Mean & Variance of Poisson Distribution

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$x \in 1, 2, \ldots \infty$$

$$\text{mean} = E(x) = \sum_x x\, P(x)$$

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

$$= \sum_x \cancel{x} \frac{\lambda^x e^{-\lambda}}{\cancel{x}(x-1)!}$$

$$= \sum_x e^{-\lambda} \frac{\lambda \lambda^{x-1}}{(x-1)!} = e^{-\lambda} \lambda \sum_x \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \cancel{e^{-\lambda}} \lambda \cancel{e^{\lambda}}$$

$$\boxed{\text{mean} = \lambda}$$

Now, $E(x(x-1)) = \sum_x x(x-1) P(x)$

$$= \sum_x x(x-1) \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \sum_x \cancel{x(x-1)} \frac{\lambda^x e^{-\lambda}}{\cancel{x(x-1)}(x-2)!}$$

$$= e^{-\lambda} \sum \lambda^2 \lambda^{(x-2)}$$

Scanned with CamScanner

$$= \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2$$

$$E(x(x-1)) = \lambda^2$$

Now, $E(x^2) = E(x(x-1)) + E(x)$

$$= \lambda^2 + \lambda$$

Variance $(x) = E(x^2) - (E(x))^2$

$$= \lambda^2 + \lambda - (\lambda)^2 = \lambda$$

$$\boxed{\begin{aligned} \sigma^2 &= \text{Variance} = \lambda \\ SD &= \sigma = \sqrt{\lambda} \end{aligned}}$$

Prove that $P(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ is probability mass fn and also find mgf, cf and p.g.f of poisson distribution.

$P(x)$ is pmf.

(i) $P(x) \geq 0$.

(ii) $\sum_x P(x) = 1$

$$\sum P(x) = \sum \frac{\lambda^x e^{-\lambda}}{x!}$$

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

$$= e^{-\lambda} \cdot e^{\lambda} = e^0 = 1$$

Proved

Scanned with CamScanner

Moment generating function:

$$m_x(t) = E(e^{xt}) = \sum e^{xt} P(x)$$

$$= \sum_x e^{xt} \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= e^{-\lambda} \sum_x \frac{(\lambda e^t)^x}{x!}$$

$$= e^{-\lambda} \cdot e^{\lambda e^t} = e^{-\lambda + \lambda e^t}$$

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

Characteristic function:

$$\phi_x(t) = E(e^{ixt}) = e^{\lambda(e^{it} - 1)}$$

Probability generating function:

$$T_x(t) = E(z^x) = e^{\lambda(z-1)}$$

Calculating mean & variance by mgf.

$$E(x) = \frac{d(M_x(t))}{dt} = \frac{d(e^{-\lambda} e^{\lambda e^t})}{dt}\Big|_{t=0}$$

$$= e^{-\lambda}(e^{\lambda e^t} \lambda e^t)_{t=0} = e^{-\lambda} e^{\lambda} \lambda$$

Mean $= E(x) = \lambda$

Scanned with CamScanner

$\boxed{\lambda = np}$ = Mean

$$E(x^2) = \frac{\partial^2 (M_x(t))}{\partial t^2} = e^{-\lambda} \frac{d(e^{\lambda e^t} \lambda e^t)}{dt}$$

$$= e^{-\lambda}(e^{\lambda}\lambda^2 + e^{\lambda}\lambda)$$

$$E(x^2) = \lambda^2 + \lambda.$$

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

$$= \cancel{\lambda^2} + \lambda - \cancel{(\lambda)^2} = \lambda$$

Q) Given that 2% of the fuses manufactured by a firm are defective.

Find the probability that a box containing 200 fuses has

(i) atleast 1 defective fuse

(ii) 3 or more defective fuse

(iii) no defective fuse

→ $n = 200$ $\quad p = 0.02$

$\lambda = np = 200(0.02) = 4$

$$P(x) = \frac{\lambda^n e^{-\lambda}}{n!} = \frac{(4)^n e^{-4}}{n!}$$

(i) atleast one defective $\quad P(x \geq 1)$

$= 1 - P(x < 1)$

$= 1 - P(0)$

$= 1 - \frac{(4)^0 e^{-4}}{0!} = 1 - e^{-4}$

Scanned with CamScanner

(ii) $P(x \geq 3) = 1 - (P(0) + P(1) + P(2))$

$= 1 - \left(\frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!}\right)$

$= 1 - 13e^{-4}$

(iii) $P(0) = e^{-4}$

Q) In a certain factory turning out blades. There is a small chance of 0.002 for any blade to be defective.

The blades are supplied in packets of 10. Using Poisson distribution, find approximate no. of packets containing.

(i) No defective blades
(ii) One defective blade.

in a consignment of 10000 packet

→ $n = 10$ $p = 0.002$

$N = 10000$.

$\lambda = np = 10 \times 0.002 = 0.02$

$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(0.02)^x e^{-0.02}}{x!}$

Scanned with CamScanner

(i) $P(0) = \frac{(0.02)^0 e^{-0.02}}{0!} = e^{-0.02}$

$= 0.9802$

$N = 10000$

$\therefore$ No of patients $= 0.9802 \times 10000$

$= 9802$.

(ii) $P(1) = \frac{(0.02)^1 e^{-0.02}}{1!} = 0.02 \times 0.9802$

No. of patients $= 10000 (P(1)) = 0.02 \times 9802$

Q) If probability of a dead reaction from a certain injection is 0.01. Find the chance that out of 200 individuals more than 2 will get bad reaction.

→ $p = 0.01$ $\quad n = 200$

$\lambda = np = 2$.

$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(2)^x e^{-2}}{x!}$.

$P(x > 2) = 1 - P(x \leq 2)$

$= 1 - [P(0) + P(1) + P(2)]$

$= 1 - \left[\frac{2^0 e^2}{0!} + \frac{2^1 e^2}{1!} + \frac{2^2 e^2}{2!}\right]$

$= 1 - 5e^{-2}$

Scanned with CamScanner

Fitting of Poisson Distribution

Q: A skilled typist on routine work kept a record of mistakes made per day during 300 working days.

mistakes/day	0	1	2	3	4	5	6
no. of days	143	90	42	12	9	3	1

$\frac{0.89^x e^{-0.89}}{x!}$

Mistakes/day x	f	Σxf	$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	$300 \times P(x)$
0	143	0	$P(0) = 0.411$	123
1	90	90	$P(1) = 0.365$	110
2	42	84	$P(2) = 0.163$	48.9 ~ 49
3	12	36	$P(3) = 0.048$	14
4	9	36	$P(4) = 0.011$	3.3
5	3	15	$P(5) = 0.02$	1
6	1	6	$P(6) = 0.003$	0

$\Sigma f = 300$ $\Sigma xf = 267$ $\Sigma f = 300$

$$\text{Mean} = \frac{\Sigma xf}{\Sigma f} = \lambda = \frac{267}{300} = 0.89$$

Exponential Distribution

A continuous random variable x which has the following p.d.f :

$$f(x) = \lambda e^{-\lambda x} \qquad \lambda > 0, \quad 0 < x < \infty$$

is called exponential variable & its distribution is called exponential distribution.

Mean & Variance of exponential distribution

$$E(x) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{0}^{\infty} x \lambda e^{-\lambda x}\,dx$$

$$= \lambda \int_{0}^{\infty} x^1 e^{-\lambda x}\,dx$$

$$= \lambda \frac{\Gamma 2}{\lambda^2} = \frac{1}{\lambda}$$

$$\int_{0}^{\infty} x^{n-1} e^{-ax} = \frac{\Gamma n}{a^n}$$

$$\Gamma 2 = 1$$

$$\boxed{\text{Mean} = E(x) = \frac{1}{\lambda}}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_{0}^{\infty} x^2 \lambda e^{-\lambda x}\,dx$$

$$= \lambda \int_{0}^{\infty} x^2 e^{-\lambda x}$$

$$= \lambda \frac{\Gamma 3}{\lambda^3} = \frac{2!}{\lambda^2}$$

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

Scanned with CamScanner

$$\sigma^2 = Var(x) = E(x^2) - (E(x))^2$$

$$= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2$$

$$\boxed{\sigma^2 = \frac{1}{\lambda^2}}$$

Moment generating fn,

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

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

$$= \int_{0}^{\infty} e^{xt} \lambda e^{-\lambda x}\,dx$$

$$= \lambda \int_{0}^{\infty} e^{-x(\lambda - t)}\,dx$$

$$= \lambda \left[\frac{e^{-x(\lambda-t)}}{\lambda - t}\right]_0^{\infty}$$

$$\text{Mean} = \frac{\lambda}{t - \lambda}$$

Characteristic fn

$$\phi_x(t) = E(e^{ixt}) = \frac{\lambda}{it - \lambda}$$

Scanned with CamScanner

Q) The length of telephone conversation is an exponential variate with mean 3 min. Find probability that call end (i) less than 3 min
(ii) takes between 3 to 5 min

→ Mean $= \frac{1}{\lambda} = 3$ $\lambda = 1/3$.

$P(x) = \lambda e^{-\lambda x} = \frac{1}{3} e^{-x/3}$

(i) $P(X<3) = \cancel{1 - P(X>)} \int_0^3 f(x)\,dx$

$= \int_0^3 \frac{1}{3} e^{-x/3}\,dx$

$= \frac{1}{3}\left[\frac{e^{-x/3}}{-1/3}\right]_0^3 = e^{-1} - 1$

(ii) $P(3<x<5) = \int_3^5 f(x)\,dx$.

$= \int_3^5 \frac{1}{3} e^{-x/3}\,dx = \frac{1}{3}\left[\frac{e^{-x/3}}{-1/3}\right]_3^5$

$= e^{-5/3} - e^{-1} = e^{-1} - e^{-5/3}$

Scanned with CamScanner

Normal Distribution

A continuous random variable which has the following p.d.f

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$$

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

is called normal variate. & the distribution is called normal distribution & is denoted by $\boxed{X \sim N(\mu, \sigma)}$

Mean & Variance of Normal Distribution

$$E(x) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

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

Let $\frac{x-\mu}{\sigma} = z$. $\quad x = \mu + \sigma z$

$$dx = \sigma dz$$

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

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \mu e^{-z^2/2} dz + \int_{-\infty}^{\infty} \sigma z e^{-z^2/2} dz \right]$$

odd fn.

Hence 0

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

Let $\frac{z^2}{2} = t$

$2dz = dt$

$dz = \frac{dt}{2}$

$dz = \frac{dt}{\sqrt{2t}}$

$$= \frac{\mu}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-t} \frac{dt}{\sqrt{2t}}$$

$$\cancel{= \frac{\mu}{\sqrt{2\pi}}}$$

$$= \frac{\cancel{2}\mu}{\cancel{\sqrt{2}}.\cancel{\sqrt{2}}\sqrt{\pi}} \int_{0}^{\infty} e^{-t} t^{-1/2} dt$$

$$= \frac{\mu}{\sqrt{\pi}} \Gamma(1/2)$$

$$= \frac{\mu}{\cancel{\sqrt{\pi}}} \cancel{\sqrt{\pi}} = \mu.$$

$$\int_{0}^{\infty} x^{n-1} e^{-x} dx = \Gamma n$$

$$\Gamma(1/2) = \sqrt{\pi}$$

$$\boxed{E(x) = \text{mean} = \mu}$$

Scanned with CamScanner

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

$$= \int x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

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

Let $\frac{x-\mu}{\sigma} = z \Rightarrow x = \mu + \sigma z$

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

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \mu^2 e^{-z^2/2} dz \right.$$

$$+ \int_{-\infty}^{\infty} \sigma^2 z^2 e^{-z^2/2} dz$$

$$\left. + 2\int_{-\infty}^{\infty} \mu\sigma z \cdot e^{-z^2/2} dz \right]$$

odd fn will be 0

$$= \frac{2}{\sqrt{2\pi}} \mu^2 \int_0^{\infty} e^{-z^2/2} dz$$

$$+ \frac{\sigma^2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-z^2/2} dz$$

Scanned with CamScanner

RANKA
DATE / /
PAGE

$z = \sqrt{2p}$

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

$\frac{\cancel{2} z\, dz}{\cancel{2}} = dz = dp$

$dz = \frac{dp}{z} = \frac{dp}{\sqrt{2p}}$

$= \frac{\cancel{2}}{\sqrt{2\pi}} \mu^2 \int_0^\infty e^{-p} \frac{dp}{\sqrt{2p}}$

$+ \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^\infty 2p\, e^{-p}\, dp$

$\cancel{= \frac{1}{\sqrt{2\pi}}}$

$\cancel{= \frac{\mu^2}{\sqrt{2}} + \frac{2\sigma^2}{\sqrt{2\pi}}}$

$= \cancel{2\mu^2}$

$= \frac{\mu^2}{\sqrt{\pi}} \int_0^\infty p^{-1/2} e^{-p}\, dp + \frac{4\sigma^2}{2\sqrt{\pi}} \int_0^\infty p^{1/2} e^{-p}\, dp$

$= \frac{\mu^2}{\sqrt{\pi}} \Gamma 1/2 + \frac{2\sigma^2}{\sqrt{\pi}} \Gamma 3/2$

$\Gamma 1/2 = \sqrt{\pi}$

$\Gamma 3/2 = \frac{1}{2} \cdot \Gamma 1/2 = \frac{1}{2}\sqrt{\pi}$

$= \frac{\mu^2}{\cancel{\sqrt{\pi}}} \cancel{\sqrt{\pi}} + \frac{\cancel{2}\sigma^2}{\cancel{\sqrt{\pi}}} \frac{1}{\cancel{2}} \cancel{\sqrt{\pi}}$

$E(x^2) = \mu^2 + \sigma^2$

Variance $= E(x^2) - (E(x))^2 = \mu^2 + \sigma^2 - (\mu)^2 = \sigma^2$

Scanned with CamScanner

Normal distribution

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

$$-\infty < x < \infty$$

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

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

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

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

Let $\frac{x-\mu}{\sigma} = z$. $\quad | x = \sigma z +$

$$\frac{dx}{\sigma} = dz$$

$$dx = \sigma dz$$

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

Scanned with CamScanner

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

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

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

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

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

$$= \frac{2 e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{1}{2}(z-\sigma t)^2} dz.$$

$$\frac{1}{2}(z-\sigma t)^2 = \theta$$

$$(z-\sigma t)\, dz = d\theta$$

$$dz = \frac{d\theta}{(z-\sigma t)}$$

$$dz = \frac{d\theta}{\sqrt{2\theta}}$$

$$M_x(t) = 2 e^{\mu t + \sigma^2 t^2/2} \int_{0}^{\infty} \frac{e^{-\theta}}{\sqrt{2\theta}} d\theta$$

$$= \frac{\cancel{2} e^{\mu t + \sigma^2 t^2/2}}{\cancel{2}\sqrt{\pi}} \int_{0}^{\infty} \theta^{-1/2} e^{-\theta} d\theta$$

Scanned with CamScanner

$$\boxed{\int_0^\infty x^{n-1} e^{-\theta} d\theta = \Gamma n}$$

$$\therefore = \frac{e^{\mu t + \sigma^2 t^2/2}}{\sqrt{\pi}} \sqrt{\pi/2}$$

$$\boxed{M_x(t) = e^{\mu t + \sigma^2 t^2/2}}$$

~~E(X²)~~

MGF of standard Normal variate Z:

$$Z = \frac{x - \mu}{\sigma}$$

$$M_z(t) = E(e^{zt}) = E\left(e^{\left(\frac{x-\mu}{\sigma}\right)t}\right)$$

$$= E\left(e^{\frac{xt}{\sigma}} e^{-\mu t/\sigma}\right)$$

$$= e^{-\mu t/\sigma} E\left(e^{xt/\sigma}\right)$$

$$= e^{-\mu t/\sigma} M_x(t/\sigma)$$

$$= e^{-\mu t/\sigma} e^{\mu t/\sigma + \frac{\sigma^2 t^2}{2}/\sigma^2}$$

$$= e^{\cancel{-\frac{\mu t}{\sigma}} + \cancel{\frac{\mu t}{\sigma}} + t^2/2}$$

$$M_z t = e^{-t^2/2}$$

Scanned with CamScanner

Mean & Variance by MGF

$$M_x(t) = e^{\mu t + \sigma^2 t^2/2}$$

$$E(X) = \frac{d(M_x(t))}{dt}\Big|_{t=0}$$

$$= \frac{d}{dt}\left(e^{\mu t + \sigma^2 t^2/2}\right)\Big|_{t=0}$$

$$= \left[e^{\mu t + \sigma^2 t^2/2}(\mu + \sigma^2 t)\right]_{t=0}$$

$$\boxed{\text{Mean} = \mu}$$

$$E(X^2) = \frac{d^2}{dt^2}(M_x(t))_t$$

$$= \frac{d}{dt}\left[e^{\mu t + \sigma^2 t^2/2}(\mu + \sigma^2 t)\right]_t$$

$$= \left[e^{\mu t + \sigma^2 t^2/2}(\mu + \sigma^2 t)^2 + e^{\mu t + \sigma^2 t^2/2}\sigma^2\right]_{t=0}$$

$$= \mu^2 + \sigma^2$$

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

$$= \mu^2 + \sigma^2 - (\mu)^2$$

$$\boxed{Var = \sigma^2}$$

Scanned with CamScanner

Normal Distribution :

Area Under Curve

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

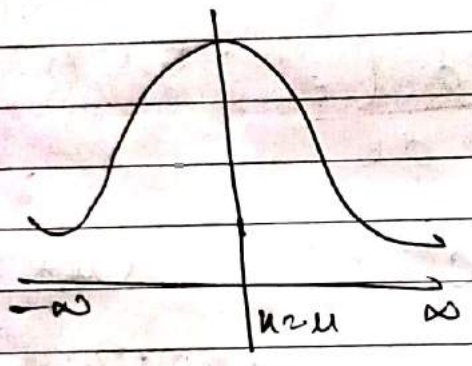
$$-\infty < x < \infty$$

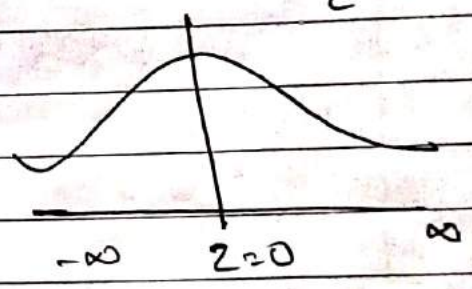
when $\mu = 0$ $\sigma = 1$

$z = \frac{x-\mu}{\sigma}$ will mark $z = x$

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

when $x = 0$
$z = 0$





If the height of 300 students are normally distributed with mean 64.5 inches & standard deviation of 3.3 inches, then how many students have height

(i) less than 5 ft.

(ii) between 5 ft & 5 ft 9 inch

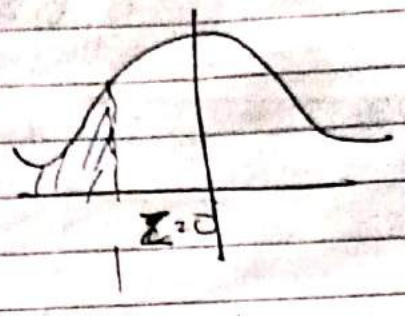
Given. $\mu = 64.5$ $\sigma = 3.3$

$$z = \frac{x-\mu}{\sigma} = \frac{x-64.5}{3.3}$$

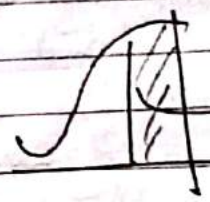
Scanned with CamScanner

(i) $P(x < 60) = P\left(z < \frac{60 - 64.5}{3.3}\right)$

$= P(z < -1.36)$



$z = -1.36$

$=$  calculate total area on left side & subtract shaded to get area.

$= 0.5 - P(-1.36 < z < 0)$

$P(x < 60) = 0.5 - P(0 < z < 1.36)$

$= 0.5 - 0.4131$ from table

$= 0.0869$

Total student with height < 5 ft $= 300 \times 0.0869$

$= 26.07 \sim 26$

(ii) $P(60 < x < 69)$

$= P(-1.36 < z < 1.36)$

$= 2 \times P(0 < z < 1.36)$

Scanned with CamScanner

$= 2 \times 0.4131$ ← from table

$= 0.8262$

300 students ; Hence, 0.8262×300
$= 248$

The distribution of 500 workers in a factory is approximately normal with mean & standard deviation ₹75 & ₹15 respectively. Find the no. of workers who receive weekly wages

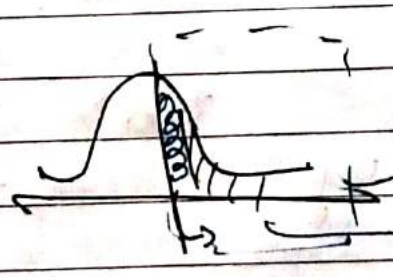
(i) more than 90
(ii) less than 45

$\mu = 75 \qquad \sigma = 15$

$$z = \frac{x-\mu}{\sigma} = \frac{x-75}{15}$$

(i) $P(X>90) = P\left(z = \frac{90-75}{15} = 1\right)$

$= P(z>1)$



$= 0.5 - P(0<z<1)$

$= 0.5 - 0.3413$ ← from table

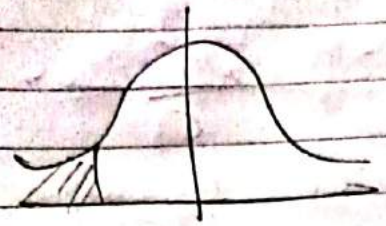
$= 0.1587$

500 workers for (i) $= 0.1587 \times 500$
$= 79$

Scanned with CamScanner

(ii) $P(X < 45) = P\left(\frac{45-90}{15}\right)$

$= P(Z < -2)$



$= 0.5 - P(-2 < Z < 0)$

$= 0.5 - P(0 < Z < 2)$

$= 0.5 - 0.4772$

$= 0.0228$

No. of workers $= 500 \times 0.0228$

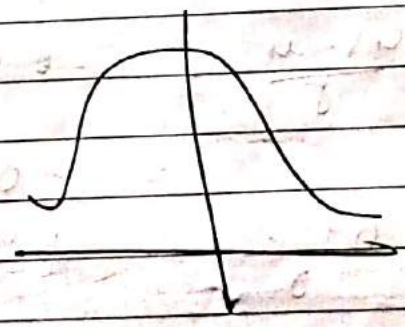
$= 500\ P(X < 45)$

$= 500 \times 0.0228$

$= 11$

Q1 In a normal distribution 31% of items are under 45 & 8% are over 64. Find mean & SD of distribution.

$\rightarrow$

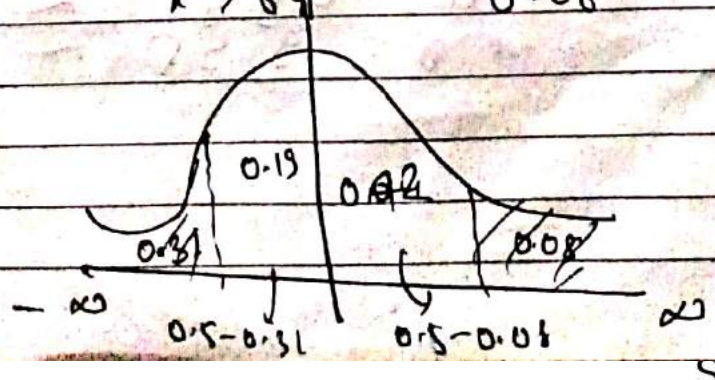


$x < 45$	0.31
$x > 64$	0.08

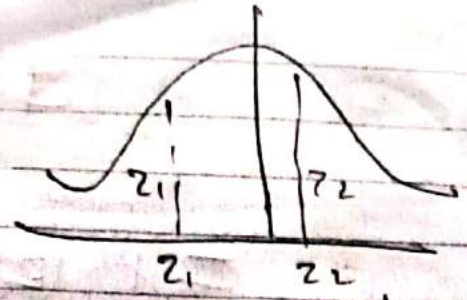
$x < 45$ 0.31

$45 < 64 \rightarrow 1 - 0.31 - 0.08 = 0.61$

$x > 64$ 0.08



Scanned with CamScanner



$z_1 = \frac{x_1 - \mu}{\sigma}$

$= \frac{45 - \mu}{\sigma}$

$P(z_1 < z < 0) = 0.19$

$z_2 = \frac{x_2 - \mu}{\sigma}$

$z_2 = \frac{64 - \mu}{\sigma}$

$P(0 < z < z_2) = 0.42$

From table when ~~$z_1 = 0$~~ $P(z_1 < z < 0) = 0.19$

$z_1 = -0.5$

$P(0 < z < z_2) = 0.42$

$z_2 = 1.41$

$\therefore z_1 = \frac{45 - \mu}{\sigma} = -0.5 = \frac{45 - \mu}{\sigma}$

$= -0.5\sigma = 45 - \mu$ — (i)

$z_2 = \frac{64 - \mu}{\sigma} = 1.41 = \frac{64 - \mu}{\sigma}$

$\Rightarrow 1.41\sigma = 64 - \mu$ — (ii)

Solving (i) & (ii)

$\mu = 49.97 \sim 50$

$\sigma = 9.95 \sim 10$

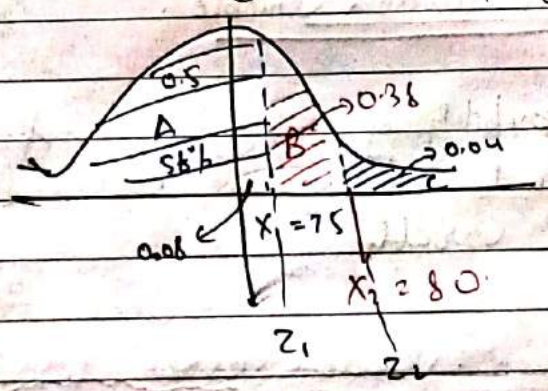
Scanned with CamScanner

Q. If skulls are classified as A, B & C according as length, breadth under as under 75, between 75 & 80, over 80.

Find approximately the mean & standard deviation of a series in which A are 58%, B are 38%, & C are 4%, being given of

$A(0.20) = 0.08$
$A(1.75) = 0.46$

A	$x < 75$	0.58
B	$75 < 80$	0.38
C	$x > 80$	0.04



$z_1 = \frac{x_1 - \mu}{\sigma}$ $\qquad$ $z_2 = \frac{x_2 - \mu}{\sigma}$

$z_1 = \frac{75 - \mu}{\sigma}$ $\qquad$ $z_2 = \frac{80 - \mu}{\sigma}$

$P(0 < z < z_1) = 0.08$ $\qquad$ $P(0 < z < z_2)$

$\frac{75 - \mu}{\sigma} = 0.08$ $\qquad$ $= 0.08 + 0.38$

$= 0.46$

$0.46 = \frac{80 - \mu}{\sigma}$

Solving both

$\mu = 74.4 \quad \sigma = 3.5$

JOINT DISTRIBUTION

BIVARIATE RANDOM VARIABLE

Let S be a sample space associated with random experiment.

Let $X(t)$ & $y(t)$ be 2 random fn each assigning a real no. to each outcome $t \in S$

Then (x, y) is called a bivariate or two dimensional random variable.

(i) **Bivariate discrete random variable.**

If the possibility of (x, y) are finite or countably infinite (x, y) is called a 2-D discrete random variable

(ii) **Bivariate Continuous random variable**

If (x, y) can be assumed all values in specified region $R(x, y)$ is called a 2D continuous random variable

Scanned with CamScanner

JOINT PROBABILITY MASS FUNCTION of (x, y)

If (x, y) is a 2D discrete random variable, where $x = x_i$ & $y = y_j$, $i, j = 1, 2, 3 \ldots$

then $P(X = x_i, Y = y_j) = P_{ij}$ is called probability mass fn of (x, y) provided

(i) $P_{ij} \geq 0$

(ii) $\sum_{i=1}^{m} \sum_{j=1}^{n} P_{ij} = 1$.

The set of triple $(x_i, y$

Marginal Probability mass function of x

$P(x = x_i) = \sum_j P_{ij} = P_{i1} + P_{i2} + \ldots$

The collection of pair (x_i, P_i) is called marginal pmf of x

Similarly marginal pmf of y.

$P(y = y_j) = \sum_i P_{ij} = P_{1j} + P_{2j} + \ldots$

The pair (y_j, P_j) is called marginal pmf of y

If x & y are independent

If $P_{xy}(x, y) = P_x(x) \cdot P_y(y)$

Scanned with CamScanner

Two balls are selected at random from a box containing 3 red, 2 green & 4 white. If x & y are the no. of red balls & green balls respectively included among 2 balls drawn from box, find

(i) Joint probability of X & Y.

(ii) Marginal probability of X & Y.

(iii) Conditional distribution of X given Y = 1

3R
2G
4W

X = 0, 1, 2, 3

Y = 0, 1, 2

X is red

Y is green.

X \ Y	0	1	2	Total
0	1/6	2/9	1/36	15/36
1	1/3	1/6	0	16/36
2	1/12	0	0	3/36
3	0	0	0	0
Total	21/36	14/36	1/36	1

P(0,0) → both balls are white

∴ 0 red & 0 green balls taken out

$$= \frac{{}^4C_2}{{}^9C_2} = \frac{1}{6}$$

P(0,1) → 0 X ball 1 green ball $= \frac{{}^2C_1\,{}^4C_1}{{}^9C_2} = \frac{2}{9}$

P(0,2) → 0 red ball 2 green balls $\frac{{}^2C_2}{{}^9C_2} = \frac{1}{36}$

P(1,0) → 1 red 0 green → 1 white $= \frac{{}^3C_1\,{}^4C_1}{{}^9C_2} = \frac{4}{3}$

P(1,1) = 1 red 1 green $= \frac{{}^3C_1\,{}^2C_1}{{}^9C_2} = \frac{1}{6}$

P(1,2) = 0

Only 2 balls can be taken out

P(2,0) → 2 red ball $= \frac{{}^3C_2}{{}^9C_2} = \frac{1}{12}$

Scanned with CamScanner

Now, (ii) Marginal pmf of x Marginal pmf of y

x	$P_x(x)$
0	15/36
1	18/36
2	3/36
3	0

y	$P_y(y)$
0	21/36
1	14/36
2	1/36
3	

(iii) Conditional pmf of X given Y = 1

$P\left(\frac{X}{Y=1}\right)$

X	$P\left(\frac{x}{y=1}\right)$
0	4/7
1	3/7
2	0
3	0

~~P(2/3)~~ $P\left(\frac{x=0}{y=1}\right)$

↳ hidden at 2/9

Total $P(Y=1) = 14/36$

$\therefore P\left(\frac{x}{y=1}\right) = \frac{2/9}{14/36} = \frac{4}{7}$

The joint probability mass function of (x, y) is given by $P(x,y) = k(2x+3y)$

$x = 0, 1, 2$ & $y = 1, 2, 3$

Find (i) k (ii) marginal pmf of x & y

(iii) conditional pmf of x given y = 1

(iv) conditional pmf of y given x = 2.

(v) probability distribution of x, y

Scanned with CamScanner

P x \ y	1	2	3	Total
0	3k	6k	9k	18k
1	5k	8k	11k	24k
2	7k	10k	13k	30k
Total	15k	24k	33k	72k

$P(x,y) = k(2x+3y)$

$72k = 1$

$k = 1/72$

(ii) ~~P x \ y 1 2 3 Total~~

(iii) $P_x(x)$.

x	$P_x(x)$
0	$18k = 18/72$
1	$24k = 24/72$
2	$30k = 30/72$

$P_y(y)$

y	$P_y(y)$
1	$15k = 15/72$
2	$24k = 24/72$
3	$33k = 33/72$

(iv) Conditional pmf of x given y = 1.

x	$P\left(\frac{x}{y=1}\right)$
0	$3k/15k = 3/15$
1	$5k/15k = 5/15$
2	$7k/15k = 7/15$

Scanned with CamScanner

Cond'nal pmf of y given x = 2

x	$P(\frac{y}{x=2})$
1	$7k/30k = 7/30$
2	$10k/30 = 10/30$
3	$13k/30k = 13/30$

(iv) Probability distribution of $z = x + y$

z	1	2		3			4		5
x	0	0	1	0	1	2	1	2	2
y	1	2	1	3	2	1	3	2	3
P(x,y)	3k	6k+5k = 11k		9k+8k+7k = 24k			11k+10k = 21k		13k

$z = x+y$	P
1	$3k = 3/72$
2	$11k = 11/72$
3	$24k = 24/72$
4	$21k = 21/72$
5	$13k = 13/72$

Are x & y as independent ?

Marginal pmf of x.

X	$P_x(x)$
0	
1	
2	

Same as dist (ii)

By definition of independence.

$$P_{x,y}(x,y) = P_x(x) \cdot P_y(y)$$

Scanned with CamScanner

$P_{xy}(0,1) = P_x(0) \times P_y(1)$

LHS = 3K = $P(x=0, y=1)$
$= 3/72$

RHS = $P_x(0) = 18/72$
$P_y(1) = 15/72$

$P_x(0) \cdot P_y(1) = \frac{18}{72} \times \frac{15}{72}$
$= \frac{5}{96}$

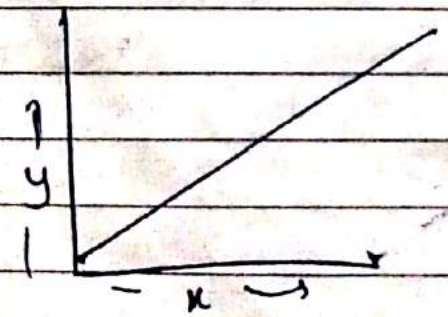
LHS $\neq$ RHS

$\therefore$ x & y are not independent

CORRELATION

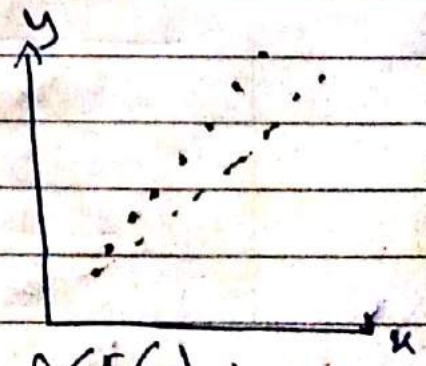
Karl Pearson coefficient of Correlation

- Correlated data - if two data are related with each other
- Uncorrelated data - if two data are not related with each other.



$r = +1$

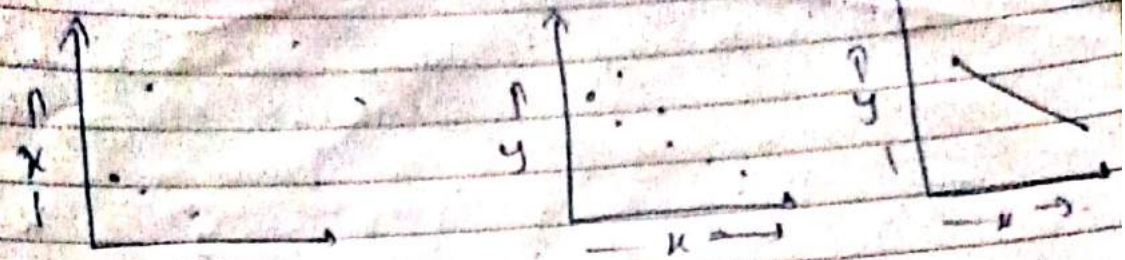
↳ Increment of x results in increment of y



$0 < r < 1$

↳ Increment but not uniform.

Scanned with CamScanner



$r=0$ $\quad$ $-1<r<0$ $\quad$ $r=-1$

Ex)

x	y
1	25
2	30
3	40
4	50
5	60

$r<1$

x	y
1	50
2	40
3	30
4	20
5	10

$r=-1$

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

$$\operatorname{cov}(x,y) = E(xy) - E(x)\,E(y)$$
$$= \frac{\sum xy}{n} - \frac{\sum x}{n}\,\frac{\sum y}{n}$$

$$\sigma_x = \sqrt{E(x^2) - [E(x)]^2}$$
$$= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$\sigma_y = \sqrt{E(y^2) - [E(y)]^2}$$
$$= \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}$$

Scanned with CamScanner

Calculate the correlation coefficient for following height (in inches) of father (x) & their sons (y)

x	y	x^2	y^2	xy
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
$\Sigma x = 554$	$\Sigma y = 552$	Σx^2 37028	Σy^2 38132	Σxy 37560

$$\text{cov}(x, y) = E(xy) - E(x)E(y)$$

$$\cancel{\sigma_x = \sqrt{}}$$

$$= \frac{\Sigma xy}{n} - \frac{\Sigma x}{n} \frac{\Sigma y}{n}$$

$$= \frac{37560}{8} - \frac{554}{8} \cdot \frac{552}{8}$$

$$= 3$$

$$\sigma_x = \sqrt{E(x^2) - [E(x)]^2}$$

$$= \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{37028}{8} - \left(\frac{554}{8}\right)^2} = 2.121$$

Scanned with CamScanner

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2}$$

$$= \sqrt{\frac{36032}{8} - \left(\frac{552}{8}\right)^2} = 2.345$$

$$r = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} = \frac{3}{(2.121)(2.345)} = 0.6032$$

Another method.

x	y	$u = x - 68$	$v = y - 69$	u^2	v^2	uv
68	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	65	-1	-4	1	16	4
68	72	0	-1	0	1	1
69	72	1	3	01	9	0
70	69	2	3	4	0	3
72	71	4	02	16	04	8
		$\Sigma u = 0$	$\Sigma v = 0$	36	44	28

from moment

$$\text{cov } uv = \frac{\Sigma uv}{n} - \frac{\Sigma u}{n}\frac{\Sigma v}{n}$$

$$= \frac{24}{8} - 0 = 3$$

$$\sigma_u = \sqrt{\frac{\Sigma u^2}{n} - \left(\frac{\Sigma u}{n}\right)^2} = \sqrt{\frac{36}{8} - 0} = 2.121$$

$$\sigma_v = \sqrt{\frac{\Sigma v^2}{n} - \left(\frac{\Sigma v}{n}\right)^2} = \sqrt{\frac{44}{8} - 0} = 2.345$$

$$r = \frac{3}{2.121 \times 2.345} = 0.6032$$

Scanned with CamScanner

Q. Find coefficient of correlation between x & y

x	y	$u = x - 5$	$v = y - 12$	u^2	v^2	uv
1	9	-4	-3	16	9	12
2	8	-3	-4	9	16	12
3	10	-2	-2	4	4	4
4	12	-1	0	1	0	0
5	11	0	-1	0	1	0
6	13	1	1	1	1	1
7	14	2	2	4	4	4
8	16	3	4	9	16	12
9	15	4	3	11	9	12
$\Sigma x = 45$	$\Sigma y = 106$	$\Sigma u = 0$	$\Sigma v = 0$	$\Sigma u^2 = 60$	$\Sigma v^2 = 60$	$\Sigma uv = 57$

$\rightarrow \bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$

$\bar{y} = \frac{\Sigma y}{n} = \frac{106}{9} = 12$

$r = \frac{\text{cov}(u, v)}{\sigma u \, \sigma v}$

$\text{cov}(u, v) = \frac{\Sigma uv}{n} - \frac{\Sigma u}{n} \cdot \frac{\Sigma v}{n}$

$= \frac{57}{9} - 0 = \frac{57}{9} = 6.33$

$\sigma u = \sqrt{\frac{\Sigma u^2}{n} - \left(\frac{\Sigma u}{\Sigma n}\right)^2} = \sqrt{\frac{60}{9} - 0}$

$= 2.582$

$\sigma v = \sqrt{\frac{\Sigma v^2}{n} - \left(\frac{\Sigma v}{\Sigma n}\right)^2} = \sqrt{60/9}$

$= 2.582$

$r = 0.95$

Scanned with CamScanner

Regression

If the curve of regression is a straight line then it is said to be line of regression.

If the curve of regression is not straight then the curve is said to be curvilinear regression or non linear regression.

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

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

$$\boxed{b_{yx} = r\frac{\sigma_y}{\sigma_x}}$$

$$\boxed{b_{xy} = r\frac{\sigma_x}{\sigma_y}}$$

$$r = \sqrt{b_{xy}\, b_{yx}}$$

The point at which x or y intersects y on x is called mean.

Q1 Calculate the coefficient of correlation & obtain line of regression for following data:-

Ans

x	y	$u = x - 4$	$v = y - 12$	u^2	v^2	uv
1	9	-4	-3	16	9	12
2	8	-3	-4	9	16	12
3	10	-2	-2	4	4	4
4	12	-1	0	1	0	0
5	11	0	-1	0	1	0
6	13	1	1	1	1	1
7	14	2	2	4	4	4
8	16	3	4	9	16	12
9	15	4	3	16	9	12
$\Sigma x = 45$	$\Sigma y = 108$	$\Sigma u = 0$	$\Sigma v = 0$	$\Sigma u^2 = 60$	$\Sigma v^2 = 60$	Σuv

Scanned with CamScanner

$\bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$

$\bar{y} = \frac{\Sigma y}{n} = \frac{108}{9} = 12$

$r_{xy} = r_{uv} = \frac{\text{cov}(u,v)}{\sigma_u \sigma_v}$

$\text{cov}(u,v) = \frac{\Sigma uv}{n} - \frac{\Sigma u}{n}\frac{\Sigma v}{n} = \frac{57}{9} - 0 = 6.33$

$\sigma_u = \sqrt{\frac{\Sigma u^2}{n} - \left(\frac{\Sigma u}{n}\right)^2} = \sqrt{\frac{60}{9} - 0} = 2.582$

$\sigma_v = \sqrt{\frac{\Sigma v^2}{n} - \left(\frac{\Sigma v}{n}\right)^2} = \sqrt{\frac{60}{9} - 0} = 2.582$

$r_{xy} = r_{uv} = \frac{\text{cov}(u,v)}{\sigma_u \sigma_v} = \frac{6.33}{2.582 \times 2.583}$

$= 0.95$

Now, line of regression

$\bar{u} = \bar{x} - 5$

$\bar{x} = \bar{u} + 5$

$\bar{v} = \bar{y} - 12$

$\bar{y} = \bar{v} + 12$

" " " "

<u>X on Y</u> $\Rightarrow$ $x - \bar{x} = b_{xy}(y - \bar{y})$

$\Rightarrow x - (\bar{u} + 5) = r_{uv} \frac{\sigma_u}{\sigma_v}(y - (\bar{v} + 12))$

$\Rightarrow x - 5 = 0.95 \times \frac{2.582}{2.582}(\bar{y} - 12)$

$\Rightarrow x = 0.95y - 6.4$

<u>Y on x</u> $y - \bar{y} = b_{yx}(x - \bar{x})$

$\Rightarrow y - (\bar{v} + 12) = r_{uv} \frac{\sigma_v}{\sigma_u}(x - (\bar{u} + 5))$

$\Rightarrow y - 12 = 0.95 \times \frac{2.58}{2.58}(x - 5)$

$\Rightarrow y = 0.95x + 0.726$

Scanned with CamScanner

Q For bivariate distribution vs 18,

$\Sigma x^2 = 60$ $\Sigma y^2 = 96$

$\Sigma x = 12$ $\Sigma y = 18$ $\Sigma xy = 48$

$\rightarrow$ $\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{60}{18} - \left(\frac{12}{18}\right)^2} = \sqrt{2.884}$

$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2} = \sqrt{\frac{96}{18} - \left(\frac{18}{18}\right)^2} = \sqrt{4.3333}$

$\text{cov}(x, y) = \frac{\Sigma xy}{n} - \frac{\Sigma x}{n} \frac{\Sigma y}{n}$

$= \frac{48}{18} - \frac{12}{18} \cdot \frac{18}{18} = 1.9997$

$r = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{1.9997}{\sqrt{2.884} \sqrt{4.333}} = 0.57$

Line of regression

$\bar{x} = \frac{\Sigma x}{n} = \frac{12}{18}$

x on y

$\bar{y} = \frac{\Sigma y}{n} = \frac{18}{18}$

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

$\Rightarrow b_{xy} = r_{xy} \frac{\sigma_y}{\sigma_x} = \frac{0.57 \sqrt{2.884}}{\sqrt{4.333}}$

$= 0.4615$

$\Rightarrow x - 0.67 = 0.4615(y - 1)$

y on x

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

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

$y - 1 = 0.692 (x - 0.67)$

Scanned with CamScanner

Q) You are given following data

	x	y
Arithmetic mean	36	85
Std. deviation	11	8

Correlation coefficient between x & y = 0.66

(i) Find 2 regression equation.
(ii) Estimate value of x when y = 75.

→ Given, $\bar{x} = 36$ $\bar{y} = 85$ $r = 0.66$
$\sigma_x = 11$ $\sigma_y = 8$

y on x

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

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

$$\Rightarrow y - 85 = 0.66 \times \frac{8}{11}(x - 36)$$

$$\Rightarrow y - 85 = 0.48(x - 36)$$

$$\Rightarrow y = 0.48x + 67.72$$

x on y

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

$$\Rightarrow x - 36 = r\frac{\sigma_x}{\sigma_y}(y - 85)$$

$$\Rightarrow x - 36 = 0.66 \times \frac{11}{8}(y - 85)$$

$$\Rightarrow x = 0.9075y - 41.1375$$

Scanned with CamScanner

(ii) Using x on y

$y = 75$

$x = 0.8075 \times 75 - 41.875$

$x = 26.925$

Q1 In a partially destroyed laboratory, record of an analysis of correlation data then following results only are legible $var(x) = 9$, regression equations $8x - 10y + 66 = 0$

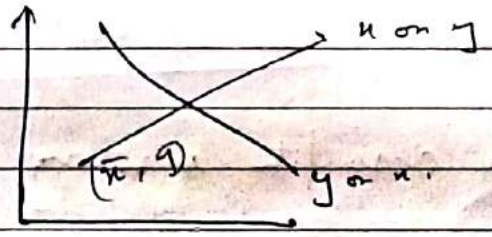
$40x - 18y = 214$

(i) find mean values of x & y.

(ii) coefficient of correlation

(iii) the standard deviation of y

→



$8\bar{x} - 10\bar{y} + 66 = 0$

$40\bar{x} - 18\bar{y} - 214 = 0$

(i) $\bar{x} = 13$ $\bar{y} = 17$

(ii) x on y

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

y on x

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

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

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

$r = \pm\sqrt{b_{xy} b_{yx}}$

$b_{xy} b_{yx} \geq 0$ $r > 0$

$b_{xy} b_{yx} < 0$ $r < 0$

Scanned with CamScanner

This mean b_{xy} & b_{yx} both
which is possible

Hence,

$8x - 10y + 66 = 0$

Let's take it as x on y

$x = \frac{18y}{8} - \frac{66}{8}$

$b_{xy} = \frac{18}{8}$

$40x - 18y = 214$

Let's take it as y on x

$18y = 40x - 214$

$y = \frac{40}{18}x - \frac{214}{18}$

$b_{yx} = \frac{40}{18}$

$r = \sqrt{b_{xy} b_{yx}}$

$= \sqrt{\frac{18}{8} \times \frac{40}{8}} = \frac{\sqrt{720}}{8} > 1$

$\boxed{\text{We know } r \text{ lies between } -1 \le r \le 1.}$

Hence assumption is wrong.

$8x - 10y + 66 = 0$

Let's take it as y on x

$\frac{8}{10}x + \frac{66}{10} = y$

$b_{yx} = 8/10$

$40x - 18y - 214 = 0$

Let's take it as x on y.

$x = \frac{18}{40}y + 214$

$b_{xy} = 18/40$

$r = \sqrt{b_{xy} b_{yx}} = \sqrt{8/10 \times 18/40}$

$= \frac{12}{20} < 1.$

$\boxed{r = +0.6}$

(iii) $\sigma_x^2 = 9$ find σ_y

$\sigma_x = 3$

Scanned with CamScanner

$$b_{yx} = r\frac{\sigma_y}{\sigma_x} = 0.6 \times \frac{\sigma_y}{3}$$

$$\Rightarrow \frac{4}{10} = 0.6 \times \frac{\sigma_y}{3}$$

$$\Rightarrow 4 = \sigma_y$$

Q) Lines $2x + 3y = 10$ & $4x + 5y = 18$ are lines of regression obtained from variables x & y.

Decide which one is the line of regression of x on y.

Given $x = 5$ find y.
$y = 2$ find x

→ Let's take it as x on y

$2x + 3y = 10$

$x = -\frac{3}{2}y + \frac{10}{2}$

$b_{xy} = \frac{-3}{2}$

$4x + 5y = 18$

$y = \frac{18}{5} - \frac{4x}{5}$

$b_{yx} = \frac{-4}{5}$

$$r = \pm\sqrt{b_{xy}\, b_{yx}} = \pm\sqrt{-\frac{3}{2} \times \frac{-4}{5}}$$

$$r = \sqrt{\frac{6}{5}} > 1$$ ✗ Not valid.

Now opposite.

y on x

$2x + 3y = 10$

$y = \frac{-2}{3} + \frac{10}{3}$

$b_{yx} = -2/3$

Ans

x on y

$4x + 5y = 18$

$x = -\frac{5}{4}y + \frac{18}{4}$

$b_{xy} = -5/4$

$$r = \sqrt{2/3 \times \frac{5}{4}} = \sqrt{5/6} = \text{True}.$$

Scanned with CamScanner

$x = 5$ find y

y on x $\quad y = -\frac{2}{3}(5) + 10/3$

$y = 0$

$y = 2$ find x

x on y $\quad x = \frac{-5}{4} y + 15/4$

$x = 2$

Angle between 2 lines of regression

Q1 ~~Show tan θ, the acute angle between 2 lines of regression is given by~~

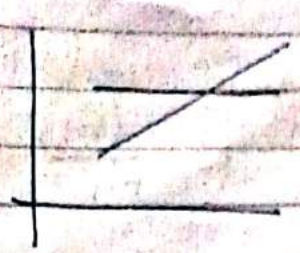
tan

Q1 Show tan θ, the acute angle between 2 lines of depression is given by.

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

Interpret the case when $r = 0$, $r = \pm 1$

Scanned with CamScanner



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

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

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

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

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

$$y-\bar{y} = \underbrace{\boxed{r\frac{\sigma_y}{\sigma_x}}}_{m_2}(x-\bar{x})$$

$$\tan\theta = \pm\frac{m_1-m_2}{1+m_1m_2}$$

$$y-\bar{y} = \underbrace{\boxed{\frac{1}{r}\frac{\sigma_y}{\sigma_x}}}_{m_1}(x-\bar{x})$$

$$\tan\theta = \pm\frac{\frac{1}{r}\frac{\sigma_y}{\sigma_x} - r\frac{\sigma_y}{\sigma_x}}{1+\frac{1}{r}\frac{\sigma_y}{\sigma_x}\,r\frac{\sigma_y}{\sigma_x}}$$

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

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

(ii) when $r=0$

$\tan\theta = 0 \longrightarrow r=0$ uncorrelated data.

$$\tan\theta = \left(\frac{1-0}{0}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right) = \infty$$

$$\theta = 90^\circ$$

Scanned with CamScanner

(ii) when $r = \pm 1$

$$\tan\theta = \left(\frac{1-(\pm 1)^2}{\pm 1}\right)\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2} = 0$$

$\tan\theta = 0$ $\qquad$ $\tan\theta = 0$

$\theta = 0$ $\qquad$ $\theta = \pi$

SAMPLING

- Sampling saves time & is less costly.
- Sampling from is a list of items that make up population.
- ~~Outliers~~ - Inaccurate results will be obtained if certain parts are excluded from the frame.
- Using different frame will give different results

Probability Sampling - chance of selection of every element is equal.

Non probability sampling - chance of selection of every element is not equal.

Convenient sampling - easy, inexpensive & convenient to sample.

Judgmental sampling - based on experience of person choosing sampling

Quota sampling - based on group.
7000 UG, 3000 PG
∴ sample of 10 will have 7 & 3 UG & PG.

— Snowball — Non probability
↳ similar sampling collect

SRS. (Simple Random sample).

- Each individual item has equal chance of getting selected
- Selection may be with or without replacement

Systematic random sample.

Sample size = n. Total individuals = N

∴ No. of groups = $k = N/n$

Select every k^{th} individual randomly

Stratified sample

1. Divide population into groups.
2. Select using SRS from each group according to population size
3. Combine samples into one

Grp 4000 UG 1500 M Tech 2500 Ph

Total = 8000

Sample = 1800

$UG = \frac{4000}{8000} \times 100 = 50$