Q. The time in hours required to repair a machine is exponentially distributed with parameter $\lambda = 1/20$. What is the pb that the required time
i) exceeds 30 hours
ii) is b/w 16 hours & 24 hours.
iii) atmost 10 hours.

→ Let X represent time required for repairing the machine : $\lambda = 1/20$.

$f(x) = \lambda e^{-\lambda x} = \frac{1}{20} e^{-1/20 \times x} \; ; \; x > 0.$

i) $P(X > 30) = \int_{30}^{\infty} f(x)\,dx = \int_{30}^{\infty} \frac{1}{20} e^{\frac{-x}{20}} dx$

$= \frac{1}{20}\left[\frac{e^{-x/20}}{-1/20}\right]_{30}^{\infty} = \frac{1}{20} \times \frac{20}{1}\left(e^{-\infty} + e^{-3/2}\right)$ $\;\; -e^{-\infty} + e^{-3/2}$

$= e^{-3/2} = 0.2231$

ii) $P(16 < X < 24) = \int_{16}^{24} \frac{1}{20} e^{\frac{-1}{20}x} dx$

$= -\left[e^{-x/20}\right]_{16}^{24} = \left(-e^{\frac{-24}{20}} + e^{-16/20}\right)$

$= e^{-4/5} - e^{-6/5} = 0.148$

iii) $P(X \leq 10) = \int_{0}^{10} f(x)\,dx = \int_{0}^{10} \frac{1}{20} e^{-1/20\, x} dx$

$= -\left[e^{\frac{-x}{20}}\right]_{0}^{10} = -e^{-1/2} + e^{0}$

$= 1 - e^{-1/2} = 0.3934$

Q. The amount of time that a watch will run without having to be reset is a random variable having an exponential distrib^n with mean 120 days. find pb that such a watch will (i) have to be set in less than 24 days.
ii) not have to be reset in atleast 180 days

→ Let X denote amt of time that a watch will run without reseting.

Mean $= \frac{1}{\lambda} = 120$.

pdf of X is, $f(x) = \frac{1}{120} e^{\frac{-x}{120}} \; ; \; x > 0.$

i) P(watch has to be reset in less than 24 days) = P(watch will run for less than 24 days) $= P(X < 24)$

$= \int_{0}^{24} f(x)\,dx = \int_{0}^{24} \frac{1}{120} e^{\frac{-x}{120}} dx$

$= \frac{-1}{120} \times \frac{120}{1}\left[e^{\frac{-x}{120}}\right]_{0}^{24} = -e^{\frac{-24}{120}} + e^{0}$

$= 1 - e^{-0.2} = 0.1813$

Page No :
Date :

ii) P(watch will not have to be reset in atleast 180 days)

$= P(\text{watch will run for atleast } 180 \text{ days})$

$= P(X > 180)$

$= \int_{180}^{\infty} f(x)\,dx = \int_{180}^{\infty} \frac{1}{120} e^{\frac{-x}{120}}\,dx$

$= -\left[e^{\frac{-x}{120}}\right]_{180}^{\infty} = -e^{-\infty} + e^{\frac{-180}{120}}$

$= 0 + e^{-3/2} = \underline{0.2231}$

Q. The mileage which a car owner gets with a certain kind of tyre is a r.v having an ~~unexpected~~ exponential distrbn with mean 40000km. Find pb that one of these tyres will last.

a) atleast 20,000 kms

b) atmost 30,000 kms

$\rightarrow f(x) = \frac{1}{40000} e^{\frac{-x}{40,000}} ;\ x > 0$

a) $P(x \geq 20,000) = \int_{20,000}^{\infty} \frac{1}{40,000} e^{\frac{-x}{40,000}}\,dx$

$= -\left[e^{\frac{-x}{40,000}}\right]_{20,000}^{\infty} = -e^{-\infty} + e^{\frac{-2}{4}}$

$= 0 + e^{-1/2} = \underline{0.6065}$

Page No :
Date :

b) $P(x \leq 30,000) = \int_{0}^{30,000} \frac{1}{40,000} e^{\frac{-x}{40,000}}\,dx$

$= -\left[e^{\frac{-x}{40,000}}\right]_{0}^{30,000} = -e^{\frac{-3}{4}} + e^{0}$

$= 1 - e^{-3/4} = \underline{0.5276}$

Uniform Distribution [Continuous] :-

The pdf of a uniform distbn in the interval (a, b) is

$$f(x) = \begin{cases} \frac{1}{b-a} ; & a \leq x \leq b \\ 0 ; & \text{otherwise} \end{cases}$$

Distribution funcn :-

$F(x) = 0$; if $x < a$;

$= \int_{-\infty}^{x} f(x)\,dx = \int_{a}^{x} \frac{1}{b-a}\,dx$

$= \frac{1}{b-a}[x]_a^x = \frac{1}{b-a}(x-a)$; $a \leq x \leq b$

$F(x) = 1$; $x \geq b$

Mean & Variance :-

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

$= \int_{b}^{a} x \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_b^a$

Page No :
Date :

$$= \frac{1}{2(b-a)}(b^2-a^2) = \frac{\cancel{(b-a)}(b+a)}{2\cancel{(b-a)}}$$

$$= \frac{b+a}{2}$$

$$E(x^2) = \int_a^b x^2 f(x)\,dx = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)}$$

$$= \frac{\cancel{(b-a)}(a^2+ab+b^2)}{3\cancel{(b-a)}} = \frac{a^2+ab+b^2}{3}$$

$$Var(x) = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{a^2+ab+b^2}{3} - \frac{(a^2+2ab+b^2)}{4}$$

$$= \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12}$$

$$= \frac{a^2-2ab+b^2}{12} = \frac{(a-b)^2}{12}$$

9. A random variable X has a uniform distrbⁿ over $(-3,3)$. Compute

a) $P(X<2)$

b) $P(|X|<2)$

c) $P(|X-2|<2)$

d) Find k for which $P(X>k) = 1/3$

→ pdf of X is

$$f(x) = \begin{cases} \frac{1}{3-(-3)} = \frac{1}{6} ; & -3<x<3 \\ 0 ; & \text{elsewhere} \end{cases}$$

a) $P(X<2) = \int_{-3}^{2} f(x)\,dx = \int_{-3}^{2} \frac{1}{6}\,dx = \frac{1}{6}[x]_{-3}^{2}$

$$= \frac{1}{6}(2+3) = 5/6$$

b) $P(|X|<2) = P(-2<X<2) = \int_{-2}^{2}\frac{1}{6}\,dx$

$$= \frac{1}{6}(2+2) = 4/6 = 2/3$$

c) $P(|X-2|<2) = P(-2<X-2<2)$

$$= P(-2+2<X<2+2)$$

$$= P(0<X<4)$$

$$= P(0<X<3) + P(3<X<4)$$

$$= \int_0^3 f(x)\,dx + \int_3^4 f(x)\,dx$$

$$= \int_0^3 \frac{1}{6}\,dx + \int_3^4 0\cdot dx = \frac{1}{6}\times 3 = \frac{1}{2}$$

d) $P(X>k) = 1/3 \Rightarrow \int_k^3 f(x)\,dx = 1/3$

$$\Rightarrow \int_k^3 \frac{1}{6}\,dx = \frac{1}{3} \Rightarrow \frac{1}{6}[x]_k^3 = \frac{1}{3}$$

$$\Rightarrow \frac{1}{\cancel{6}_2}(3-k) = \frac{1}{\cancel{3}} \Rightarrow 3-k = 2 \Rightarrow k = 1$$

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

Q: If X has a uniform distrib^n in $(-k,k)$; $k>0$. find k if $P(|x|<1) = P(|x|>1)$

→ pdf of X is $f(x) = \begin{cases} \frac{1}{k-(-k)} = \frac{1}{2k} & ; -k<x<k \\ 0 & ; \text{otherwise} \end{cases}$

$P(|x|<1) = P(|x|>1) = 1-(P|x|\leq 1)$

$\Rightarrow 2P(|x|\leq 1) = 1$

$\Rightarrow 2P(-1<x<1) = 1$

$\Rightarrow 2\int_{-1}^{1} f(x)\,dx = 1 \Rightarrow 2\int_{-1}^{1} \frac{1}{2k}\,dx = 1$

$\Rightarrow 2\frac{1}{2k}[x]_{-1}^{1} = 1 \Rightarrow \frac{1}{k}(1-(-1)) = 1$

$\Rightarrow \frac{1}{k}\times 2 = 1 \Rightarrow k = 2$

Q: If X is uniformly distributed rv with mean $=1$ and variance $= \frac{4}{3}$. find $P(X<0)$.

→ Mean $= \frac{a+b}{2} = 1 \Rightarrow a+b = 2$ —— (1)

Variance $= \frac{(a-b)^2}{12} = \frac{4}{3} \Rightarrow (a-b)^2 = 16$

$\Rightarrow a-b = \pm 4$

$a+b=2$
$a-b=-4$
$2a=-2$
$a=-1$

Since $a<b$, $a-b=-4$ —— (2)

(1) + (2) $\Rightarrow 2a = -2 \Rightarrow a = -1$

From (2) $\Rightarrow b = a+4 = -1+4 \Rightarrow b = 3$

$\therefore f(x) = \begin{cases} 1/4 & ; -1<x<3 \\ 0 & ; \text{otherwise} \end{cases}$

$P(X<0) = \int_{-1}^{0} f(x)\,dx = \int_{-1}^{0} \frac{1}{4}\,dx = \frac{1}{4}[x]_{-1}^{0}$

$= \frac{1}{4}(0+1) = 1/4$

⇒ **Normal Distribution:-**

Its the most important pb distrib^n in statistics. It is a continuous distrib^n.

Def^n ⇒ A continuous rv X is said to follow a normal distrib^n if its pdf $f(x)$ is given by;

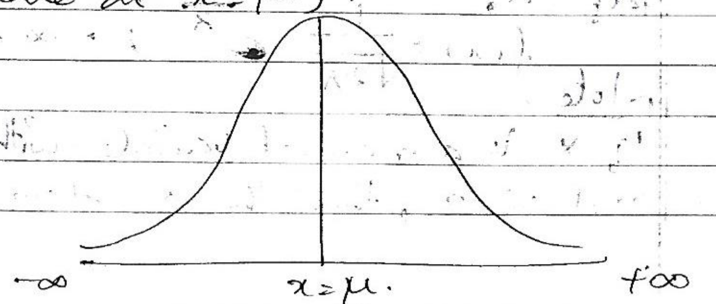
$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} ; \quad -\infty<x<\infty$$

and μ & σ are parameters with $\sigma>0$ and its denoted by $N(\mu,\sigma)$.

Note:-

1) The mean & SD of normal distrib^n are μ & σ respectively, $-\infty<x<\infty$; $\sigma>0$

2) The normal curve is bell shaped and symmetrical about its mean (The ordinate at $x=\mu$).



Page No:
Date:

3) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

ie; total area under normal curve is 1.
The ordinate at $x=\mu$ divides the area into 2 equal parts.

4) $P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)dx$

$P(X < x_2) = \int_{-\infty}^{x_2} f(x)\, dx$

$P(X \geq x_1) = \int_{x_1}^{\infty} f(x)\, dx$

5) Since normal pb density funcⁿ cannot be integrated in the closed form b/w every pair of limits x_1 & x_2, we use a table called std normal table to find the integral.

⇒ Standard Normal Distbⁿ :-

A normal distbⁿ with mean = 0 & SD = 1 is called std normal distbⁿ if its pdf is,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \; ; -\infty < x < \infty$$

Note:-
If X is a normal variate with mean μ and SD σ, then the random variable

$z = \frac{x-\mu}{\sigma}$ follows std normal distbⁿ. Z is called std normal variate.

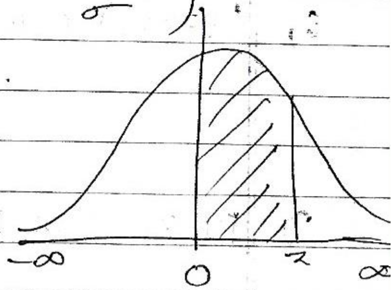
ie; if $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$, the table given value of integral $\int_0^z \phi(z)\, dz$.

ie; $P(x_1 < X < x_2) = P\left(\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{x_2-\mu}{\sigma}\right)$

$= P\left(\frac{x_1-\mu}{\sigma} < z < \frac{x_2-\mu}{\sigma}\right)$

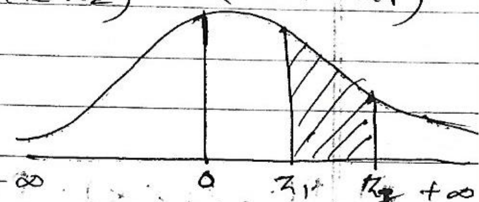
$= \int_{\frac{x_1-\mu}{\sigma}}^{\frac{x_2-\mu}{\sigma}} \phi(z)\, dz.$

Note:-

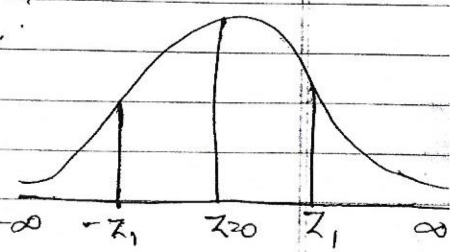


1) $P(0 < z < z_1)$ can be taken from table.

2) $P(z_1 < z < z_2) = P(0 < z < z_2) - P(0 < z < z_1)$ if $z_1, z_2 > 0$.

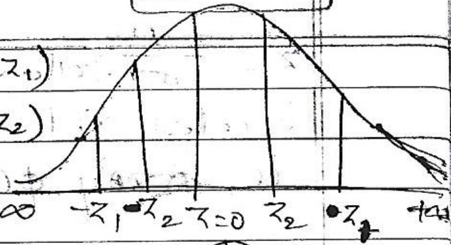


3) $P(-z_1 < z < 0) = P(0 < z < z_1)$, since curve is symmetric about zero.



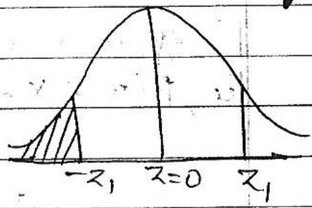
4) $P(-z_1 < z < -z_2) = P(z_2 < z < z_1)$

$= P(0 < z < z_1) - P(0 < z < z_2)$

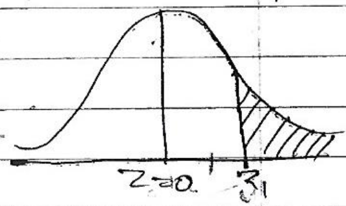


5) $P(z < -z_1) = 0.5 - P(-z_1 < z < 0)$

$= 0.5 - P(0 < z < z_1)$

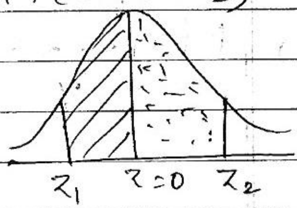


6) $P(z > z_1) = 0.5 - P(0 < z < z_1)$



7) $P(-z_1 < z < z_2) = P(-z_1 < z < 0) + P(0 < z < z_2)$

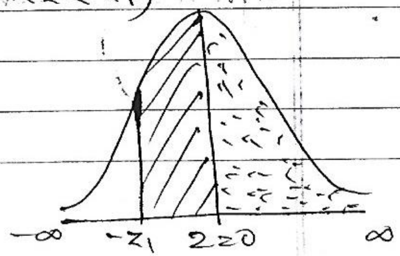
$= P(0 < z < z_1) + P(0 < z < z_2)$

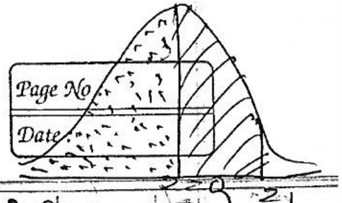


8) $P(-z_1 < z < \infty) = P(-z_1 < z < 0) + P(0 < z < \infty)$

$= 0.5 + P(-z_1 < z < 0)$

$= 0.5 + P(0 < z < z_1)$





9) $P(-\infty < z < z_1) = P(-\infty < z < 0) + P(0 < z < z_1)$

$= 0.5 + P(0 < z < z_1)$

Q. If X is a normal variate with mean 30 and SD 5. find

(i) $P(26 \le X \le 40)$ (ii) $P(X \ge 45)$

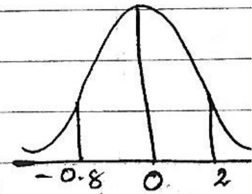
(iii) $P(|X - 30| > 5)$

→ $\mu = 30$ and $\sigma = 5$

Put $z = \frac{X-\mu}{\sigma} = \frac{X-30}{5}$

i) $P(26 \le X \le 40) = P\left(\frac{26-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{40-\mu}{\sigma}\right)$

$= P\left(\frac{26-30}{5} \le z \le \frac{40-30}{5}\right)$



$= P(-0.8 \le z \le 2)$

$= P(-0.8 \le z \le 0) + P(0 \le z \le 2)$

$= P(0 \le z \le 0.8) + P(0 \le z \le 2)$

$= 0.2881 + 0.4772 = 0.7653$

ii) $P(X \ge 45) = P\left(\frac{X-\mu}{\sigma} \ge \frac{45-\mu}{\sigma}\right)$

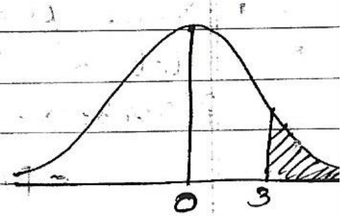
$= P\left(z \ge \frac{45-30}{5}\right)$

$= P(z \ge 3)$

$= 0.5 - P(0 \le z \le 3)$

$= 0.5 - 0.4987$

$= 0.0013$



iii) $P(|x-30|>5) = P((x-30)>5 \text{ or } (x-30)<-5)$

$= P(X>35) + P(X<25)$

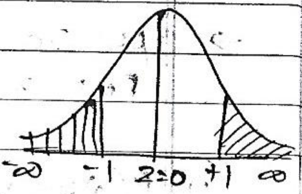
$= P\left(\frac{x-\mu}{\sigma} > \frac{35-\mu}{\sigma}\right) + P\left(\frac{x-\mu}{\sigma} < \frac{25-\mu}{\sigma}\right)$

$= P\left(z > \frac{35-30}{5}\right) + P\left(z > \frac{25-30}{5}\right)$

$= P(z>1) + P(z<-1)$

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

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

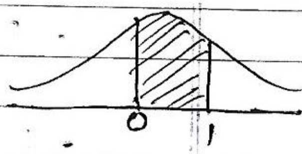


0.4669
0.2684
0.0401
0.2266

$= 1 - 2\,P(0<z<1)$

$= 1 - 2 \times 0.3413$

$= 0.3174$



Q. For a normally distributed variate with mean 1 and SD 3, find pb that $3.43 \le X \le 6.19$.

→ Since $\mu = 1$ and $\sigma = 3$

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

$P(3.43 \le X \le 6.19) = P\left(\frac{3.43-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{6.19-\mu}{\sigma}\right)$

$= P\left(\frac{3.43-1}{3} \le z \le \frac{6.19-1}{3}\right)$

$= P(0.81 \le z \le 1.73)$

$= P(0 \le z \le 1.73) - P(0 \le z \le 0.81)$

$= 0.4582 - 0.2910$ (from normal table)

$= 0.1672$

Q. If x is normally distributed with mean 8 and SD 4. find (a) $P(5 \le X \le 10)$ (b) $P(10 \le X \le 15)$ (c) $P(X \ge 15)$ (d) $P(X \le 5)$

→ $\mu = 8$; $\sigma = 4$; $z = \frac{x-\mu}{\sigma}$

a) $P(5 \le x \le 10) = P\left(\frac{5-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{10-\mu}{\sigma}\right)$

$= P\left(\frac{5-8}{4} \le z \le \frac{10-8}{4}\right)$

$= P\left(-\frac{3}{4} \le z \le \frac{1}{2}\right)$

$= P(-0.75 \le z \le 0.5)$

$= P(0 \le z \le 0.75) + P(0 \le z \le 0.5)$

$= 0.2734 + 0.1915 = 0.4649$ 0.4649

b) $P(10 \le x \le 15) = P\left(\frac{10-\mu}{\sigma} \le \frac{x-\mu}{\sigma} \le \frac{15-\mu}{\sigma}\right)$

$= P\left(\frac{10-8}{4} \le z \le \frac{15-8}{4}\right)$

$= P(0.5 \le z \le 1.75)$

~~$= P(0 \le z \le 0.5) + P(0 \le z \le 1.75)$~~

~~$= 0.1915 + 0.4599$~~ 0.4599

$= P(0 \le z \le 1.75) - P(0 \le z \le 0.5)$

$= 0.4599 - 0.1915$

$= 0.2684$

c) $P(x \ge 15) = P\left(\frac{x-\mu}{\sigma} \ge \frac{15-\mu}{\sigma}\right)$

$= P\left(\frac{x-\mu}{\sigma} \geq \frac{15-8}{4}\right)$

$= P(z \geq 7/4)$

$= P(z \geq 1.75)$

$= 0.5 - P(0 \leq z \leq 1.75)$

$= 0.5 - 0.4599$

$= 0.04011$

d) $P(x \leq 5) = P\left(\frac{x-\mu}{\sigma} \leq \frac{5-\mu}{\sigma}\right)$

$= P\left(\frac{x-\mu}{\sigma} \leq \frac{5-8}{4}\right)$

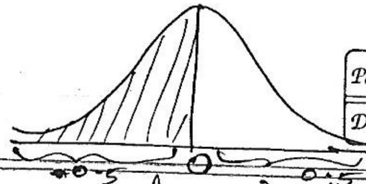
$= P(z \leq -3/4)$

$= P(z \leq -0.75)$

$= 0.5 - P(0 \leq z \leq 0.75)$

$= 0.5 - 0.2734$

$= 0.2266$



Note:- In a normal variant, the total area to the left ordinate of $x = \mu$ is 0.5

Imp Q) In a normal distrib^n 7% of items are under 35 and 89% are under 63. What are the mean & SD of distrib^n?

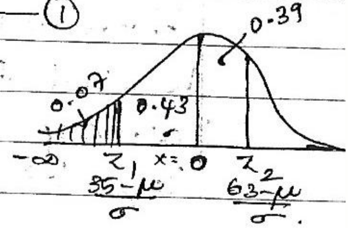
→ Let $z = \frac{x-\mu}{\sigma}$, x be the normal variate with mean μ & SD σ.

Since only 7% are below 35

i.e; $P(X<35) = \frac{7}{100} = 0.07$ —(1)

Since 89% are below 63

i.e; $P(X<63) = 0.89$ —(2)



from (1);

$P(x<35) = 0.07$

$P\left(\frac{x-\mu}{\sigma} < \frac{35-\mu}{\sigma}\right) = 0.07$

$P\left(z < \frac{35-\mu}{\sigma}\right) = 0.07$

$\Rightarrow 0.5 - P\left(\frac{35-\mu}{\sigma} < z < 0\right) = 0.07$

$P\left(\frac{35-\mu}{\sigma} < z < 0\right) = 0.05 - 0.07 = 0.43$

from std table, we get;

$\frac{35-\mu}{\sigma} = -1.48$ { ∵ $x = 35$ is located to left of $x = \mu$ & consequently corresponding value of z are -ve

$\Rightarrow 35-\mu = -1.48\sigma$ —(3)

From (2)

$P(X < 63) = 0.89$

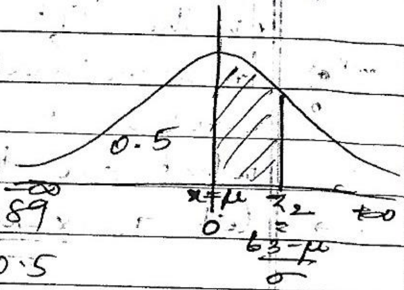
$P\left(\frac{X-\mu}{\sigma} < \frac{63-\mu}{\sigma}\right) = 0.89$

$P\left(Z < \frac{63-\mu}{\sigma}\right) = 0.89$

$0.5 + P\left(0 < Z < \frac{63-\mu}{\sigma}\right) = 0.89$

$P\left(0 < Z < \frac{63-\mu}{\sigma}\right) = 0.89 - 0.5$

$= \underline{0.39}$



From table, $\frac{63-\mu}{\sigma} = 1.23$ [$\because x = 63$ is located to right of $x = \mu$ & consequently corresponding value of z are +ve]

ie, $\frac{63-\mu}{\sigma} = 1.23$

$\Rightarrow 63 - \mu = 1.23\sigma$ —— (4)

Solving (3) & (4) we get;

$-28 = -2.71\sigma$

$\frac{28}{2.71} = \sigma$

$63 = 12.7$

(3) $\Rightarrow 35 - \mu = -1.48\sigma$

(4) $\Rightarrow 63 - \mu = 1.23\sigma$

(3)-(4) $\Rightarrow -28 = -2.71\sigma$

$\sigma = \frac{28}{2.71} = \underline{\underline{10.33}}$

$\sigma = 10.33$ in (4) $\Rightarrow$

$63 - \mu = 1.23 \times 10.33 = 12.7$

$63 - 12.7 = \mu \Rightarrow \mu = \underline{\underline{50.3}}$

Q. In a test on 2000 electric bulbs, it was found that the life of a particular bulb was normally distributed with an avg of 2040 hours and SD of 60 hours. Estimate no of bulbs likely to burn for

a) more than 2150 hours.

b) less than 1950 hours.

c) more than 1920 hours but less than 2160 hours.

$\rightarrow$ $Z = \frac{X-\mu}{\sigma}$; $\mu = 2040$; $\sigma = 60$.

a) $P(X > 2150) = P\left(\frac{X-\mu}{\sigma} > \frac{2150-\mu}{\sigma}\right)$

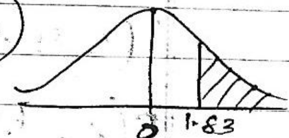
$= P\left(Z > \frac{2150-2040}{60}\right)$

$= P(Z > 1.83)$

$= 0.5 - P(0 < Z < 1.83)$

$= 0.5 - 0.4664$

$= 0.0336$



required no. of bulbs $= 0.0336 \times 2000$

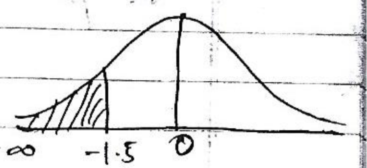
$= 67$

b) $P(X < 1950) = P(Z < -1.5)$

$= 0.5 - P(0 < Z < 1.5)$

$= 0.5 - 0.4332$

$= \underline{0.0668}$



required no of bulbs $= 0.0668 \times 2020$

$= 134$

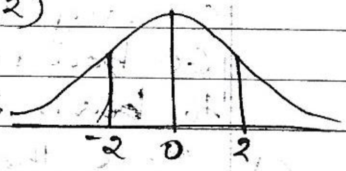
c) $P(1920 < x < 2160) = P(-2 < z < 2)$

$= P(-2 < z < 0) + P(0 < z < 2)$

$= 2\,P(0 < z < 2)$

$= 2 \times 0.4772$

$= 0.9544$



required no of bulbs $= 0.9544 \times 2020$

$= 1909$

9. In a normal distrib^n 5% of items are under 60 & 40% are b/w 60 & 65. find mean & SD of distrib^n.

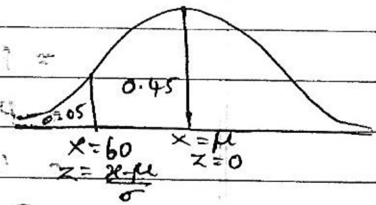
$\Rightarrow$ $P(x < 60) = 5\% = 0.05$ — ①

$P\left(\frac{x-\mu}{\sigma} < \frac{60-\mu}{\sigma}\right) = 0.05$

$P\left(z < \frac{60-\mu}{\sigma}\right) = 0.05$

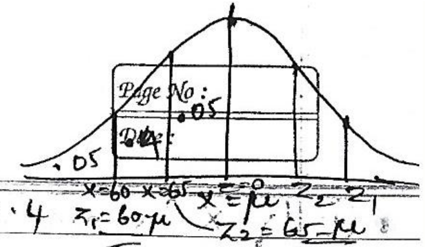
$0.5 - P\left(\frac{60-\mu}{\sigma} < z < 0\right) = 0.05$

$P\left(\frac{60-\mu}{\sigma} < z < 0\right) = 0.45$ — Ⓐ



from table, $\frac{60-\mu}{\sigma} = -1.65$ [The points x = 60 are located to left of $x = \mu$ & consequently corresponding value of z is -ve]

$\Rightarrow 60-\mu = -1.65\,\sigma$ — ②



$P(60 < x < 65) = 40\% = 0.4$

$P\left(\frac{60-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{65-\mu}{\sigma}\right) = 0.4$

$P\left(\frac{60-\mu}{\sigma} < z < \frac{65-\mu}{\sigma}\right) = 0.4$

$P\left(0 < z < \frac{60-\mu}{\sigma}\right) - P\left(0 < z < \frac{65-\mu}{\sigma}\right) = 0.4$

$\Rightarrow P\left(0 < z < \frac{65-\mu}{\sigma}\right) = P\left(0 < z < \frac{60-\mu}{\sigma}\right) - 0.4$

from Ⓐ $= 0.45 - 0.4 = 0.05$

$\Rightarrow P\left(0 < z < \frac{65-\mu}{\sigma}\right) = 0.05$

from table, we get

$\frac{65-\mu}{\sigma} = -0.12$ [$\because$ both points x = 60 & x = 65 are located to left to $x = \mu$ & consequently the corr. values of z are -ve

$\therefore 65-\mu = -0.12\,\sigma$ — ③

② $\Rightarrow$ $60-\mu = -1.65\,\sigma$

③ $\Rightarrow$ $65-\mu = -0.12\,\sigma$

② - ③ $\Rightarrow$ $-5 = -1.53\,\sigma$

$\Rightarrow \frac{5}{1.53} = \sigma \Rightarrow \sigma = 3.3$

$\sigma = 3.3$ in ② $\Rightarrow$

$60-\mu = -1.65 \times 3.3 = -5.445$

$\mu = 60 + 5.445 = 65.4$

1. If X is normally distributed with mean 4 and SD 3. find
a) $P(2.5 \leq X \leq 5.5)$ (b) $P(X \leq 7.9)$
2. At an examination 10% of the students got less than 30 marks and 97% got less than 62 marks. Assuming normal distrbⁿ, find σ & μ.

Module - 3.
CURVE FITTING.

⇒ Curve Fitting:-
Let (x_1, y_1) (x_2, y_2) ... (x_n, y_n) be 'n' pairs of observations. The general problem of finding a relation of the form $y = f(x)$, which fits best to given data is called Curve Fitting.

* The process of determining the best values of the parameters statistically is known as Curve Fitting and the principle that is most commonly used for curve fitting is Principle of Least Squares.

⇒ Principle of Least Squares:-
Let (x_1, y_1), (x_2, y_2) ... (x_n, y_n) be the given 'n' observations. Let $y = f(x)$ be a best fit curve. y_i is called the observed value of y corr. to x_i and $f(x_i)$ is called the expected value of y corr. to x_i.
Let $E_i = y_i - f(x_i)$ for $i = 1, 2 \ldots n$;
E_i is called the error or residual for y_i. The principle of least squares states that for a best fit curve, the sum of

Page No :
Date :

the squares of the residuals is a ~~maximum~~ Minimum

ie; $E = \sum_{i=1}^{n} [y_i - f(x_i)]^2$ is a minimum

① Fitting a straight line $y = a + bx$:-

Let $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ be 'n' observations. Our problem is to determine the best values of a & b.

Consider $y = a + bx$

Then $E = \sum_{i=1}^{n} [y_i - f(x_i)]^2$

$$= \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

$$= \sum_{i=1}^{n} (y_i - a - bx_i)^2$$ and it is a function of a & b.

By the principle of least squares the best value of a & b are obtained by solving the following eqns:

$\frac{\partial E}{\partial a} = 0$ and $\frac{\partial E}{\partial b} = 0$ [∵ E is min]

$$\frac{\partial E}{\partial a} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - a - bx_i)(-1) = 0.$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \quad \text{——①}$$

Page No :
Date :

$$\frac{\partial E}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - a - bx_i)(-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \cancel{(y_i x_i - a x_i - b x_i^2)} = 0.$$

$$\Rightarrow \sum_{i=1}^{n} [y_i x_i - a x_i - b x_i^2] = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \quad \text{——②}$$

Eqns ① & ② are called normal eqns. Solving these eqns we get the best values of a & b.

Q. By the method of least squares, find the best fitting straight line $y = a + bx$ to the data given below.

x	y	x^2	xy
5	15	25	75
10	19	100	190
15	23	225	345
20	26	400	520
25	29	625	750
75	113	1375	1880

$n = 5$

Given $y = a + bx$.

normal eqns $\Rightarrow \Sigma y = na + b\Sigma x$

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

ie; $5a + 75b = 113$ ——①

$75a + 1375b = 1880$ ——②

Solving ① & ② we get

$a = 11.5$; $b = 0.74$

$\therefore y = 11.5 + 0.74x$

Q. Fit a straight line of the form $y = ax+b$

x	y	x^2	xy
1	5	1	5
2	7	4	14
3	9	9	27
4	10	16	40
5	11	25	55
15	42	55	141

$n = 5$

$y = ax+b$.

$$\left.\begin{aligned}\Sigma y &= a\Sigma x + nb \\ \Sigma xy &= a\Sigma x^2 + b\Sigma x\end{aligned}\right\} \Rightarrow \begin{aligned}15a+5b&=42\\55a+15b&=141\end{aligned}$$

$a = 1.5$; $b = 3.9$

$\therefore y = 1.5x + 3.9$

② Fitting a Parabola $y = a+bx+cx^2$:-

$E = \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2)^2$ which is a func^n of a, b, c.

By principle of least squares the best value of a, b, c are obtained by solving the following eqns :-

$\frac{\partial E}{\partial a} = 0$; $\frac{\partial E}{\partial b} = 0$; $\frac{\partial E}{\partial c} = 0$.

Now $\frac{\partial E}{\partial a} = 0 \Rightarrow 2\Sigma(y_i - a - bx_i - cx_i^2)(-1) = 0$

$\Rightarrow \Sigma y_i = na + b\Sigma x_i + c\Sigma x_i^2$ —— ①

$\frac{\partial E}{\partial b} = 0 \Rightarrow 2\Sigma[y_i - a - bx_i - cx_i^2](-x_i) = 0$

$\Rightarrow \Sigma x_i y_i = a\Sigma x_i + b\Sigma x_i^2 + c\Sigma x_i^3$ —— ②

$\frac{\partial E}{\partial c} = 0 \Rightarrow 2[\Sigma y_i - a - bx_i - cx_i^2](x_i^2) = 0$

$\Rightarrow \Sigma x_i^2 y_i = a\Sigma x_i^2 + b\Sigma x_i^3 + c\Sigma x_i^4$ —— ③

Eqns ①, ②, ③ are called normal eqns. Solving these eqns we get the values of a, b, c.

Q. Fit a parabola by the method of least squares to the following data.

x	1	2	3	4	5
y	5	12	26	60	97

Let $y = a+bx+cx^2$ be the parabola of best fit. Then the normal eqns are:

$\Sigma y = na + b\Sigma x + c\Sigma x^2$ —— ①

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

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

① $\Rightarrow 5a + 15b + 55c = 200$

② $\Rightarrow 15a + 55b + 225c = 832$

③ $\Rightarrow 55a + 225b + 979c = 3672$

Page No :
Date :

x	y	x^2	x^3	x^4	xy	x^2y
1	5	1	1	1	5	5
2	12	4	8	16	24	48
3	26	9	27	81	78	234
4	60	16	64	256	240	960
5	97	25	125	625	485	2425
15	200	55	225	979	837	3672

Solving eqns, we get

$a = 10.3$; $b = -11$; $c = 5.7$

$\therefore y = 10.3 - 11x + 5.7x^2$ //

Q. Fit a second degree curve of the form $y = ax^2 + bx + c$ to the following data.

x	y	$X = x-1913$	X^2	X^3	X^4	Xy	X^2y
1911	10	-2	4	-8	16	-20	40
1912	12	-1	1	-1	1	-12	12
1913	8	0	0	0	0	0	0
1914	10	1	1	1	1	10	10
1915	14	2	4	8	16	28	56
		0	10	0	34	6	118

Let $X = x - 1913$

$\therefore y = aX^2 + bX + c$ —— (1)

normal eqns are :-

$\Sigma y = a\Sigma X^2 + b\Sigma X + nc$ —— (2)

$\Sigma Xy = a\Sigma X^3 + b\Sigma X^2 + c\Sigma X$ —— (3)

$\Sigma X^2y = a\Sigma X^4 + b\Sigma X^3 + c\Sigma X^2$ —— (4)

234
16
1276
24
336

314
25
155
62
775

Page No :
Date :

ie; $10a + 0b + 5c = 54$

$0a + 10b + 0c = 6$

$34a + 0b + 10c = 118$.

$\therefore a = 0.714$; $b = 0.6$; $c = 9.37$

(1) $\Rightarrow y = 0.714X^2 + 0.6X + 9.37$.

$y = 0.714(x-1913)^2 + 0.6(x-1913) + 9.37$ //

Q. Fit a second degree curve for the following data:

x	1	2	3	4	5
y	3	9	13	21	31

134 × 2
9
117

$\Sigma x = 15$; $\Sigma y = 77$; $\Sigma x^2 = 55$; $\Sigma x^3 = 225$

$\Sigma x^4 = 979$; $\Sigma xy = 299$; $\Sigma x^2y = 1267$; $n = 5$

x	y	x^2	x^3	x^4	xy	x^2y
1	3	1	1	1	3	3
2	9	4	8	16	18	36
3	13	9	27	81	39	117
4	21	16	64	256	84	336
5	31	25	125	625	155	775
15	77	55	225	979	299	1267

$n = 5$.

$y = a + bx + cx^2$

$5a + 15b + 55c = 77$

$15a + 55b + 125c = 299$

$55a + 125b + 979c = 1267$.

$a = 2.058$; $b = 0.744$; $c = 1.01$

Page No :
Date :

Q. Fit a second degree parabola :

x	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885
y	88	87	81	72	74	79	85	84	90	92	100

— Take $X = x - 1880$; $Y = y - 87$.

$y = 0.62X^2 + 1.18X + 79.01$; $X = x - 1880$.

Q. Convert the eqn $y = ax + bx^2$ to a linear form and write the corr. normal equations to fit it.

— The given eqn is $y = ax + bx^2$

∴ Dividing by 'x', we get

$\frac{y}{x} = a + bx$ —— ①

Put $y/x = Y$

∴ ① ⟹ $Y = a + bx$ which is linear in x and Y.

The normal eqns are:

$\Sigma Y = na + b\Sigma x$

$\Sigma xY = a\Sigma x + b\Sigma x^2$ //

Q. Use method of least squares to determine a and b ~~[illegible]~~ in the formula $y = ax + bx^2$ for the following data:

x	1	2	3	4	5
y	1.8	5.1	8.9	14.1	19.8

— Normal eqns are:

$\Sigma yx = a\Sigma x^2 + b\Sigma x^3$ $n = 5$

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

x	y	x^2	x^3	x^4	xy	x^2y
1	1.8	1				
2	5.1	4				
3	8.9	9				
4	14.1	16				
5	19.8	25				

Page No :
Date :

$\therefore y = 1.38x + 0.527x^2$

? Fit a parabola of the form $y = a + bx^2$ to the following data:

x	1	2	3	4	5
y	0.43	0.83	1.4	2.33	3.42

- Normal eqns are:

$\Sigma y = na + b\Sigma x^2$

$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3$

x	y	x^2	x^3	xy	x^2y
1	0.43	1	1	0.43	0.43
2	0.83	4	8		
3	1.4	9	27		
4	2.33	16	64		
5	3.42	25	125		

$\therefore y = 0.307 + 0.125x^2$ //

? Fit a curve of the form $y = a + bxy$ to the following data:

x	-4	1	2	3
y	4	6	10	8

Page No :
Date :

- Normal eqns are:-

$\Sigma y = na + b\Sigma xy$

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

x	y	x^2	y^2	xy	xy^2
-4	4				
1	6				
2	10				
3	8				

$a = 5.9$; $b = 0.128$.

$\therefore y = 5.9 + 0.128xy$ //

③ Fitting a Curve of the form $y = ae^{bx}$:-

- Given $y = ae^{bx}$

Taking log on both sides we get:

$\log_{10} y = \log_{10} a + bx \log_{10} e$

put $\log_{10} y = Y$; $\log_{10} a = A$; $b\log_{10} e = B$.

Then

$Y = A + Bx$ which is linear in x & Y

Normal eqns are:-

$\Sigma Y = nA + B\Sigma x$ —①

$\Sigma xY = A\Sigma x + B\Sigma x^2$ —②

Solving ① & ②, we get A & B.

$\therefore a = 10^{A}$

and $B = b \log_{10} e \Rightarrow b = \frac{B}{\log_{10} e}$

(ii) where $\log_{10} e = 0.4343$ //

Q. Fit $y = ae^{bx}$ to the following data:

x	y	$Y = \log y$	x^2	xy
0	5.012	0.7	0	0
2	10	1	4	2
4	31.62	1.45	16	5.8
6		3.15	20	7.8

– Given $y = ae^{bx}$

$\log_{10} y = \log_{10} a + bx \log_{10} e$

$\therefore Y = A + Bx.$

put $Y = \log_{10} y$

$A = \log_{10} a$

$B = b \log_{10} e$

$n = 3$

Normal eqns are:-

$\Sigma Y = nA + B\Sigma x$

$\Sigma xY = A\Sigma x + B\Sigma x^2$

ie; $3.15 = 3A + 6B$ —①

$7.8 = 6A + 20B$ —②

Solving ① & ②, we get

$A = 0.675$ and $B = 0.1875$

$\therefore \log_{10} a = 0.675$ & $b \log_{10} e = 0.1875$

$\therefore a = 10^{0.675} = 4.75$ //

$b = \frac{0.1875}{\log_{10} e} = \frac{0.1875}{0.4343} = 0.43$ //

$\therefore$ Required eqn is $y = 4.73\, e^{0.43x}$ //

(4) Fitting a Curve $y = ax^b$:-

Given $y = ax^b$.

Then $\log y = \log a + b \log x$ —①

Take $Y = \log y$; $A = \log a$; $X = \log x$.

$\therefore$ ① $\Rightarrow Y = A + bX$; This is linear in Y and X. Normal eqns are:-

$\Sigma Y = nA + b\Sigma X$ —②

$\Sigma XY = A\Sigma X + b\Sigma X^2$ —③

Solving ② & ③ we get A & b.

$a = \text{Antilog}(A)$.

Q. For a randomly selected observations, the following data were recorded.

Observn no	1	2	3	4	5	6	7	8	9	10
Overtime hours (X)	1	1	2	2	3	3	4	5	6	7
Additional unit (Y)	2	7	7	10	8	12	10	14	11	14

Determine the coefficients a, b, c and fit the parabola $y = a + bx + cx^2$.

– Normal eqns are:-

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

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

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

x	y	x^2	x^3	x^4	xy	x^2y
1	2	1	1	1	2	2
1	7	1	1	1	7	7
2	7	4	8	16	14	
2	10	4	8	16	20	
3	8	9	27	81	24	
3	12	9	27	81	36	
4	10	16	64	256	40	
5	14	25	125	625		
6	11	36			66	
7	14	49				
34	95	154	820	4774	377	1849

$n = 10$

$\therefore 10a + 34b + 154c = 95$

$34a + 154b + 820c = 377$

$154a + 820b + 4774c = 1849$

$a = 1.80;\ b = 3.48;\ c = -0.27$

$\therefore y = 1.8 + 3.48x - 0.27x^2$ //

9. Fit an exponential curve of the form $y = ab^x$ to the following data:

x	1	2	3	4	5	6	7	8
y	1.0	1.2	1.8	2.5	3.6	4.7	6.6	9.8

$\Rightarrow \quad y = 0.6821(1.38)^x$ //

$\Rightarrow$ Correlations and Regressions :-

In this section, we shall consider the methods that are suitable to study the connection, dependence or relation b/w 2 measured variables.

Let us consider two characteristics [variables] of the units of population. Let $(x_1, y_1)(x_2, y_2) \ldots (x_n, y_n)$ be 'n' pairs of values taken from the population, where x and y are the variables. Based on the data we want to know "How the variables related?" "How much they related?".

The study of the answer to the first question is Regression Analysis.

Regression Analysis gives the nature of relationship b/w the variables.

Answer to the second question gives the study of Correlation Analysis.

ie; Correlation gives the degree relationship b/w the variables.

Whenever two variables are related in such a way that a change in the one is followed directly or inversely by a

change in the other; they are said to be correlated. Such a data connecting 2 variables is called bivariate data.
If the 2 variables deviate in the same direction, ie; if the increase (or decrease) in one results corresponding increase (or decrease) in the correlation is said to be direct or positive.
eg:- Correlation b/w income and expenditure is positive.

If the ↑se or ↓se in one corresponds to the ↓se or ↑se in the other, the correlation is said to be negative. Here the 2 variables deviate in opposite direction.
eg:- The correlation b/w price and demand is negative.
Correlation is said to be perfect if the deviation in one variable is followed by a corresponding proportional deviation in the other. If there is no relationship indicated b/w the variables, they're said to be independent or uncorrelated.

⇒ Bivariate Distribution:-
Data containing 2 sets of measurements or obsrvns is called bivariate distbn.

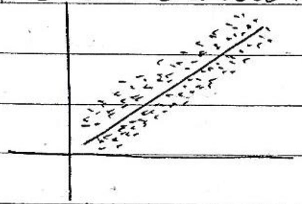
⇒ Scatter or Dot Diagrams:-

It is the simplest method of the diagramatic reprsn of bivariate data. Let (x_i, y_i), for $i = 1, 2 \ldots n$ be a bivariate distbn. Let the values of this variables x and y be plotted along this x-axis and y-axis on a suitable scale. Then corr. to every ordered pair, there corresponds a point or dot in the xy-plane. The diagrams of dots so obtained is called a dot or scatter diagrams.

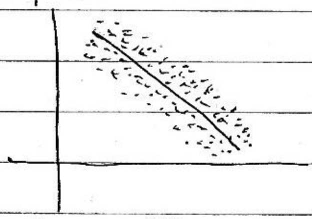
From this scatter diagrams, we can guess roughly how the variables x & y are correlated.

If the dots are very close to each other and the no. of obsrvns is not very large, a fairly good correlation is expected.

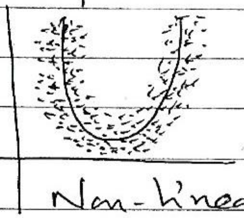
If the dots are widely seated a poor correlation is expected. If all the points in the scatter diagrams seem to be near a line, then there is corr^n b/w the variables and the corr^n is linear. If all the points seem to cluster round some curve such as parabola, the correlation is called non-linear.



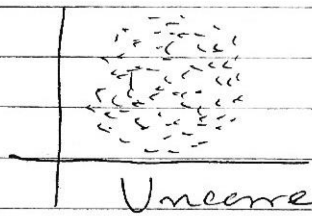
Linear
+ve correlation



linear
-ve correlation



Non-linear
correlation



Uncorrelated.

=> Karl Pearson's Coefficient of Corr^n:-

Correlation Coeff b/w 2 variables x & y, usually denoted by $r(x,y)$ or r_{xy} or simply r is defined by :

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$\boxed{r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}x_iy_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2-(\bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_i^2-(\bar{y})^2}}} \qquad \boxed{\begin{array}{l}\bar{x} = \sum_{i=1}^{n} x_i \\ \bar{y} = \sum_{i=1}^{n} y_i\end{array}}$$

and r_{xy} is called Karl Pearson's Coefficient Correlation b/w x & y.

* Note:-

$r_{xy} = \frac{P_{xy}}{\sigma_x \sigma_y}$ where P_{xy} or σ_{xy} is the covariance of x & y. $Cov(x,y) = P_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$

and $\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$ and

$$\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

* Schwartz's Inequality:-

$\sum_{i=1}^{n} A_i^2 \cdot \sum_{i=1}^{n} B_i^2 \geq \left(\sum_{i=1}^{n} A_iB_i\right)^2$, where