \qquad [\because q = 1-p]$$

$$+ (x-np)^r \left[x p^{x-1} q^{n-x} - (n-x) p^x q^{n-x-1} \right] \Big]$$

$$= -nr \sum_{x=0}^{n} \binom{n}{x} (x-np)^{r-1} p^x q^{n-x}$$

$$+ \sum_{x=0}^{n} \binom{n}{x} (x-np)^r p^x q^{n-x} \left\{ \frac{x}{p} - \frac{n-x}{q} \right\}$$

$$= -nr \sum_{x=0}^{n} (x-np)^{r-1} p(x) + \sum_{x=0}^{n} (x-np)^r p(x) \left(\frac{x-np}{pq} \right)$$

$$\left[\because P(x) = \binom{n}{x} p^x q^{n-x} \right]$$

$$= -nr \sum_{x=0}^{n} (x-np)^{r-1} p(x) + \frac{1}{pq} \sum_{x=0}^{n} (x-np)^{r+1} p(x)$$

$$\therefore \frac{d\mu_r}{dp} = -nr\,\mu_{r-1} + \frac{1}{pq}\mu_{r+1}$$

(i.e) $\mu_{r+1} = pq\left[nr\,\mu_{r-1} + \frac{d\mu_r}{dp} \right]$ —— ①

This gives the recurrence relation for the moment of Binomial distribution

Put $n=1$ in (1), we get

$$\mu_2 = pq\left[n\mu_0 + \frac{d\mu_r}{dp}\right] = npq \quad [\because \mu_1 = 0,\ \mu_0 = 1]$$

Put $n=2$ in (1), we get

$$\mu_3 = pq\left[2n\mu_1 + \frac{d\mu_2}{dp}\right]$$

$$= pq\frac{d\mu_2}{dp}$$

$$= pq\frac{d[npq]}{dp}$$

$$= \frac{pq\, d[np(1-p)]}{dp}$$

$$= npq[p(-1) + (1-p)]$$

$$= npq(-p+q)$$

$$= npq(q-p)$$

Show that the r^{th} moment μ_r' about the origin of the binomial distribution of degree 'n' is given by

Proof:

$$\mu_r' = \left(p\frac{\partial}{\partial p}\right)^r (q+p)^n$$

we shall prove this result by mathematical induction

We know that

$$(q+p)^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}$$

$$\frac{\partial}{\partial p}[(q+p)^n] = \sum_{x=0}^{n} \binom{n}{x} \frac{\partial}{\partial p}(p^x) q^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} q^{n-x} . x p^{x-1}$$

$$\therefore p.\frac{\partial}{\partial p}(q+p)^n = p\sum_{x=0}^{n} \binom{n}{x} q^{n-x} . x p^{x-1}$$

$$= \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} . x$$

$$= \mu_1$$

Thus the result is true for $r = 1$. Now let us assume that the result is true for $r = k$.

(i.e) $\left(p.\frac{\partial}{\partial p}\right)^k (q+p)^n = \mu_k' = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} x^k$

Differentiating partially w.r.t. 'p' we get

$$\frac{\partial}{\partial p}\left[\left(p\frac{\partial}{\partial p}\right)^k (q+p)^n\right] = \sum_{x=0}^{n} \binom{n}{x} \frac{\partial}{\partial p}(p^x) q^{n-x} x^k$$

$$= \sum_{x=0}^{n} \binom{n}{x} x\, p^{x-1} q^{n-x} x^k$$

$$= \sum_{x=0}^{n} \binom{n}{x} x^{k+1} p^{x-1} q^{n-x}$$

$$p\frac{\partial}{\partial p}\left[\left(p\frac{\partial}{\partial p}\right)^k (q+p)^n\right] = \sum_{x=0}^{n} \binom{n}{x} x^{k+1} p^x q^{n-x}$$

$$= \mu_{k+1}'$$

Thus the result is true for $r = k+1$. (i.e) if the result is true for $r = k$, it is also true for $r = k+1$. Hence by induction method, it is true for all positive integral values of k.

(i.e) $\mu_r' = \left(p.\frac{\partial}{\partial p}\right)^r (q+p)^n$

Ex: 01 The mean and variance of a binomial distribution are 4 and $4/3$. Find $P(x \geq 1)$

Sol:- W.K.T Mean of Binomial distribution $= np$ and variance is npq.

Given $np = 4$

$nqp = 4/3$

$$\frac{①}{②} \Rightarrow \frac{np}{npq} = \cancel{4} \times \frac{3}{\cancel{4}} = 3$$

$$\frac{1}{q} = 3 \quad \therefore \boxed{q = 1/3}$$

But we know $p + q = 1$

Then $\Rightarrow p = 1 - q \Rightarrow p = 1 - 1/3 \Rightarrow P = \frac{3-1}{3}$

$$\boxed{p = 2/3}$$

Mean $= np = 4$

$$\Rightarrow n(2/3) = 4$$

$$n = \overset{2}{\cancel{4}}(3/\cancel{2})$$

$$\boxed{n = 6}$$

Now

$$P(x \geq 1) = 1 - p(0)$$

$$= 1 - nC_0\, p^0 q^6 - 0 \quad [P(x) = nC_x p^x q^{n-x}]$$

$$= 1 - q^6 = 1 - (1/3)^6$$

$$P(x \geq 1) = 0.998$$

Ex: 02 10 coins are thrown simultaneously Find the probability of ability atleast 7 heads

Sol: $p = 1/2$; $q = 1/2$; $n = 10$.

P(getting x successess) $= p(x) = nC_x\, p^x q^{n-x}$

P(getting atleast 7 heads)

$= p(x \geq 7)$

$= p(7) + p(8) + p(9) + p(10)$

$= 10C_7 (1/2)^7 (1/2)^3 + 10C_8 (1/2)^8 (1/2)^2$

$+ 10C_9 (1/2)^9 (1/2) + 10C_{10} (1/2)^{10}$.

$= \frac{1}{2^{10}} [10C_7 + 10C_8 + 10C_9 + 10C_{10}]$

$= \frac{1}{2^{10}} [120 + 45 + 10 + 1]$

$= \frac{176}{2^{10}} = \frac{176}{1024} = 0.171875$.

Poisson Distribution

Poisson distribution is a limiting case of binomial distribution under the following assumptions

(i) The number of trails 'n' should be indefinitely large. (i.e) $n \to \infty$

(ii) The probability of successes 'p' for each trail is indefinetly small.

(iii) $np = \lambda$, should be finite where λ is a constant.

Now know that the binomial distribution is

$$P(X=x) = nC_x \, p^x q^{n-x}$$

$$= \frac{n!}{(n-x)!\, x!} p^x (1-p)^{n-x}$$

$$= \frac{1 \cdot 2 \cdot 3 \cdots (n-x)(n-x+1)(n-x+2) \cdots}{1 \cdot 2 \cdot 3 \cdots (n-x) \cdot x!} \cdot \left(\frac{p}{1-p}\right)^x (1-p)^n$$

$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!} \left(\frac{\frac{\lambda}{n}}{1-\frac{\lambda}{n}}\right)^x \left(1-\frac{\lambda}{n}\right)^n$$

$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!} \frac{\lambda^x}{n^x} \frac{1}{\left(1-\frac{\lambda}{n}\right)^x} \left(1-\frac{\lambda}{n}\right)^n$$

$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!} \cdot \frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$p(X=x) = \frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left\{1-\left(\frac{x-1}{n}\right)\right\}}{x!}$$

$$\cdot\left(1-\frac{\lambda}{n}\right)^{n-x}$$

We know that $\underset{n\to\infty}{Lt}\left(1-\frac{\lambda}{n}\right)^{n-x} = e^{-\lambda}$

and $\underset{n\to\infty}{Lt}\left(1-\frac{1}{n}\right) = \underset{n\to\infty}{Lt}\left(1-\frac{2}{n}\right) = 1$

$\therefore$ When $n\to\infty$, the R.H.S of (1) gives

$$\frac{\lambda^x \cdot e^{-\lambda}}{x!} \longrightarrow (2)$$

Substituting (2) in (1), we get

$$\therefore p(X=x) = \frac{\lambda^x}{x!} e^{-\lambda} \cdot x = 0, 1, 2, \ldots \infty$$

Hence the probability function of a random variable 'X' which follows poisson distribution is given by.

$$p(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} ; \quad x = 0, 1, 2 \ldots, \infty$$

$$= 0, \text{ otherwise.}$$

Moment of Generating Function of the Poisson Distribution

We know that the m.g.f of random variable 'x' is given by

$$M_x(t) = \sum_{x=0}^{n} e^{tx} \cdot f(x) \quad [f(x) \text{ is P.d.f of poisson Distr.}]$$

$$= \sum_{x=0}^{n} e^{t}\left(\frac{e^{-\lambda} \cdot \lambda^{x}}{x!}\right)$$

$$= \sum_{x=0}^{n} \frac{e^{-\lambda}(\lambda e^{t})^{x}}{x!} = e^{-\lambda} \sum_{x=0}^{n} \left(\frac{(\lambda e^{t})^{x}}{x!}\right)$$

$$= e^{-\lambda}\left\{1 + \lambda e^{t} + \left(\frac{\lambda e^{t}}{2!}\right)^{2} + \cdots\right\}$$

$$\left[\because e^{x} = 1 + \frac{x}{1!} + \frac{x^{2}}{2!} + \cdots\right]$$

$$= e^{-\lambda} \cdot e^{\lambda e^{t}} = e^{\lambda(e^{t}-1)}$$

Hence $M_x(t) = e^{\lambda(e^{t}-1)}$

$\therefore$ Moment generating function of the random variable 'x' is

$$M_x(t) = e^{\lambda(e^{t}-1)}$$

<u>To find mean and variance of the poisson distribution</u>

We know that, for discrete probability and distribution, mean is given by

$$\mu_1' = E(x)$$

$$= \sum_{x=0}^{\infty} x \cdot P(x) \quad [P(x) - \text{p.d.f}]$$

$$= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^{x}}{x!} - \sum_{x=0}^{n} \frac{x e^{-\lambda} \cdot \lambda \cdot \lambda^{x-1}}{x!}$$

$M_x(t) = \sum_{x=0}^{n} e^{tx} f(x)$ [$f(x)$ is p.d.f of Poisson distribution]

$= \sum_{x=0}^{n} e^{t}\left(\frac{e^{-\lambda}.\lambda^{x}}{x}\right)$

$= \sum_{x=0}^{n} \frac{e^{-\lambda}(\lambda e^{t})^{x}}{x!} = e^{-\lambda}\sum_{x=0}^{n} \left(\frac{(\lambda e^{t})x}{x!}\right)$

$= e^{-\lambda}\left\{1+\lambda e^{t}+\left(\frac{\lambda e^{t}}{2!}\right)^{2}+\cdots\right\}$

$\left[\because e^{x} - 1+\frac{x}{1!}+\frac{x^{2}}{2!}+\cdots\right]$

$= e^{-\lambda}.e^{\lambda e^{t}} = e^{\lambda(e^{t}-1)}$

Hence $M_x(t) = e^{\lambda(e^{t}-1)}$

$\therefore$ Moment of generating function of the random variable 'x' is

$M_x(t) = e^{\lambda(e^{t}-1)}$

<u>To find Mean and variance of the poisson distribution</u>

we know that for discrete Probability and distribution Mean is given by

$\mu_1' = E(x)$

$= \sum_{x=0}^{\infty} x.\mu(x)$ [$p(x)$ - p.d.f]

$= \sum_{x=0}^{\infty} x.\frac{e^{-\lambda}\lambda^{x}}{x!} - \sum_{x=0}^{n} \frac{xe^{-\lambda}.\lambda.\lambda^{x-1}}{x!}$

Ex: 01 If x is a poission variate $P(X=2) = 9P(X=4) + 90P(X=6)$.

Find (i) Mean of x (ii) variance of x.

Sol: $P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$ $x = 0, 1, 2 \ldots$

Given $P(X=2) = 9(X=4) + 90P(X=6)$

$$\frac{e^{-\lambda}\lambda^{2}}{2!} = 9\frac{e^{-\lambda}\lambda^{4}}{4!} + 90\frac{e^{-\lambda}\lambda^{6}}{6!}$$

$$\cancel{\tfrac{1}{2}} = e^{-\lambda}\lambda^{2}\left(\frac{9\lambda^{2}}{4!} + \frac{90\lambda^{4}}{6!}\right)$$

$$\frac{1}{2} = \frac{9\lambda^{2}}{4!} + \frac{90\lambda^{4}}{6!}$$

$$\frac{1}{2} = \frac{3\lambda^{2}}{8} + \frac{\lambda^{4}}{8}$$

$$\frac{3\lambda^{2}}{8} + \frac{\lambda^{4}}{8} - \frac{1}{2} = 0$$

$$\lambda^{4} + 3\lambda^{2} - 4 = 0$$

$$\lambda^{2} = \frac{-3 \pm \sqrt{9+16}}{2}$$

$$= \frac{3 \pm 5}{2}$$

$$\lambda^{2} = 1 \text{ or } \lambda^{2} = -1$$

$$\lambda = \pm 1$$

Mean $= \lambda = 1$

variance $= \lambda = 1$

standard derivation $= 1$.

Ex: 02 If x is a poission variate such that $p(x=1) = 3/10$ and $p(x=2) = 1/5$ find $p(x=0)$ and $p(x=3)$

sol: $p(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

$p(x=1) = e^{-\lambda}\lambda = \frac{3}{10}$ — (1)

$p(x=2) = \frac{e^{-\lambda}\cdot\lambda^{2}}{2!} = 1/5$ — (2)

(1) $\Rightarrow$ $e^{-\lambda}\lambda = \frac{3}{10}$ — (3)

(2) $\Rightarrow$ $e^{-\lambda}\lambda^{2} = \frac{2}{5}$ — (4)

$\frac{(3)}{(4)} \Rightarrow \frac{1}{\lambda} = \frac{3}{10} \times \frac{5}{2} = \frac{3}{4}$

$\boxed{\lambda = 4/3}$

$p(x=0) = \frac{e^{-\lambda}\lambda^{0}}{0!} = e^{-4/3}$

$p(x=3) = \frac{e^{-\lambda}\lambda^{3}}{3!} = \frac{e^{-4/3}(4/3)^{3}}{3!}$

Fitting of poisson distribution:

Ex:01 Fit a poisson distribution to the following data and calculate the theoretical frequencies.

Deaths	0	1	2	3	4
Frequency	122	60	15	2	1

Solution.

x	f	fx	Theoretical frequencies
0	122	0	121
1	60	60	61
2	15	30	15
3	2	6	3
4	1	4	0
	$N =$	$\Sigma fx = 100$	200

$$\text{Mean } \bar{x} = \frac{\Sigma fx}{N} = \frac{100}{N} = 0.5$$

Theoretical distribution is given by

$$= N \times p(x)$$

$$= 200 \times \frac{e^{-\lambda} \lambda^{x}}{x!}$$

Hence the theoretical frequencies are given by.

$$f(x) = 200 \frac{e^{-0.5}(0.5)^x}{x!}$$

Putting $x = 0, 1, 2, 3, 4$ in (1) We get

$$\therefore f(0) = \frac{200 \times e^{-0.5}(0.5)^0}{0!} = 121$$

$$f(1) = \frac{200 \times e^{-0.5}(0.5)^1}{1!} = 61$$

$$f(2) = \frac{200 \times e^{-0.5}(0.5)^2}{2!} = 15$$

$$f(3) = \frac{200 \times e^{-0.5}(0.5)^3}{3!} = 3$$

$$f(4) = \frac{200 \times e^{-0.5}(0.5)^4}{4!} = 0.$$

Continuous Distributions.

Normal Distribution

A random variable x is said to follow normal distribution with mean μ and variance σ^2 if its density function is given by the probability law.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}} \rightarrow (1)$$

$$-\infty < x < \infty,\ \sigma > 0,\ -\infty < \mu < \infty$$

The total area bounded by the above curve is 1.

$$\text{Area} = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}}\,dx$$

put $\frac{x-\mu}{\sigma} = z$; $\sigma\,dz = dx$.

when $x = \infty$, $z = \infty$

$x = -\infty$, $z = -\infty$

$$\therefore \text{Area} = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}\,\sigma\,dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2}\,dz$$

$$= \frac{1}{\sqrt{2\pi}}\, 2\int_{0}^{\infty} e^{-z^2/2}\,dz$$

$$= \frac{1}{\sqrt{2\pi}}\, 2\int_{0}^{\infty} e^{-\left(\frac{z^2}{2}\right)}\,dz$$

put $\frac{z}{\sqrt{2}} = u$, $dz = \sqrt{2}\,du$

$$\therefore \text{Area} = \sqrt{2/\pi} \int_{0}^{\infty} e^{-u^2}\cdot\sqrt{2}\,du$$

$$= \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} e^{-u^2}\,du$$

$$= \frac{2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2} = 1 \qquad \left[\because \int_0^\infty e^{-u^2} du = \frac{\sqrt{\pi}}{2}\right]$$

Area bounded by the normal curve is 1.

$$p(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)\, dx,$$

$$= \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

put $z = \frac{x-\mu}{\sigma}$ when $x = x_1,\ z_1 = \frac{x_1-\mu}{\sigma}$

$\sigma dz = dx$ when $x = x_2 .\ z_2 = \frac{x_2-\mu}{\sigma}$

$$\therefore p(z_1 < z < z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{z^2}{2}} dz - \phi(z)$$

The integral $\frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{z^2}{2}} dz$ is called the probability integral. The values of these integrals for different values of z are given in the table. The curve given by.

$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{z^2}{2}} dz$ is called the standard normal curve and it is bell shaped and symmetrical about the line $z = 0$.

Note : 1. Mean, Median and Mode of the normal distribution coincide.

2. $QD : MD : SD = 10 : 12 : 15$

M.G.F of Normal Distribution

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$= \int_{-\infty}^{\infty} e^{tx} . \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

Put.

$$z = \frac{x-\mu}{\sigma} \qquad \Big| \text{ when } x = -\infty,\ z = -\infty$$

$$\sigma\, dz = dx \qquad \Big| \quad x = \infty,\ z = \infty$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma z + \mu)} \cdot e^{-\frac{z^2}{2}}\, dz$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z^2 - 2t\sigma z)}{2}}\, dz$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-\sigma t)^2 + \frac{\sigma^2 t^2}{2}}\, dz$$

$$= e^{\mu t + \frac{t^2\sigma^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-\sigma t)^2}\, dz$$

$$= e^{\mu t + \frac{t^2\sigma^2}{2}}\, \frac{1}{\sqrt{2\pi}} \times 2\int_{0}^{\infty} e^{-u^2/2}\, du$$

$$= e^{\mu t + \frac{t^2\sigma^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}\, 2\, \frac{\sqrt{\pi}}{2} \left[\int_0^{\infty} e^{-u^2/2}\, du = \frac{\sqrt{\pi}}{2}\right]$$

$$\therefore \cancel{M_x(t) = e^{\mu r + \frac{r^2\sigma^2}{2}}}$$

$$\therefore M_x(t) = e^{\mu r + \frac{r^2\sigma^2}{2}}$$

Moments of Normal distribution

Odd order moments about mean

given by,

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} f(x)\, dx$$

$$= \int_{-\infty}^{\infty} (x-\mu)^{2n+1} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2n+1} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

put $z = \frac{x-\mu}{\sigma}$

$\Rightarrow \sigma z = x - \mu \Rightarrow \sigma\, dz = dx$

$$\mu_{2n+1} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \cdot e^{\frac{-(\sigma z)^2}{2\sigma^2}} \cdot \sigma dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \cdot e^{\frac{-z^2}{2}} \cdot dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \cdot e^{-\frac{z^2}{2}} dz$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} \cdot e^{\frac{-z^2}{2}} dz$$

Even order moments about mean are given by,

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x)\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma . z)^{2n} \cdot e^{-\frac{z^2}{2}} dz$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n} \cdot e^{-\frac{z^2}{2}} \cdot dz$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}} = 2\int_{0}^{\infty} z^{2n} \cdot e^{-z^2/2} dz$$

$\left(\because z^{2n} . e^{\frac{-z^2}{2}}\right.$ is an even function).

Put $\frac{z^2}{2} = t \Rightarrow z\,dz = dt$

$$\therefore \mu_{2n} = \frac{\sigma^{2n}}{\sqrt{2\pi}} \cdot 2\int_0^\infty (2t)^{2n} \cdot e^{-\frac{2t}{2}} \cdot \frac{dt}{\sqrt{2t}}$$

$$= \frac{2^n}{\sqrt{\pi}} \sigma^{2n} \int_0^\infty e^{-t} t^{n-\frac{1}{2}} dt$$

$$= \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_0^\infty e^{-t} t^{(n+\frac{1}{2})-1} dt$$

$$\mu_{2n} = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \Gamma\left(n+\tfrac{1}{2}\right) \left(\because \Gamma_n = \int_0^\infty e^{-x} x^{n-1} dx\right)$$

Changing n to $n-1$, we get

$$\mu_{2n-2} = \frac{2^{n-1}}{\sqrt{\pi}} \sigma^{2n-2} \Gamma\left(n-\tfrac{1}{2}\right)$$

$$\therefore \frac{\mu_{2n}}{\mu_{2n-2}} = 2\sigma^2 \frac{\Gamma(n+\frac{1}{2})}{\Gamma(n-\frac{1}{2})} - 2\sigma^2\left(n-\tfrac{1}{2}\right)$$

$$[\because \Gamma n = (n-1)\Gamma(n-1)]$$

(i.e) $\mu_{2n} = \sigma^2 (2n-1)\mu_{2n-2}$.

Which gives the recurrence relation for the moments of normal distribution.

Ex: 01. x is normality distributed and the mean of x is 12 and the S.D is 4. Find out the probability of the following (i) $x \geq 20$, (ii) $x \leq 20$ (iii) $0 \leq x \leq 12$.

Given :- $\mu = 12$, $\sigma = 4$

(i) $P(x \geq 20) = P(z \geq 2)$

$= 0.5 - P(0 \leq z \leq 2)$ — (1)

$= 0.5 - 0.4772$

$= 0.0228$

(ii) To find $P(x \leq 20)$

When $x = 20$, $z = \frac{x-\mu}{\sigma} = \frac{20-12}{4} = 2$

$\therefore P(x \leq 20) = P(z \leq 2)$

$= 1 - P(z \geq 2)$

$= 1 - 0.0228$

$= 0.9772$

(iii) To find $P(0 \leq x \leq 12)$

When $x = 0$, $z = \frac{x-\mu}{\sigma} = \frac{0-12}{4} = -3$

When $x = 12$, $z = \frac{x-\mu}{\sigma}$ $\frac{12-12}{4} = 0$.

$P(0 \leq x \leq 12) = P(-3 \leq z \leq 0)$

$= 0.4987$ (from table).