Aditya Kundu

21CS011004

Probability

Poission distribution :-

Let the random variable x is called poisson distribution with parameter μ with distribution.

X : 0 1 2 3 . . . . . . ∞

f_i : f_0 f_1 f_2 f_3 . . . . ∞

where, probability mass function $f_i = \dfrac{e^{-\mu}\mu^i}{i!}$, $i = 1, 2, 3, \dots, \infty$

① $$m = E(x) = \sum_{i=0}^{\infty} i f_i = \sum_{i=0}^{\infty} i \cdot \frac{e^{-\mu}\mu^i}{i!} = e^{-\mu}\sum_{i=1}^{\infty}\frac{\mu^i}{(i-1)!}$$

$$= e^{-\mu}\cdot\mu\sum_{i=1}^{\infty}\frac{\mu^{i-1}}{(i-1)!} = e^{-\mu}\cdot\mu\left[1 + \frac{\mu}{1!} + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \dots + \infty\right]$$

$$= e^{-\mu}\cdot\mu\cdot e^{\mu} = \mu e^0 = \mu$$

$$Var(x) = E(x(x-1)) - m(m-1)$$

$$= \sum_{i=0}^{\infty} i(i-1) f_i - \mu(\mu-1)$$

$$= \sum_{i=0}^{\infty} i(i-1)\frac{e^{-\mu}\mu^i}{i!} - \mu(\mu-1)$$

$$= e^{-\mu}\sum_{i=2}^{\infty}\frac{\mu^i}{(i-2)!} = \mu^2 \cdot e^{-\mu}\sum_{i=2}^{\infty}\frac{\mu^{i-2}}{(i-2)!} - \mu(\mu-1)$$

$$= \mu^2 \cdot e^{-\mu}\left[1 + \frac{\mu}{1!} + \frac{\mu^2}{2!} + \dots\right] - \mu(\mu-1)$$

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

⊛ Special Case :-

Let the random variable $X(t)$ denote the number of changes in a certain time interval $(0,t)$ the probability mass function is defined as,

$$P(X(t)=i) = f_i = \frac{e^{-\lambda t} \cdot (\lambda t)^i}{i!}, \quad i = 1, 2, 3 \ldots \infty$$

$\boxed{\mu = \lambda t}$

Where λ is the no of changes per unit time.

Q. A radio active source emits on average 2.5 particle per sec. Calculate the probability that 2 or more particle will be emitted in one interval of 4 seconds.

Soln: λ = no of changes emits the particle on the average per unit time = 2.5

Let $X(t)$ be the random variable defined as no. of changes on the average on the time interval $(0,t)$

$\therefore \lambda t = 2.5 \times 4 = 10$

$$f_i = \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

$$\therefore P(X \geq 2) = 1 - P(X<2)$$
$$= 1 - P(X=0) - P(X=1)$$
$$= 1 - e^{-10} \cdot \frac{10^0}{0!} - e^{-10} \cdot \frac{10^1}{1!}$$

Limiting Case of Binomial distribution :-

Binomial $\downarrow$ n finite $\downarrow$ Large

$X \sim B(n, p)$

$f_i = {}^nC_i \, p^i (1-p)^{n-i}$

$\lim_{n\to\infty} = {}^nC_i \, p^i (1-p)^{n-i}$

$= \frac{e^{-np} (np)^i}{i!} = \frac{e^{-\mu} \cdot \mu^i}{i!}$

Q. Six coins were tossed 6400 times. Using Poisson distribution find the approximate probability of getting 6 heads 8 times.

Soln : Let the random variable X is defined as no of getting 6 heads in the toss of 6 coins.

$p = \left(\frac{1}{2}\right)^6$, $n = 6400$, $np = \frac{1}{2^6} \times 6400 = 100$

$P(X=8) = {}^{6400}C_8 \left(\frac{1}{2^6}\right)^8 \cdot \left(1 - \frac{1}{2^6}\right)^{6392} \simeq \frac{e^{-100} \cdot (100)^8}{8!}$

Continuous distribution :-

- Probability density function (PDF)

Let x be a continuous random variable with distribution $F(x) = P(-\infty < X \leq x)$. Then PDF is defined as

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

$\Rightarrow F'(x) = f(x) \geq 0$

$F(-\infty) = 0$,

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

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

- **Properties** :-

$$P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b)$$
$$= P(a < x < b)$$
$$= \int_a^b f(x)\,dx = F(b) - F(a)$$
$$F(b) = \int_{-\infty}^{b} f(x)\,dx$$
$$F(a) = \int_{-\infty}^{a} f(v)\,dv$$

14/Sep/2022

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

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

$$\sum_{i=0}^{n} i f i$$

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

Example: Q. The PDF of X is $f(x) = K(x-1)(2-x)$, for $1 \leq x \leq 2$

① Determine the value of K

② The distribution function

③ $P(5/4 \leq x \leq 3/2)$ & $P(5/4 < x < 3/2)$

Soln: (1) We know that $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\Rightarrow \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{2} K(x-1)(2-x)\,dx + \int_{1}^{\infty} f(x)\,dx = 1$$

$$\Rightarrow K\int_{1}^{2} (3x - x^2 - 2)\,dx = 1$$

$$\Rightarrow K\left[\frac{3x^2}{2} - \frac{x^3}{3} - \frac{2x}{1}\right]_1^2 = 1$$

$$\Rightarrow K\left[\left(3.\frac{4}{2} - \frac{8}{3} - 4\right) - \left(\frac{3}{2} - \frac{1}{3} - 2\right)\right] = 1$$

$$\Rightarrow K = 6 \quad \text{[Ans:-]}$$

(2) $F(x) = 0 \qquad , \quad x < 1$

$$= \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{x} f(x)\,dx, \qquad 1 \leq x < 2$$

$$= 0 + \int_{1}^{x} 6(x-1)(2-x)\,dx$$

$$= 6\int_{1}^{x} (3x - x^2 - 2)\,dx$$

$$= 6\left[\frac{3}{2}x^2 - \frac{x^3}{3} - 2x\right]_1^x$$

$$= 6\left[\left(\frac{3}{2}x^2 - \frac{x^3}{3} - 2x\right) - \left(\frac{3}{2} - \frac{1}{3} - 2\right)\right]$$

$$F(x) = \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx, \quad x \geq 2$$

$$= 0 + 1 + 0 = 1$$

③ $P(5/4 < x < 3/2)$

$= P(5/4 \leq x \leq 3/2)$

$= 11/32$

Q. X is a continuous random variable having pdf, $f(x) = 4x/5$; $0 < x \leq 1$

$= 2/5(3-x)$, $1 < x \leq 2$

$= 0$, elsewhere.

Find E(x).

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

$= \int_0^1 x \frac{4x}{5}\, dx + \int_1^2 x \frac{2}{5}(3-x)\, dx$

$= \left| \frac{4}{5} \cdot \frac{x^3}{3} \right|_0^1 + \frac{2}{5}\left[\frac{3x^2}{2} - \frac{x^3}{3}\right]_1^2 = \frac{17}{15}$

Q. Find mean and variance of X having pdf $f(x) = |-1|-x|$, $0 < x < 2$

$= 0$, elsewhere.

$|x| = x$, $x > 0$

$= -x$, $x < 0$

$= 0$, $x = 0$

$|1-x| = 1-x$, $0 < x < 1$

$= -(1-x)$, $x < 0$

$= 0$, $x = 1$

Soln: $m = E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$

$= \int_0^2 x(1-|1-x|)\, dx$

$= \int_0^1 x(1-(1-x))\, dx + \int_1^2 x(1+(1-x))\, dx$

$= 1$

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

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

$= \int_0^1 x^2(1-(1-x))\, dx + \int_1^2 x^2(1+(1-x))\, dx - 1$

$= \frac{7}{6} - 1 = \frac{1}{6}$

Q. Find the value of constant K such that $f(x) = Kx(1-x),\ 0 < x \leq 1$
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = 0$, elsewhere.
is a pdf of x. Compute $E(x)$, $Var(X)$ & $P(x > 1/2)$

Soln:

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

$$\Rightarrow \int_0^1 f(x)\,dx \Rightarrow \int_0^1 Kx(1-x)\,dx = 1$$

$$\Rightarrow K \int_0^1 (x - x^2)\,dx = 1$$

$$\Rightarrow K\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 1$$

$$\Rightarrow K\left[\frac{1}{2} - \frac{1}{3}\right] = 1$$

$$\Rightarrow K = 6$$

$P(x > 1/2)$

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

$= 1 - P(-\infty < x \leq 1/2)$

$= 1 - \int_{-\infty}^{1/2} f(x)\,dx$

$= 1 - \int_0^{1/2} 6x(1-x)\,dx$

$= 1 - 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^{1/2}$

$= 1 - 6\left[\frac{1}{8} - \frac{1}{24}\right] = 1 - 6\left(\frac{2}{24}\right) = 1 - \frac{1}{2} = \frac{1}{2}$

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

$$= \int_{0}^{-\infty} 6x(1-x)\,dx = ?$$

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

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

$$= \int_{0}^{1} x^2\, 6x(1-x)\,dx - (E(X))^2$$

21/Sep/2022

- **Normal distribution :-**

Let X be a continous random variable having normal distribution $N(m, \sigma)$, if the density function is given by,

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

(i) $f(x) \geq 0$ (ii) $\int_{-\infty}^{\infty} f(x)\,dx$

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

$$= 1$$

L.H.S $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-m)^2}{2\sigma^2}} \cdot dx$

let, $z = \frac{(x-m)^2}{2\sigma}$

$z^2 = t$

$2z\,dz = dt$

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

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

$$\int_{-a}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx = 0$$

$$= 2\int_{0}^{\infty} \frac{1}{\sqrt{\pi}} \frac{dt}{2\sqrt{t}} \cdot e^{-t}\, dt$$

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

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

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

$$\Rightarrow \Gamma\tfrac{1}{2} = \int_{0}^{\infty} e^{-x} x^{-1/2}\, dx$$

$$= \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi}$$

$= 1$ [R.H.S] Proved

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

$$\Gamma n = (n-1)!$$

When n is a integer.

Standard Normal Distribution :-

$$\boxed{X \sim N(0,1)}$$

When x is a normal distribution function with parameter 0 and 1. ie, $X \sim N(0,1)$, then X is called standard normal distribution.

If, $X \sim N(m, \sigma)$

$m = E(X) = m$

$Var(X) = \sigma^2$

Standard division $= \sqrt{Var(X)} = \sigma$

Solve the problem of normal distribution from standard normal distribution

Let, $X \sim N(m, \sigma)$ and $Z = \frac{X-m}{\sigma}$, then $Z \sim N(0,1)$

$\Rightarrow$ Since, $X \sim N(m, \sigma)$

So, $E(X) = m$, $Var(X) = \sigma^2$

$$E(ax+b) = a E(x) + b$$

$$E(z) = E\left(\frac{x-m}{\sigma}\right)$$

$$= \frac{1}{\sigma}[E(x) - m]$$

$$= \frac{1}{\sigma}(m-m)$$

$$\boxed{E(z) = 0}$$

$$Var(z) = Var\frac{(x-m)}{\sigma}$$

x $\to$ $Var(x) = E(x-m)^2$

$\therefore Var(x) = E(x-m)^2$

$$= E\left\{\left(\frac{x-m}{\sigma} - 0\right)^2\right\}$$

$$= E\left(\frac{x-m}{\sigma}\right)^2$$

$$= \frac{1}{\sigma^2} E(x-m)^2$$

$$= \frac{1}{\sigma^2} Var(x) = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$

$\therefore$ $\underline{Var(z) = 1}$

Important!

REMEMBER

Q. The length of bolts produced by a machine is normally distributed with mean 4 and S.D 0.5. A bolt is defective if its length does not lie in the interval (3.8, 4.3). Find the percentage of defective bolts produced by machine.

Soln:

$$\text{Given, } \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0.6} e^{-t^2/2}\, dt = 0.7257$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0.4} e^{-t^2/2}.\, dt = 0.6554$$

$$X \sim N(0,1)$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

Let x be the random variable defined as "the length of the bolts".

$$X \sim N(4, 0.5), \quad Z = \frac{X-4}{0.5}$$

$X = 4.3$

$$Z = \frac{4.3-4}{0.5} = 0.6$$

$$Z = \frac{3.8-4}{0.5} = -0.4$$

$$P(3.8 < x < 4.3)$$

$$= P(-0.4 < Z < 0.6)$$

$$= \int_{0.4}^{0.6} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

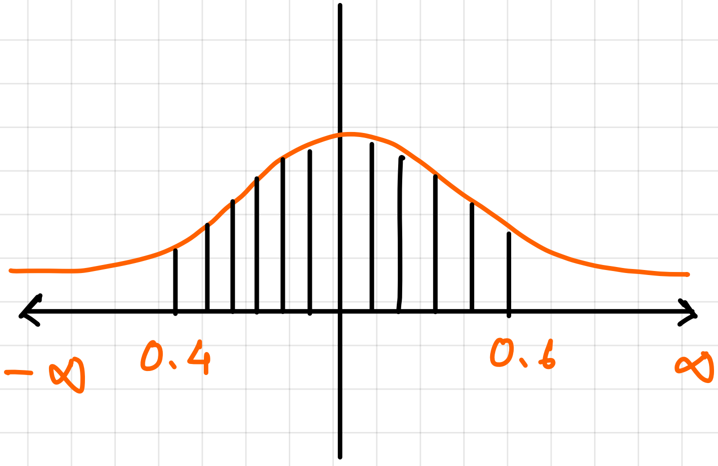
$$= \int_{-\infty}^{0.6} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt - \int_{-\infty}^{-0.4} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.\, dt$$

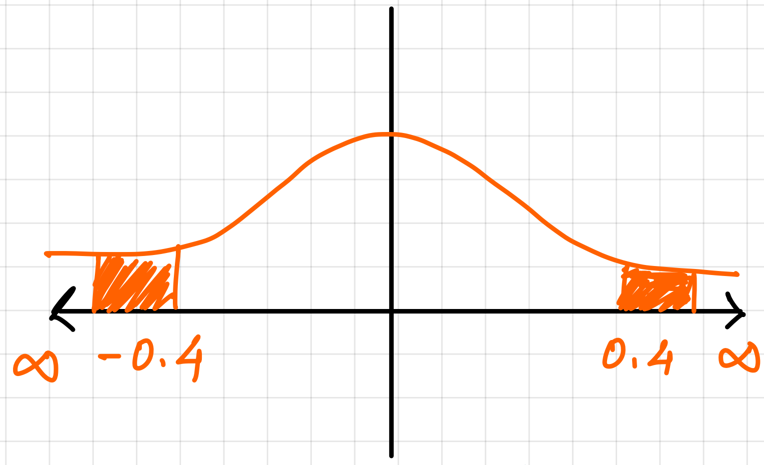
$$P(a < x < b) = \int_a^b f(x)\, dx$$

$$= 0.7257 - \int_{0.4}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

$$= 0.7257 - \left(1 - \frac{1}{\sqrt{2\pi}} \int_{\infty}^{0.4} e^{-t^2/2}.\, dt\right)$$

$$= 0.7257 - (1 - 0.6554) = 0.3811$$





So, the probability of the defective bolt is $(1-0.3811) = 0.6189$

∴ Thus percentage of defective bolt,

$$= 0.6189 \times 100 = 61.89\,\%$$ Ans

- Exponential Distribution :-

Let x be a continuous random variable, having exponential distribution if Pdf is given by $f(x) = \lambda e^{-\lambda x}$, $(x>0)$

$=0$ elsewhere

$$X \sim e(\lambda)$$

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

$$= \int_0^{\infty} x \cdot \lambda e^{-\lambda x}\, dx$$

$$= \lambda \int_0^{\infty} x \cdot e^{-\lambda x}\, dx = \int_0^{\infty} x \cdot \frac{e^{-\lambda x}}{-\lambda} - \int_0^{\infty} \frac{e^{-\lambda x}}{-\lambda}\, dx$$

$$= x \frac{e^{-\lambda x}}{-\lambda} + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x}\, dx$$

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

Gamma function

$$= \int_0^{\infty} \frac{z}{\cancel{\lambda}} \cdot \cancel{\lambda} \cdot e^{-z} \cdot \frac{dz}{\lambda}$$

Let,

$\lambda x = z$

$\lambda dx = dz$

$\Gamma n = \int_0^{\infty} e^{-x} \cdot x^{n-1} dx$

$\Gamma n = (n-1)!$

$$= \frac{1}{\lambda} \int_0^{\infty} e^{-z} \cdot z\, dz$$

$$= \frac{1}{\lambda} \int_0^{\infty} e^{-z} \cdot z^{2-1}\, dx$$

$$= \frac{1}{\lambda} \Gamma 2 = (2-1)! \times \frac{1}{\lambda} = \frac{1}{\lambda}$$

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

$$= \left(\int_{-\infty}^{\infty} x^2 f(x)\,dx\right) - \left(\frac{1}{\lambda}\right)^2$$

$$= \int_0^{\infty} x^2 f(x)\,dx - \left(\frac{1}{\lambda}\right)^2$$

$$= \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx - \left(\frac{1}{\lambda}\right)^2 \quad \Big| \; \lambda x = z, \; \lambda dx = dz$$

$$= \int_0^{\infty} x^2 \lambda e^{-x\lambda}\,dx - \left(\frac{1}{\lambda}\right)^2$$

$$= \int_0^{\infty} \frac{z^2}{\cancel{\lambda^2}} \cancel{\lambda} e^{-z} \frac{dz}{\lambda}$$

$$= \frac{1}{\lambda^2}\int_0^{\infty} z^2 e^{-z}\,dz = \frac{1}{\lambda^2}\int_0^{\infty} e^{-z} z^2\,dz = \frac{1}{\lambda^2}$$

$$L_3^3 = 1$$
$$3!$$

(i) $P(x > 10) = \int_{10}^{\infty} \frac{1}{10} e^{-x/10}\,dx$

$$= \frac{1}{10}\int_{10}^{\infty} e^{-x/10}.dx$$

$$= \frac{1}{10}\int_{10}^{\infty} \frac{e^{-x/10}}{-1/10}\,dx = -1 \lim_{p\to\infty}\left[e^{-x/10}\right]_{10}^{p} dx$$

$$= -\lim_{p\to\infty}\left[e^{-p/10} - e^{-1}\right]$$

$$= 0 + e^{-1} = 1/e$$

(ii) $P(10 < x < 20) = \frac{1}{10}\int_{10}^{20} \frac{e^{-x/10}}{-1/10}\,dx$

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

Q. 10 cards are drawn from a well shuffled pack of cards. Find the probability at least one of them to be a diamond.

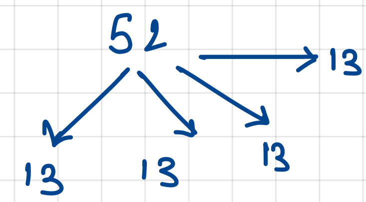
Soln:-

<u>No diamond</u>

$$\frac{^{39}C_{10}}{^{52}C_{10}}$$

<u>1 diamond</u>

$$1-\frac{^{39}C_{10}}{^{52}C_{10}}$$



Q. Let x be normally distributed with mean 0 & unit variance. Find $E(x)^2$.

Soln: $m = E(x) = 0$, $Var(x) = 1$

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

$$\Rightarrow E(x^2) = 1 - 0 = 1$$

STATISTICS

<u>02/Nov/2022</u>

Defination: An aggregate of fact which are affected by a no. of causes/ reasons and which are expressed as numerically to the reasonable extent of accuracy and which are collected in a systematic manner for specific purpose is called statistics.

- **Variable:-**

A variable is a symbol. eg. x or y which assume every prescribed value in a set.

- **Data/Observation :-**

A value assumed by any variable is called data or observation.

- **Frequency :** Frequency is the occurance of number of observations.

X :	2	3	4	5	6
f :	2	5	10	20	5

- **Group Frequency distribution :-**

X :	40-50 kg	51-60 kg	61-70 kg
Y :	10	50	60

$\Rightarrow$ **Some terms associated with group frequency distribution :-**

(I) Class interval

(II) Class limits → Lower class limit (LCL) ; Upper class limit (UCL)

(III) Class Boundaries → lower class boundary $= LCL - \frac{d}{2}$; upper class boundary $= UCL + \frac{d}{2}$

{ where, d is the difference between two class intervals.

$\therefore$ d = UCL of next class interval − UCL of previous class interval. }

(IV) Mid Value : $\frac{1}{2}(LCL + UCL)$

(V) Width of Class : $UCB - LCB$

Arithmetic Mean (A.M) :-

$$x_1, x_2, x_3 \longrightarrow x_n$$

then, $$A.M = \frac{x_1 + x_2 + x_3 + \_\_\_ + x_n}{\text{no of observation}}$$

$$= \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

X : x_1 x_2 x_3 --- x_n

f : f_1 f_2 f_3 ---- f_n

$$A.M = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \_\_\_ + x_n f_n}{f_1 + f_2 + f_3 + \_\_\_ f_n}$$

$$= \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

Arithmatic Mean (A.M) of group frequency distribution :-

$$A.M = \frac{\overline{x_1} f_1 + \overline{x_2} f_2 + \dots + \overline{x_n} f_n}{f_1 + f_2 + \_\_\_ + f_n}$$

$\overline{x_i}$ = mid value of class interval

Eg :-

X :	1	2	3	4	5	6
f :	2	3	5	10	2	3

$$A.M = \frac{1\times2 + 2\times3 + 3\times5 + 4\times10 + 5\times2 + 6\times3}{2+3+5+10+2+3} = \frac{91}{25}$$

eg.

X:	1–10	11–20	21–30	31–40
f:	2	3	5	10

$$A.M = \frac{\frac{(1+10)}{\cancel{2}} \times \cancel{2} + \frac{(11+20)}{2} \times 3 + \frac{(21+30)}{2} \times 5 + \frac{(31+40)}{\cancel{2}} \times \cancel{10}^{5}}{2+3+5+10}$$

$$= \frac{11 + 46.5 + 127.5 + 355}{20} = 27$$

- **Result** :–

If x and y are two variable and $y = \frac{x-c}{d}$, where c and d are constants, then $\bar{y} = \frac{\bar{x}-c}{d}$, $\bar{x}, \bar{y}$ are A.M of distribution.

- **Median and Mode** :–

→ Median : Median is the middle most value of the observation which are arranged in increasing order or decreasing order in magnitude.

→ Mode : Mode is the observation having maximum frequency.

X:	3	5	7	10	12
f:	2	1	3	2	4

∴ Mode = 12 (observation with max freq.)

X	f	C.F (Cumulative frequency)
3	2	2
5	1	3
7	3	6
10	2	8
12	4	12

N = total frequency = 12

Median : Corresponding observation of $\frac{N+1}{2}$ in cumulative frequency.

$$\frac{N+1}{2} = \frac{12+1}{2} = \frac{13}{2} = 6.5$$

If not equal to any term in CF, then we consider next observation of $\frac{N+1}{2}$ in CF,

Next CF value of 6.5 is 8, then median = 10

Suppose you have observations without frequency.

X : 1 2 3 4 5 6 7

If we have odd no. of observations,

Then median $= \left(\frac{n+1}{2}\right)^{th}$ observation. $= \left(\frac{7+1}{2}\right) =$ 4th observation

$= 20$

Then if we have even no. of observations,

Then median $= \frac{1}{2}\left[\frac{n}{2}^{th} \text{ observation} + \left(\frac{n}{2}+1\right)^{th} \text{ observation}\right]$

$= \frac{1}{2}\left[4^{th} \text{ observation} + 5^{th} \text{ observation}\right]$

$= \frac{1}{2}\left[15 + 20\right] = 17.5$

- **For grouped frequency distribution :-**

$$\text{Median} = LM + \frac{N/2 - F}{f_m} x_i$$

Where,

LM = lower boundary of median class.

N = Total frequency

F = Cumulative frequency of class preceding to the median class.

f_m = frequency of median class

i = width of median class.

$$\text{Mode} = LM + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Where,

Lm = lower boundary of modal class

f_1 = frequency of the modal class

f_0 = frequency of the class preceding modal class.

f_2 = frequency of class succeeding the modal class.

i = width of each class.

- **Relation b/w medial & modal :-**

Mean - Mode = 3 (Mean - Median)

Calculate the mean, median and mode :- 04/Nov/2022

Score	Frequency (f)	C.F	MidValue (x)	Class Boundary
10 – 14	3	3	12	9.5 – 14.5
15 – 19	9	12	17	14.5 – 19.5
20 – 24	6	18	22	19.5 – 24.5
25 – 29	12	30	27	24.5 – 29.5
30 – 34	10	40	32	29.5 – 34.5
35 – 39	7	47	37	34.5 – 39.5
40 – 44	3	50	42	39.5 – 44.5

(N) = total frequency = 50

$y = \frac{x-27}{5}$	fy	fy^2
−3	−9	27
−2	−18	36
−1	−6	6
0	0	0
1	10	10
2	14	28
3	9	27

$\Sigma fy = 0$; $\Sigma fy^2 = 134$

$\therefore$ **Mean of y** $= \frac{\Sigma fy}{\Sigma f}$

$= \frac{0}{50} = 0 = \bar{y}$

So, $\bar{y} = \frac{\bar{x} - c}{d}$

$\Rightarrow 0 = \frac{\bar{x} - 27}{5} \Rightarrow \bar{x} = 27$

Median $= L_m + \frac{N/2 - F}{f_m} \times i$

$= 24.5 + \frac{50/2 - 18}{12} \times 5 = 27.42$

Mode $= L_m + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$

$= 24.5 + \frac{12 - 10}{2 \times 12 - 6 - 10} \times 5 = 24.5 + \frac{\cancel{2}}{\cancel{8}4} \times 5 = 25.75$

Remember

For group frequency distribution median class is $N/2$th position in CF or next higher position in CF. And modal class is the class with maximum frequency.

And, Standard Deviation, SD of $y = \sqrt{var(y)}$

$$= \sqrt{\frac{\sum f_i (y-\bar{y})^2}{\sum f}}$$

$$= \sqrt{\frac{\sum f y^2}{\sum f} - \left(\frac{\sum f y}{\sum f}\right)^2} = \sqrt{\frac{134}{50} - 0}$$

$$= \sqrt{\frac{134}{50}}$$

$$= 1.64$$

SD of $x = \sigma x$

$$= \sigma y \times d = \sqrt{\frac{134}{50}} \times 5$$

$$= 1.64 \times 5 = 8.2$$

Q. The given distribution is — 09/Nov/2022

Class interval :	0–10	10–20	20–30	30–40	40–50
Frequency :	3	α	20	12	β

If median and mode are 27 and 26. Find α and β.

Soln: Since, the mode is 26.

$\therefore$ Modal class = 20 - 30

$\therefore$ Mode $= 26 = LM + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$

$$\Rightarrow 26 = 20 + \frac{20 - \alpha}{40 - \alpha - 12} \times 10$$

$$\Rightarrow 26 = 20 + \frac{20 - \alpha}{28 - \alpha} \times 10$$

$$\Rightarrow 6 = \frac{10(20 - \alpha)}{28 - \alpha}$$

$$\Rightarrow 168 - 6x = 200 - 10\alpha \Rightarrow 4\alpha = 32$$

$$\Rightarrow \alpha = 8$$

Class interval	frequency	CF
0 – 10	3	3
10 – 20	8	11
20 – 30	20	31
30 – 40	12	43
40 – 50	B	43+B=N

Median = 27

$$\Rightarrow 27 = LM + \frac{N/2 - F}{f_m} \times i$$

$$\Rightarrow 27 = \frac{\frac{43+B}{2} - 11}{\cancel{20}} \times \cancel{10}$$

$$\Rightarrow 7 = \frac{43+B-22}{4}$$

$$\Rightarrow 7 = \frac{21+B}{4}$$

$$\Rightarrow B = 7$$

Q. Class interval : 30-39 40-49 50-59 60-69 70-79 80-89 90-99

Frequency : 2 3 11 20 x 25 7

If A.M is 72.5. Then find the value of x.

Soln :

$$\overline{x_1} f_1 = \frac{(30+39)}{2} \times 2 = 69$$

$$\overline{x_2} f_2 = \frac{40+49}{2} \times 3 = 133.5$$

$$\overline{x_3} f_3 = \frac{50+59}{2} \times 11 = 599.5$$

$$\overline{x_4} f_4 = \frac{60-69}{2} \times 20 = 1290$$

$$\overline{x_5} f_5 = \frac{70-79}{2} \times x = 74.5x$$

$$\overline{x_6} f_6 = \frac{80+89}{2} \times 25 = 2112.5$$

$$\overline{x_7} f_7 = \frac{99+90}{2} \times 7 = 661.5$$

$$AM = \frac{69 + 133.5 + 599.5 + 1290 + 74.5x + 2112.5 + 661.5}{68 + x}$$

$$\Rightarrow 72.5 = \frac{4866 + 74.5x}{68 + x}$$

$$\Rightarrow 4930 + 72.5x = 4866 + 74.5x$$

$$\Rightarrow 2x = 64 \Rightarrow \boxed{x = 32}$$

TO DO

Q. Class interval : 3-5 5-7 7-9 9-11 11-13 total

frequency : 32 p 57 q 25 200

If A.M = 7.74. Find the missing term.

Soln :

$\bar{x}_1 f_1 = \frac{3+5}{2} \times 32 = 128$

$\bar{x}_2 f_2 = \frac{5+7}{2} \times p = 6p$

$\bar{x}_3 f_3 = \frac{7+9}{2} \times 57 = 456$

$\bar{x}_4 f_4 = \frac{9+11}{2} \times q = 10q$

$\bar{x}_5 f_5 = \frac{11+13}{2} \times 25 = 300$

$32 + p + 57 + q + 25 = 200$

$\Rightarrow p + q + 114 = 200$

$\Rightarrow p + q = 200 - 114$

$\Rightarrow q = 86 - p$

$7.74 = \frac{128 + 6p + 456 + 10q + 300}{200}$

$\Rightarrow 7.74 = \frac{884 + 6p + 10q}{200}$

$\Rightarrow 1548 = 884 + 6p + 10q$

$\Rightarrow 6p + 10q = 1548 - 884$

$\Rightarrow 6p + 10(86 - p) = 664$

$\Rightarrow 6p + 860 - 10p = 664$

$\Rightarrow 4p = 196 \Rightarrow \boxed{p = 49}$

$\therefore \boxed{q = 86 - 49 = 37}$

TO DO

Q. Find the mean and median.

Marks		No of students (CF)	Class interval	frequency
Below	10	3	0 − 9	3
"	20	8	10 − 19	8 − 3 = 5
"	30	17	20 − 29	17 − 8 = 9
"	40	20	30 − 39	20 − 17 = 3
"	50	22	40 − 49	22 − 20 = 2

• Skewness & Kurtosis :-

r^{th} central moment about mean of a distribution

$$= m_r = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^r,$$

Where, $\bar{x}$ = A.M of distribution of n term

f_i is the corresponding frequency.

N = Total frequency.

$$m_2 = \text{Variance} = 1/N \sum_{i=1}^{n} f_i (x_i - \bar{x})^2$$

$$m_3 = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^3, \quad m_4 = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^4$$

∴ Skewness $= \frac{m_3}{\sigma^3}$, where $\sigma = SD = \sqrt{Var(x)} = \sqrt{m_2}$

Kurtosis $= \frac{m_4}{\sigma^4} - 3$

Skewness $= \frac{\text{Mean} - \text{mode}}{\text{Standard deviation}}$

$\Rightarrow \quad = \frac{3(\text{mean} - \text{median})}{\text{Standard deviation}}$

16/Nov/2022

Covariance and Corelation Coefficient :-

$X : \quad x_1 \quad x_2 \quad x_3 \text{ ——— } x_n$

$Y : \quad y_1 \quad y_2 \quad y_3 \text{ ——— } y_n$

Definition : Covariance of x, y = $\text{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i y_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\bar{y}\right)$$

$\left\{ \begin{array}{l} \bar{x} = \text{Arith. mean of } x \\ \bar{y} = \text{Arith. mean of } y \end{array} \right\}$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{y}\frac{1}{n}\sum_{i=1}^{n} x_i - \bar{x}\frac{1}{n}\sum_{i=1}^{n} y_i + \bar{x}\bar{y}\frac{1}{n}\sum_{i=1}^{n} 1$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \cancel{\bar{y}\bar{x}} - \bar{x}\bar{y} + \cancel{\bar{x}\bar{y}} \cdot \frac{1}{n} \cdot n$$

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

Definition : Corelation coefficient of x, y = $r_{x,y} = \dfrac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$,

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

$\left\{ \begin{array}{l} \sigma_x = \text{S.D of } x \\ \sigma_y = \text{SD of } y \end{array} \right\}$

Result : Suppose X and y are two observation and if $u = ax + b$ and $v = cy + d$, then $r_{uv} = \dfrac{ac}{|a||c|} r_{x,y}$

Result : $\boxed{-1 \leq r_{x,y} \leq 1}$

Q. Find the corelation coeff. between two observation

x: 65 66 67 67 68 69 70 72

y: 67 68 65 68 72 72 69 71

Soln.:

x_i	y_i	$u_i = x_i - 67$	$v_i = y_i - 67$	$u_i v_i$	u_i^2	v_i^2
65	67	-2	-1	2	4	1
66	68	-1	0	0	1	0
67	65	0	-3	0	0	9
67	68	0	0	0	0	0
68	72	1	4	4	1	16
69	72	2	4	8	4	16
70	69	3	1	3	9	1
72	71	5	3	15	25	9
		$\sum u_i = 8$	$\sum v_i = 8$	$\sum u_i v_i = 32$	$\sum u_i^2 = 44$	$\sum v_i^2 = 52$

$$Cov(u,v) = \frac{1}{8}\sum_{i=1}^{8} u_i v_i - \bar{u}\bar{v}$$

$$= \frac{1}{8}\cdot 32 - \left(\frac{8}{8}\cdot\frac{8}{8}\right) = 4-1 = 3$$

$$\sigma_u^2 = \frac{1}{8} \times 44 - \left(\frac{8}{8}\right)^2 = \frac{1}{\cancel{8}_2} \times \cancel{44}^{11} - 1 = 4.5$$

$$\therefore \sigma_u = \sqrt{4.5}$$

$$\sigma_v^2 = \frac{52}{8} - \left(\frac{8}{8}\right)^2 = \frac{52}{8} - 1 = 5.5 \quad \therefore \sigma_v = \sqrt{5.5}$$

$$\rho_{u,v} = \frac{Cov(u,v)}{\sqrt{4.5} \times \sqrt{5.5}} = 0.60,$$

$$\rho_{uv} = \frac{ac}{|a||c|}\rho_{xy} = \frac{1\cdot 1}{|1||1|}\rho_{xy} \Rightarrow \rho_{uv} = \rho_{xy}$$

$$\therefore \rho_{xy} = 0.60$$

Ans

Q. If $u+3x=5$ and $2y-v=7$ and $r_{xy}=0.12$ then find r_{uv}.

Soln: $u=-3x+5$ $\quad\quad$ $v=2y-7$

$\Rightarrow v = 2y+(-7)$

Here, $a=-3$, $c=2$

$$r_{u,v} = \frac{ac}{|a||c|}\, r_{xy}$$

$$= \frac{(-3)(2)}{(3\times 2)} \times 0.12 = -1\times 0.12$$

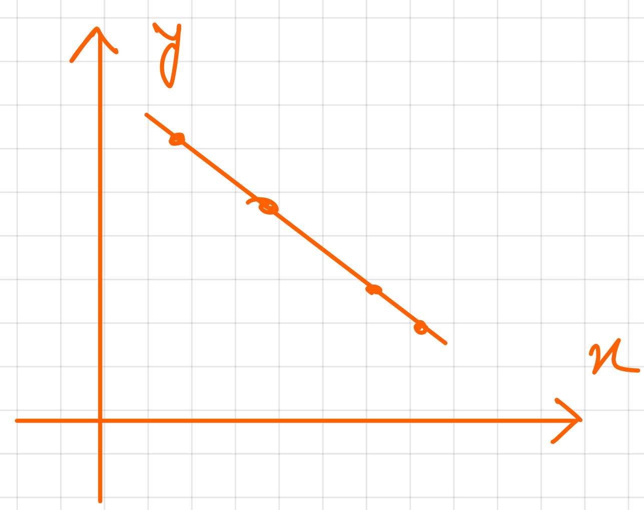
$$= -0.12 \quad \text{Ans}$$

• Regression line and curve fitting :-

X : x_1 x_2 x_3 —— x_n

Y : y_1 y_2 y_3 —— y_n

$(x_1,y_1),(x_2,y_2)$ —— $(x_n,y_n) \rightarrow$ Bivariate data



(i) Regression line of y on x : $y-\bar{y} = b_{yx}(x-\bar{x})$

(ii) " " " x on y : $x-\bar{x} = b_{xy}(y-\bar{y})$

$b_{yx} = r_{xy}\cdot\frac{\sigma_y}{\sigma_x}$ = regression coeff. of y on x.

$b_{xy} = r_{xy}\cdot\frac{\sigma_x}{\sigma_y}$ = regression coefficient of x on y

$$\boxed{b_{yx}\cdot b_{xy} = r_{xy}^2}$$

Remember

Curve fitting by least square method :-

18/Nov/2022

$X: x_1 \quad x_2 \quad x_3 \quad \text{———} \quad x_n$

$Y: y_1 \quad y_2 \quad y_3 \quad \text{———} \quad y_n$

$y = a + bx$

$y_i = a + bx_i, \; i = 1, 2, \ldots n$

$$\Rightarrow \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} a + \sum_{i=1}^{n} bx_i \quad \text{———} \; (1)$$

$xy = ax + bx^2$

$x_i y_i = ax_i + bx_i^2$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} bx_i^2 \quad \text{———} \; (11)$$

Q. Fit the straight line of the form $y = a + bx$ to the following data :—

$X: \; 2 \quad 5 \quad 6 \quad 8 \quad 9$

$Y: \; 8 \quad 14 \quad 19 \quad 20 \quad 31$

x_i	y_i	$x_i y_i$	x_i^2
2	8	16	4
5	14	70	25
6	19	114	36
8	20	160	64
9	31	279	81
$\Sigma x_i = 30$	$\Sigma y_i = 92$	$\Sigma = 639$	$\Sigma = 210$

Soln: $\sum_{i=1}^{5} y_i = \sum_{i=1}^{5} a + \sum_{i=1}^{5} bx_i$

$\Rightarrow 92 = 5a + b.30$ ——— (1)

$\Rightarrow \sum_{i=1}^{5} x_i y_i = \sum_{i=1}^{5} ax_i + \sum_{i=1}^{5} bx_i^2$

$639 = a.30 + b.210$ ——— (11)

Solving (1) and (11), we get,

$a = 1, \; b = 2.9$

$y = 1 + 2.9x$ **Ans**

$y = a + bx + cx^2$ $\quad X: x_1 \quad x_2 \text{ ——— } x_n$

$\quad Y: y_1 \quad y_2 \text{ ——— } y_n$

① $\sum y_i = \sum a + \sum b x_i + c \sum x_i^2$ ——— (i)

② $xy = ax + bx^2 + cx^3$

$\sum x_i y_i = \sum a x_i + \sum b x_i^2 + \sum c x_i^3$ ——— (ii)

③ $x^2 y = ax^2 + bx^3 + cx^4$

$\sum x_i^2 y_i^2 = \sum a x_i^2 + \sum b x_i^3 + \sum c x_i^4$ ——— (iii)

Important!

Q. Fit the 2nd degree parabola $y = a + bx + cx^2$ to the observation.

X: 0 1 2 3 4

Y: 1 5 10 22 38

Soln:

x_i	y_i	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
0	1				
1	5				
2	10				
3	22				
4	38				
$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	Σ

$5a + 10b + 30c = 76$

$10a + 30b + 100c = 243$

$30a + 100b + 354c = 851$

Solving, $a = \frac{10}{7}$, $b = \frac{17}{70}$, $c = \frac{31}{14}$

• Testing hypothesis :-

Important!

Remember

Definition : • **Statistical hypothesis :-**

Any assumption taken on a population regarding it's probability distribution or its parameter is called statistical hypothesis.

• Simple hypothesis :-

A statistical hypothesis which specifies the probability distribution and all related parameters of a population is called simple hypothesis.

• Composite hypothesis :-

A statistical hypothesis which does not specify the population completely is called composite hypothesis.

• NULL hypothesis :-

A statistical hypothesis whose possible acceptance or rejection is tested on the basis of sample observation is called null hypothesis.

Suppose, our assumption is $m = 40$,

NULL hypothesis denoted as $H_0\,(m = 40)$

23/Nov/2022

Critical Region :-

Let (H_0) be null hypothesis & 't' be an appropriate test by which we decide whether (H_0) would be accepted or not.

Let a positive number α is taken under consideration and let $P(a \leq t \leq b) = 1 - \alpha$. If computed value '$t$' lies within (a, b) then H_0 is accepted at the α level of significance & the interval (a, b) is called Region of acceptance & beyond the (a, b) is called critical region.

A null hypothesis, Ho tested on the basis of sample values only. So here two types of errors may occur.

Type ①: Occured when Ho is rejected, but observation was really true.

Type ②: Occured when Ho is accepted but observation was really false.

Assignment Problems :-

5) Let $a_1, a_2, a_3 \ldots a_n$ be n positive number.

$$A.M = \frac{a_1 + a_2 + a_3 + \ldots + a_n}{n}$$

$$H.M = \frac{n}{1/a_1 + 1/a_2 + \ldots + 1/a_n}$$

$$G.M = \sqrt[n]{a_1 . a_2 \ldots \ldots a_n}$$

Let, 2 +ve numbers are $m \times n$

$\therefore$ $G.M = \sqrt[2]{m \times n}$

$$A.M = \frac{m+n}{2} = k$$

$$H.M = \frac{2}{1/m + 1/n} = l$$

$$G.M \text{ of } k \text{ and } l = \sqrt[2]{k \times l} = \sqrt[2]{\frac{m+n}{\cancel{2}} \times \frac{\cancel{2}}{1/m + 1/n}}$$

$$= \sqrt[2]{(\cancel{m+n}) \frac{mn}{(\cancel{m+n})}} = \sqrt[2]{mn}$$

$\therefore$ G.M of k and l = G.M of m and n.

8) $X \sim N(m, \sigma)$, Let for standard normal distribution $N(0,1)$, $\therefore Z = \frac{x-m}{\sigma}$

for the sample of size n, $Z = \frac{\bar{X}-m}{\sigma/\sqrt{n}}$

9) b

10) b

11) c

12) a

13) c

15) $m_r' = E\{(x-a)^r\}$

$m_r = E\{(x-m)^r\}$

$m_1' = E\{(x-5)^1\}$

$= E(x) - 5$

$= m - 5$

Name → Aditya Kundu, Roll → 21CS011004 30/Nov/2022

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

$= \int_0^2 x(|x-1| + |1-x|)\,dx$

$= \int_0^2 x[\cancel{x}\cancel{-1}\cancel{+1}\cancel{-x}]\,dx + \int_0^2 x[x-1-1+x]\,dx$

$= \int_0^2 x(2x-2)\,dx$

$= \int_0^2 (2x^2 - 2x)\,dx$

$= 2\int_0^2 x^2\,dx - 2\int_0^2 x\,dx$

$= 2\left[\frac{x^3}{3}\right]_0^2 - 2\left[\frac{x^2}{2}\right]_0^2$

$= 2\left[\frac{8}{3}\right] - 2\left[\frac{4}{2}\right]$

$= \frac{16}{3} - 4 = \frac{4}{3}$ Ans

---

2) $X \sim N(2,1)$

$E(5x^2+3)$

$\Rightarrow E(5x^2)+3$

$= 5(E(x^2))+3$

$= 5\times5+3$

$= 25+3 = 28$

$\sigma = Var(x)$

$Var(x) = E(x^2) - m^2$

$\Rightarrow 1 = E(x^2) - \{E(x)\}^2$

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

$\Rightarrow 1 = E(x^2) - 4$

$\Rightarrow E(x^2) = 5$

3)

x_i	y_i	$x_i y_i$	x_i^2
1	5	5	1
2	6	12	4
3	8	24	9
4	10	40	16
$\Sigma x_i = 10$	$\Sigma y_i = 29$	$\Sigma x_i y_i = 81$	$\Sigma x_i^2 = 30$

$$\therefore \sum_{i=1}^{4} y_i = \sum_{i=1}^{4} a + \sum_{i=1}^{4} b x_i$$

$$\Rightarrow 29 = 4a + b \sum_{i=1}^{4} x_i$$

$$\Rightarrow 29 = 4a + b \times 10$$

$$\Rightarrow 4a + 10b = 29 \quad \text{—— (I)}$$

Similarly,

$$\sum_{i=1}^{4} x_i y_i = \sum_{i=1}^{4} a x_i + \sum_{i=1}^{4} b x_i^2$$

$$\Rightarrow 81 = a \sum_{i=1}^{4} x_i + b \sum_{i=1}^{4} x_i^2$$

$$\Rightarrow 81 = a \times 10 + b \times 30$$

$$\Rightarrow 10a + 30b = 81 \quad \text{—— (II)}$$

From (I) and (II),

$$4a + 10b = 29 \quad , \quad 10a + 30b = 81$$

$$\Rightarrow a = \frac{29 - 10b}{4} \qquad \Rightarrow a = \frac{81 - 30b}{10}$$

$$\therefore \frac{29 - 10b}{4} = \frac{81 - 30b}{10}$$

$$\Rightarrow 290 - 100b = 324 - 120b$$

$$\Rightarrow 120b - 100b = 324 - 290$$

$$\Rightarrow 20b = 34$$

$$\Rightarrow b = \cancel{\frac{34}{20}} = \frac{17}{10}$$

$$a = \frac{81 - 30b}{10} = \frac{81 - 51}{10} = \frac{3\cancel{0}}{1\cancel{0}} = 3$$

$\therefore$ Eqn of St. line $\rightarrow$ $y = a + bx$

$$\Rightarrow y = 3 + \frac{17}{10} x$$

$$\Rightarrow 10y = 30 + 17x$$

$$\Rightarrow 17x - 10y + 30 = 0$$

Ans