$A_1, A_2 \ldots A_n$ and $B_1, B_2 \ldots B_n$ are numbers.

Theorem:-

The correlation coefficient b/w x & y always lies b/w -1 & $+1$

i.e; $-1 \leq r \leq 1$

Proof:-

Let there be 'n' pairs of obsrvns (x_1, y_1) $(x_2, y_2) \ldots (x_n, y_n)$. Then

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

$$r^2 = \frac{\left[\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2} \quad \text{--- (1)}$$

Let $x_i - \bar{x} = A_i$ and $y_i - \bar{y} = B_i$

$\therefore$ (1) $\Rightarrow$

$$r^2 = \frac{\left(\sum_{i=1}^{n} A_i B_i\right)^2}{\sum_{i=1}^{n} A_i^2 \cdot \sum_{i=1}^{n} B_i^2} \quad \text{--- (2)}$$

By Schwartz's inequality:

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

$$\Rightarrow \frac{\left(\sum_{i=1}^{n} A_i B_i\right)^2}{\sum_{i=1}^{n} A_i^2 \cdot \sum_{i=1}^{n} B_i^2} \leq 1 \quad \text{--- (3)}$$

(3) in (2) $\Rightarrow r^2 \leq 1 \Rightarrow -1 \leq r \leq 1$.

Note:-

- If $r > 0$, corrln is +ve
- If $r < 0$, corrln is -ve
- If $r = 0$, no corrln
- If $r = +1$, corrln is perfect +ve
- If $r = -1$, corrln is perfect -ve.

⇒ Regression

It is the estimation of unknown values of one variable from known values of another variable.

⇒ Linear Regression:-

The eqn which gives the relationship b/w the variables is called regrn eqn. When the correlation is linear, the regrn eqn will be linear and of the form $y = a + bx$. Here 'x' is the independent variable and 'y' is the dependent variable. This is known as regrn eqn of y on x. The constant 'b' is known as regrsn coefficient.

If y is the independent variable and 'x' is the dependent variable, the

Page No:
Date:

regrs^n eqn will be of the form $x = c + dy$. This is known as the regrs^n eq^n of x on y.

⇒ Lines of Regression:-

Let (x_i, y_i) for $i = 1, 2 \ldots n$ be the given 'n' observ^{ms}. If we use the method of least squares, such that

(i) the sum of squares of deviations parallel to y-axis is minimized, we get a regr^n line of y on x.

(ii) the sum of squares of deviations parallel to the x-axis is minimised, we get regrs^n line of x on y.

Let the regrs^n line of y on 'x' be

$y = a + bx$ —— ①

we have to determine a & b, by the principle of least squares. Then

$\bar{y} = a + b\bar{x}$ —— ②

① - ② ⇒ $y - \bar{y} = b(x - \bar{x})$

The normal eqns are:-

$\Sigma y_i = na + b\Sigma x_i$ —— ③

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

Shifting the origin to $(\bar{x}, \bar{y})$, ④ ⇒

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = a\Sigma(x_i - \bar{x}) + b\Sigma(x_i - \bar{x})^2$

Page No:
Date:

$\frac{1}{n}\Sigma(x_i - \bar{x})(y_i - \bar{y}) = a\frac{1}{n}\Sigma(x_i - \bar{x}) + b\frac{1}{n}\Sigma(x_i - \bar{x})^2$ —— ⑤

We know that, the corr^n coeff.

r or $\sigma_{xy} = \dfrac{\frac{1}{n}\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \cdot \sigma_y}$

$\Rightarrow \frac{1}{n}\Sigma(x_i - \bar{x})(y_i - \bar{y}) = r \cdot \sigma_x \cdot \sigma_y$.

$\Sigma(x_i - \bar{x}) = \Sigma x_i - \Sigma\bar{x}$

$= n\bar{x} - n\bar{x} = 0$

$\dfrac{\Sigma(x_i - \bar{x})^2}{n} = \sigma_x^2 \Rightarrow \Sigma(x_i - \bar{x})^2 = n\sigma_x^2$.

∴ ⑤ ⇒ $r\sigma_x\sigma_y = a \cdot 0 + \frac{b}{n} \cdot n\sigma_x^2$

$\Rightarrow r\sigma_x\sigma_y = b\sigma_x^2$

$\Rightarrow b = \dfrac{r\sigma_y}{\sigma_x}$, which is the regrs^n coefficient of y on x.

It is denoted by b_{yx}. Hence regrs^n line of y on x is

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

The regrs^n line of x on y be

$x = c + dy$ —— ①

It passes through $(\bar{x}, \bar{y})$

ie; $\bar{x} = c + d\bar{y}$ —— ②

① - ② ⇒ $x - \bar{x} = d(y - \bar{y})$.

Page No :
Date :

Normal eqns are:-

$\Sigma x_i = nc + d\Sigma y_i$ —— (3)

$\Sigma x_i y_i = c\Sigma y_i + d\Sigma y_i^2$ —— (4)

Shifting origin to $(\bar{x}, \bar{y})$, (4) $\Rightarrow$

$\Sigma(x_i-\bar{x})(y_i-\bar{y}) = c\Sigma(y_i-\bar{y}) + d\Sigma(y_i-\bar{y})^2$

$\frac{1}{n}\Sigma(x_i-\bar{x})(y_i-\bar{y}) = c\cdot\frac{1}{n}\Sigma(y_i-\bar{y}) + d\cdot\frac{1}{n}\Sigma(y_i-\bar{y})^2$

$\sigma_{xy} = c\cdot 0 + d\cdot\sigma^2_y$.

$r\sigma_{xy} = d\sigma^2_y$.

$\Rightarrow d = \frac{r\sigma_x}{\sigma_y}$ which is the regrsn coefficient of x on y.

It is denoted by b_{xy}.

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

$\therefore$ The regrsn eqn of x on y is

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

-CONTD-

For std normal table we get,

$\frac{62-\mu}{\sigma} = 1.89$

$62-\mu = 1.89\sigma$ —— (2)

(1) & (2)

$62-\mu = 1.89\sigma$

$30-\mu = -1.29\sigma$

$32 = 3.18\sigma$

$\therefore \sigma = \frac{32}{3.18} = 10.063$

$-1.89 \times 10.063 + 62 = \mu$

$\therefore \mu = 42.98$

Page No :
Date :

10% get marks less than 30

$P(X < 30) = \frac{10}{100}$

97% get marks less than 62.

$P(X < 62) = \frac{97}{100}$

$\therefore P(X < 30) = 0.1$

$P\left(\frac{X-\mu}{\sigma} < \frac{30-\mu}{\sigma}\right) = 0.1$

$P\left(Z < \frac{30-\mu}{\sigma}\right) = 0.1$

$P(Z < Z_1) = 0.1$

$0.5 - P(Z_1 < Z < 0) = 0.1$

$P\left(\frac{30-\mu}{\sigma} < Z < 0\right) = 0.4$

From std normal table, we get,

$\frac{30-\mu}{\sigma} = -1.29$

$30-\mu = -1.29\sigma$ —— (1)

$P(X < 62) = 0.97$

$P\left(\frac{X-\mu}{\sigma} < \frac{62-\mu}{\sigma}\right) = 0.97$

$P\left(Z < \frac{62-\mu}{\sigma}\right) = 0.97$

$0.5 + P\left(0 < Z < \frac{62-\mu}{\sigma}\right) = 0.97$

$P\left(0 < Z < \frac{62-\mu}{\sigma}\right) = 0.47$

Page No :
Date :

2. If x has a uniform distrb^n and ^{n} in $(-k, k)$, $k > 0$. Find k s.t. $P(|x| < 1) = P(|x| > 1)$

→ $X \sim$ Uniform distrb^n

$\therefore f(x) = \begin{cases} \frac{1}{2k} ; & -k \le x \le k \\ 0 ; & \text{otherwise} \end{cases}$

$P(|x| < 1) = P(|x| > 1)$

$= 1 - P(|x| \le 1)$

$\Rightarrow P(|x| < 1) + P(|x| \le 1) = 1$

$2P(|x| \le 1) = 1$

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

$\Rightarrow 2\int_{-1}^{1} \frac{1}{2k}\,dx = 1$

$\Rightarrow \int_{-1}^{1} \frac{1}{k}\,dx = \left[\frac{1}{k} \times x\right]_{-1}^{1} = \frac{2}{k}$

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

3. At an examination 10% of the students get less than 30 marks and 97% got less than 62 marks. Assuming normal distrb^n, find σ & μ.

→ Let $z = \frac{\cancel{x}-\mu}{\sigma}$

$$\text{Mean}, \mu = E(x) = \int_0^1 x \cdot f(x)\, dx$$

$$= \int_0^1 5x(1-x)^4\, dx$$

$$= \left[\frac{\cancel{5}x(1-x)^5}{\cancel{5}} \times -1 - \int 5\, \frac{(1-x)^5}{5} \times -1 \right]_0^1$$

$$= \left[-x(1-x)^5 - \frac{(1-x)^6}{6} \right]_0^1$$

$$= 0 - (0 - 1/6) = 1/6$$

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

$$E(x^2) = \int_0^1 x^2 \cdot f(x)\, dx = \int_0^1 5x^2(1-x)^4\, dx$$

$$= 5x^2 \frac{(1-x)^5}{5} \times -1 - \left[\int 10x \times \frac{(1-x)^5}{5} \times -1\, dx \right]$$

$$= -x^2(1-x)^5 - \left[2x\frac{(1-x)^5}{6} - \int 2x\, \frac{(1-x)^6}{6}\, dx \right]$$

$$= \left[-x^2(1-x)^5 - \frac{x(1-x)^6}{3} - \frac{(1-x)^7}{7 \times 3} \times -1 \right]_0^1$$

$$= 1/21$$

$$\therefore Var(x) = \frac{1}{21} - \frac{1}{36} = 0.0198$$

$$F(x) = \begin{cases} 0 & ; x < 0 \\ (1-x)^5 & ; 0 \le x \le 1 \\ 1 & ; x \ge 1 \end{cases}$$

Assignment - 2

1. If $f(x) = \begin{cases} k(1-x)^4 & ; 0 \le x \le 1 \\ 0 & ; \text{otherwise} \end{cases}$ is a pdf. find

(i) K (ii) $P(x \ge 1/2)$ (iii) $P(0.2 \le x \le 0.8)$

Also find mean variance, and distrbⁿ funcⁿ.

→ Given $f(x) = k(1-x)^4 \; ; 0 \le x \le 1$

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

$$\int_0^1 k(1-x)^4\, dx = k\int_0^1 (1-x)^4\, dx$$

$$= k\left[\frac{(1-x)^5}{5} \times -1 \right]_0^1 = k(0 + 1/5) = k/5$$

$$\therefore \frac{k}{5} = 1 \Rightarrow k = 5$$

$$\therefore f(x) = \begin{cases} 5(1-x)^4 & ; 0 \le x \le 1 \\ 0 & ; \text{otherwise} \end{cases}$$

(ii) $$P(x \ge 1/2) = \int_{1/2}^1 f(x)\, dx$$

$$= \int_{1/2}^1 5(1-x)^4\, dx = \left[5\, \frac{(1-x)^5}{5} \times -1 \right]_{1/2}^1$$

$$= -5\left[0 - \frac{1}{32 \times 5} \right] = \frac{1}{32}$$

(iii) $$P(0.2 \le x \le 0.8) = \int_{0.2}^{0.8} f(x)\, dx = \int_{0.2}^{0.8} 5(1-x)^4\, dx$$

$$= \left[\frac{5(1-x)^5}{5} \times -1 \right]_{0.2}^{0.8} = -(0.2)^5 + (0.8)^5 = 0.3273$$

Page No :
Date :

To find theoratical frequencies :-

x	$f(x)$	Theoratical freq $N \times f(x)$	approx $\approx$ TF
0	$5C_0(0.57)^0(0.43)^5 = 0.0147$	$100 \times 0.0147 = 1.47$	1
1	$5C_1(0.57)^1(0.43)^4 = 0.0974$	$100 \times 0.0974 = 9.74$	10
2	$5C_2(0.57)^2(0.43)^3 = 0.2583$	$100 \times 0.2583 = 25.83$	26
3	$5C_3(0.57)^3(0.43)^2 = 0.3424$	$100 \times 0.3424 = 34.24$	34
4	$5C_4(0.57)^4(0.43)^1 = 0.2269$	$100 \times 0.2269 = 22.69$	23
5	$5C_5(0.57)^5(0.43)^0 = 0.0601$	$100 \times 0.0601 = 6.01$	6
			100.

3. Find the pb that atmost 5 defective bulbs will be found in a box of 200 bolts if it is known that 2% of such bolts are imputed to be defective.

→ $p = 2\% = \frac{2}{100}$; $n = 200$

$\lambda = np = 200 \times \frac{2}{100} = 4$

For Poisson distrbn,

$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$; $x = 0, 1, 2 \cdots \infty$

$= \frac{e^{-4}4^x}{x!}$; $x = 0, 1, 2 \cdots \infty$

Now, $p(5) = \frac{e^{-4}4^5}{5!} = 0.1562$.

Page No :
Date :

(i) $P(1 \le x \le 3) = P(x = 1, 2, 3)$

$= P(x=1) + P(x=2) + P(x=3)$

$= 1/2 + 3/10 + 1/30$

$= 5/6$

2. Fit a binomial distrbn of the following data:

x	0	1	2	3	4	5
$f(x)$	2	14	20	34	22	8

x	$f(x)$	$x \cdot f(x)$
0	2	0
1	14	14
2	20	40
3	34	102
4	22	88
5	8	40
	100	284

$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{284}{100} = 2.84$

$n = 5$

$\Sigma f = N = 100$

For binomial distribution,

$\bar{x} = np \Rightarrow p = \frac{\bar{x}}{n} = \frac{2.84}{5} = 0.568$

$= 0.57$

$\therefore q = 1 - p = 1 - 0.57 = 0.43$

$\therefore f(x) = nC_x p^x q^{n-x}$

$= 5C_x(0.57)^x(0.43)^{5-x}$

$P(0.71) = 0.26115$

Assignment - 1

1. Find k if the following is a pdf. Also find (i) $P(1 \le X \le 3)$ (ii) Mean & variance.

x	0	1	2	3
$f(x)$	1/6	1/2	k/10	1/30

$\Sigma f(x) = 1$

i.e; $\frac{1}{6} + \frac{1}{2} + \frac{k}{10} + \frac{1}{30} = 1$

$\frac{7}{10} + \frac{k}{10} = 1 \Rightarrow \frac{7+k}{10} = 1$

$7 + k = 10 \Rightarrow k = 3$

x	0	1	2	3
$f(x)$	1/6	1/2	3/10	1/30

(ii) Mean, $\mu = \sum_{x=0}^{3} x \cdot f(x)$

$= 0 \times \frac{1}{6} + 1 \times \frac{1}{2} + 2 \times \frac{3}{10} + 3 \times \frac{1}{30}$

$= 0 + \frac{1}{2} + \frac{6}{10} + \frac{3}{30} = \frac{6}{5}$

Variance $= \Sigma x^2 \cdot f(x) - \mu^2$

$\Sigma x^2 \cdot f(x) = 0 \times 1/6 + 1^2 \times 1/2 + 2^2 \times 3/10 + 3^2 \times 1/30$

$= 0 + \frac{1}{2} + \frac{6}{5} + \frac{3}{10} = 2$

$Var(x) = \Sigma x^2 f(x) - \mu^2$

$= 2 - \left(\frac{6}{5}\right)^2$

$= \frac{14}{25}$

$= P(Z \le -3/4) = P(Z \le -0.75)$

$= 0.5 - P(0 \le Z \le 0.75)$

$= 0.5 - 0.2734 = 0.2266$

4. If X is normally distributed and the mean of X is 12 and SD is 4. Find x' when $P(x > x') = 0.24$.

→ Given $\mu = 12$ & $SD = 4$

$P(x > x') = 0.24$

$P\left(\frac{x-\mu}{\sigma} > \frac{x'-\mu}{\sigma}\right) = 0.24$

$P\left(z > \frac{x'-12}{4}\right) = 0.24$

$0.5 - P\left(0 \le z \le \frac{x'-12}{4}\right) = 0.24$

$P\left(0 \le z \le \frac{x'-12}{4}\right) = 0.5 - 0.24 = 0.26$

$\frac{x'-12}{4} = 0.71$ (from table)

$x' = (0.71 \times 4) + 12$

$= 14.84$

Page No :

Date :

→ $\mu = 8 \; ; \; \sigma = 4$

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

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

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

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

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

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

$= 0.2734 + 0.1915 = 0.4649$ //

b) $P(10 \le x \le 15) = P\left(\frac{10-8}{4} \le \frac{x-\mu}{\sigma} \le \frac{15-8}{4}\right)$

$= P(1/2 \le z \le 7/4)$

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

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

$= 0.4599 - 0.1915$

$= 0.2684$ //

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

$= P(z \geq 7/4) = P(z \geq 1.75)$

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

$= 0.75 - 0.4599$

$= 0.0401$ //

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

Page No :

Date :

2. A bus arrives every 15 minutes at a bus stop. Assuming that the waiting time X for bus is uniformly distributed find the probability that a person has to wait for the bus

a) for more than 10 minutes

b) between 5 & 10 minutes.

→ a) Since Uniform distribution,

$f(x) = \frac{1}{b-a}; \; a < x < b.$

Here $a = 0; \; b = 15$

$\therefore$ pdf $= 1/15 = f(x)$

a) $P(x > 10) = \int_{10}^{15} \frac{1}{15}\, dx = \frac{1}{15}[x]_{10}^{15}$

$= \frac{1}{15}(15-10) = \frac{5}{15} = 1/3$ //

b) $P(5 < x < 10) = \int_{5}^{10} \frac{1}{15}\, dx = \frac{1}{15}[x]_{5}^{10}$

$= \frac{1}{15}(10-5) = \frac{5}{15} = 1/3$ //

3. If X is normally distributed with mean 8 & SD 4 find

a) $P(5 \le x \le 10)$ (c) $P(x \geq 15)$

b) $P(10 \le x \le 15)$ (d) $P(x \not> 5)$

Page No :
Date :

$$= \frac{1}{10}\left[\left(\frac{2^4}{2}\right)+(2^3)\right] = \frac{16}{10} = 1.6$$

$\therefore Var(X) = 1.6 - (1.13)^2 = 0.3231$

distb^n func^n :-

$$\int_0^x F(x)\,dx = \int_0^x \frac{2x+3}{10}\,dx$$

put $t = 2x+3$
$dt = 2\,dx$
$dx = dt/2$
$x = 0 \Rightarrow t = 3$
$x = x \Rightarrow t = 2x+3$

$$= \int_3^{2x+3} t \cdot \frac{dt}{2} \times \frac{1}{10}$$

$$= \frac{1}{20}\int_3^{2x+3} t \cdot dt$$

$$= \frac{1}{20}\left[\frac{t^2}{2}\right]_3^{2x+3}$$

$$= \frac{1}{40}(2x+3)^2 - 9$$

$$= \frac{1}{40}(4x^2+12x+9-9)$$

$$= \frac{1}{40}\times 4(x^2+3x)$$

$$= \frac{1}{10}(x^2+3x)$$

$$\therefore F(x) = \begin{cases} \frac{1}{10}(x^2+3x) & ; 0<x<2 \\ 0 & ; \text{Otherwise} \end{cases}$$

21/6/2021

Page No : 35
Date :

Tutorial-2.

1. If $f(x) = \begin{cases} k(2x+3) & ; 0<x<2 \\ 0 & ; \text{otherwise} \end{cases}$

is a pdf, find k, mean, variance and distb func^n $F(x)$.

— Since $f(x)$ is a pdf,

$$\int_0^2 f(x)\,dx = 1 \Rightarrow \int_0^2 k(2x+3)\,dx = 1$$

$$\Rightarrow k\int_0^2 (2x+3)\,dx = 1 \Rightarrow k\left[\frac{2x^2}{2}+3x\right]_0^2 = 1$$

$$\Rightarrow k(4+6) = 1 \Rightarrow 10k = 1 \Rightarrow k = 1/10$$

Mean; $E(X) = \int_0^2 x \cdot f(x)\,dx$

$$= \int_0^2 x \cdot \frac{1}{10}(2x+3)\,dx$$

$$= \frac{1}{10}\int_0^2 (2x^2+3x)\,dx$$

$$= \frac{1}{10}\left[\frac{2x^3}{3}+\frac{3x^2}{2}\right]_0^2$$

$$= \frac{1}{10}\left(\frac{2(8)}{3}+\frac{3}{2}\times 4\right) = 1.13$$

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

$$E(X^2) = \int_0^2 x^2 \cdot f(x)\,dx = \int_0^2 x^2 \cdot \frac{1}{10}(2x+3)\,dx$$

$$= \frac{1}{10}\int_0^2 (2x^3+3x^2)\,dx = \frac{1}{10}\left[\frac{2x^4}{4}+\frac{3x^3}{3}\right]_0^2$$

Page No :

Date :

$N \times f(0) = 420 \times e^{-1.26} = 119.13 \approx 119$

$N \times f(1) = N \times f(0) \times \frac{\bar{x}}{1} = 119.13 \times 1.26$

$= 150.1 \approx 150$

$N \times f(2) = N \times f(1) \times \frac{\bar{x}}{2} = 150.1 \times \frac{1.26}{2}$

$= 94.56 \approx 95$

$N \times f(3) = N \times f(2) \times \frac{\bar{x}}{3} = 94.56 \times \frac{1.26}{3}$

$= 39.7 \approx 40$

$N \times f(4) = N \times f(3) \times \frac{\bar{x}}{4} = 39.7 \times \frac{1.26}{4}$

$= 12.5 \approx 12$

$N \times f(5) = N \times f(4) \times \frac{\bar{x}}{5} = 12.5 \times \frac{1.26}{5}$

$= 3.15 \approx 3$

$N \times f(6) = N \times f(5) \times \frac{\bar{x}}{6} = 3.15 \times \frac{1.26}{6}$

$= 0.66 \approx 1$

$N \times f(7) = N \times f(6) \times \frac{\bar{x}}{7} = 0.66 \times \frac{1.26}{7} = 0.12$

≈ 0

x	0	1	2	3	4	5	6	7	Total
f	123	143	98	42	8	4	2	0	420
Tf	119	150	95	40	12	3	1	0	420

$\frac{1}{10}\int_0^2 2x+3\,dx$ $\begin{cases} \frac{1}{10}(2x+3); & 0 \le x \le 2 \\ 0; & \text{elsewhere} \end{cases}$

$\frac{1}{10}\left[x^2+3x\right]_0^2 = \frac{1}{10} \times 8 =$

Page No :

Date :

$N \times f(0) = 420 \times \frac{e^{-1.26}(1.26)^0}{0!}$

$= 420 \times 0.28 = 119.6 \approx 118$

$N \times f(1) = N \times f(0) \times \frac{\bar{x}}{1} = 117.6 \times 1.26$

$= 148.176 \approx 149$

$N \times f(2) = N \times f(1) \times \frac{\bar{x}}{2} = 148.176 \times \frac{1.26}{2}$

$= 93.35 \approx 93$

$N \times f(3) = N \times f(2) \times \frac{\bar{x}}{3} = 93.35 \times \frac{1.26}{3}$

$= 39.21 \approx 39$

$N \times f(4) = N \times f(3) \times \frac{\bar{x}}{4} = 39.21 \times \frac{1.26}{4}$

$= 12.35 \approx 12$

$N \times f(5) = N \times f(4) \times \frac{\bar{x}}{5} = 12.35 \times \frac{1.26}{5}$

$= 3.11 \approx 3$

$N \times f(6) = N \times f(5) \times \frac{\bar{x}}{6} = 3.11 \times \frac{1.26}{6} = 0.65$

≈ 1

$N \times f(7) = N \times f(6) \times \frac{\bar{x}}{7} = 0.65 \times \frac{1.26}{7} = 0.12$

≈ 0

x	0	1	2	3	4	5	6	7	Total
f	123	143	98	42	8	4	2	0	
Tf	118	149	93	39	12	3	1	0	

$f(5) = {}^6C_5 (0.44)^5 (0.56) = 0.055$

$N \times f(5) = 200 \times 0.055 = 11.08 \approx 11$

$f(6) = {}^6C_6 (0.44)^6 (0.56)^0 = 0.0073$

$N \times f(6) = 200 \times 0.0073 = 1.46 \approx 2$

x	0	1	2	3	4	5	6	Total
f	13	25	52	58	32	16	4	200
xf	6	29	57	60	35	11	2	200

4. Fit a poisson distrib^n to the data

x	0	1	2	3	4	5	6	7
f	123	143	98	42	8	4	2	0

→

x	f	fx
0	123	0
1	143	143
2	98	196
3	42	126
4	8	32
5	4	20
6	2	12
7	0	0
	420	529

$\bar{x} = \frac{529}{420} = 1.26$

$\lambda = \bar{x} = 1.26$

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

$= \frac{e^{-1.26} (1.26)^x}{x!} \quad ; x = 0, 1, 2 \cdots 7$

Here $n = 6$.

For a binomial distrib^n,

$\bar{x} = np \Rightarrow p = \frac{\bar{x}}{n} = \frac{2.675}{6} = 0.44$

$q = 1 - p = 1 - 0.44$
$= 0.56$

$f(x) = {}^nC_x p^x q^{n-x}$
$= {}^6C_x (0.44)^x (0.56)^{6-x} \quad ; x = 0, 1 \ldots 6$

$f(0) = {}^6C_0 (0.44)^0 (0.56)^6 = 0.031$

$N \cdot f(0) = 200 \times 0.031 = 6.2$
≈ 6

$f(1) = {}^6C_1 (0.44)^1 (0.56)^5 = 0.145$

$N \cdot f(1) = 200 \times 0.145 = 29.1 \approx 29$

$f(2) = {}^6C_2 (0.44)^2 (0.56)^4$
$= 0.285$

$N \cdot f(2) = 200 \times 0.285 = 57.12 \approx 57$

$f(3) = {}^6C_3 (0.44)^3 (0.56)^3 = 0.299$

$N \cdot f(3) = 200 \times 0.299 = 59.8$
≈ 60

$f(4) = {}^6C_4 (0.44)^4 (0.56)^2 = 0.176$

$N \times f(4) = 200 \times 0.176 = 35.26$
≈ 36

$= 5C_3 (1/2)^3 (1/2)^2$

$= 0.3125$

Required no. of boys $= N \times P(3)$

$= 800 \times 0.3125$

$= 250$

(ii) $P(5 \text{ girls}) = P(\text{no boys})$

$= P(X=0) = P(0)$

$= 5C_0 (1/2)^0 (1/2)^{5-0}$

$= 1 \times 1 \times (1/2)^5$

$= 0.03125$

required no. of girls $= N \times P(0)$

$= 800 \times 0.03125$

$= 25$

3. Fit a binomial distrbⁿ to the data

x	0	1	2	3	4	5	6
f	13	25	52	58	32	16	4

→

x	f	xf
0	13	0
1	25	25
2	52	104
3	58	174
4	32	128
5	16	80
6	4	24
	200	535

$N = \Sigma f_i = 200$

$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{535}{200}$

$= 2.675$

$\sigma^2 = \Sigma x^2 f(x) - \mu^2$

$= \left[0 + \left(1^2 \times \frac{5}{27}\right) + \left(4 \times \frac{14}{27}\right) + \left(9 \times \frac{1}{54}\right)\right] - \left(\frac{69}{54}\right)^2$

$= \frac{5}{27} + \frac{56}{27} + \frac{9}{54} - \frac{4761}{2916}$

$= \frac{10+112+9}{54} - \frac{4761}{2916}$

$= \frac{131}{54} - \frac{4761}{2916}$

$= \frac{257}{324} = 0.7932$

2. Out of 800 families of 5 children each, how many would you expect to have (i) 3 boys (ii) 5 girls

→ Required number $= N \times pb.$

Here $n = 5$; $N = 800$.

$p = $ pb for a boy $= 1/2$

$q = 1 - 1/2 = 1/2$

$P(x) = P(x \text{ boys})$

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

$= 5C_x (1/2)^x (1/2)^{5-x}$; $x = 0, 1, 2, 3, 4, 5$

(i) $P(3 \text{ boys}) = 5C_3 (1/2)^3 (1/2)^{5-3}$

1/6/2021

Page No :

Date :

Tutorial-1

1- Find k for which the following is pdf. Also find mean & variance

x	0	1	2	3
$f(x)$	$k/2$	$k/3$	$\frac{k+1}{3}$	$2k-1/6$

→ We know that

$\Sigma f(x) = 1$

$\therefore \Sigma f(x) = \frac{k}{2} + \frac{k}{3} + \frac{k+1}{3} + \frac{2k-1}{6}$

$= \frac{3k+2k+2(k+1)+2k-1}{6}$

$= \frac{3k+2k+2k+2+2k-1}{6}$

$= \frac{9k+1}{6}$

$\therefore \frac{9k+1}{6} = 1 \Rightarrow 9k+1 = 6$

$\Rightarrow 9k = 5 \Rightarrow k = 5/9$

x	0	1	2	3
$f(x)$	$5/18$	$5/27$	$14/27$	$1/54$

$\mu = \Sigma x \cdot f(x)$

$= \left(0\times\frac{5}{18}\right) + \left(1\times\frac{5}{27}\right) + \left(2\times\frac{14}{27}\right) + \left(3\times\frac{1}{54}\right)$

$= \frac{5}{27} + \frac{28}{27} + \frac{3}{54} = \frac{10+56+3}{54} = \frac{69}{54}$

37

$\frac{-2\alpha \pm \sqrt{(2\alpha)^2 - 4\times1\times(2\alpha+3)}}{2}$

$-\alpha \pm \sqrt{\alpha^2 - 2\alpha + 3}$

$4 - 4\times1\times-3$

$4+12$

$\alpha^2 - 2\alpha - 3 \geq 0$

$\alpha = 2 \pm \sqrt{16}$

2 ± 4

$\alpha = 2+4 = 6$ $\quad\quad \alpha = 2-4 = -2$

$\alpha^2 - 2\alpha - 3 = (\alpha+2)(\alpha-6)$

$\alpha^2 - 2\alpha - 3 \geq 0$ iff $(\alpha+2)(\alpha-6) \geq 0$

$0 \leq \alpha \leq 7 \Rightarrow \alpha^2 - 2\alpha - 3 \geq 0$ iff $\alpha - 6 \geq 0$

$P(\alpha \geq 6) =$

$\int_2^5 \frac{1}{5} \quad \frac{1}{5}(5-2) = \frac{3}{5}$

Module-3 - contd -

⇒ Theorem:-

Correlation coefficient is the geometric mean of the regression coefficients.

ie; $r = \pm\sqrt{b_{xy} \cdot b_{yx}}$

Proof:-

The regrs^n coeff. of y on x is

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

The regrs^n coeff of x on y is

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

$$(1) \times (2) \Rightarrow b_{yx} \cdot b_{xy} = \frac{r\sigma_y}{\sigma_x} \times \frac{r\sigma_x}{\sigma_y}$$

$$= r^2$$

$$\Rightarrow r^2 = b_{yx} \cdot b_{xy}$$

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

Note:-

Since σ_x & σ_y are +ve, r has the same sign as that of the regrs^n coefficients. ie; r is +ve if b_{yx} and b_{xy} are both +ve and r is -ve if b_{yx} and b_{xy} are -ve.

Q. If θ is the angle b/w the 2 regrs^n lines. Then show that:

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

→ The regrs^n eq^n of y on x is:

$$y-\bar{y} = \frac{r\sigma_y}{\sigma_x}(x-\bar{x}) \quad \text{—— (1)}$$

Let m_1 = slope of line (1)

$= \text{coeff of } x = \dfrac{r\sigma_y}{\sigma_x}$

The regrs^n eq^n of x on y is:

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

$$y-\bar{y} = \frac{\sigma_y}{r\sigma_x}(x-\bar{x}) \quad \text{—— (2)}$$

Let m_2 = slope of line (2)

$= \text{coeff of } x$

$= \dfrac{\sigma_y}{r\sigma_x}$

Let θ be the angle b/w regrs^n lines.

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

$$= \pm\frac{\left(\frac{r\sigma_y}{\sigma_x}\right)-\left(\frac{\sigma_y}{r\sigma_x}\right)}{1+\left(\frac{\cancel{r}\sigma_y}{\sigma_x}\times\frac{\sigma_y}{\cancel{r}\sigma_x}\right)}$$

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

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

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

? Obtain the eq^n of lines of regrs^n from the following data:-

x	1	2	3	4	5	6	7
y	9	8	10	12	11	13	14

→

x	y	$u=x-4$	$v=y-11$	u^2	v^2	uv
1	9	-3	-2	9	4	6
2	8	-2	-3	4	9	6
3	10	-1	-1	1	1	1
4	12	0	1	0	1	0
5	11	1	0	1	0	0
6	13	2	2	4	4	4
7	14	3	3	9	9	9
		0	0	28	28	26

$\bar{v} = \dfrac{\Sigma v}{n} = \dfrac{0}{7} = 0$; $\bar{u} = \dfrac{\Sigma u}{n} = \dfrac{0}{7} = 0$

$$\sigma_u = \sqrt{\frac{\Sigma u^2}{n} - (\bar{u})^2}$$

$$= \sqrt{\frac{28}{7} - 0^2} = \sqrt{4} = 2$$

$$\sigma_v = \sqrt{\frac{\Sigma v^2}{n} - (\bar{v})^2}$$

$$= \sqrt{\frac{28}{7} - 0} = \sqrt{4} = 2$$

$$r = \frac{\frac{1}{n}\Sigma uv - \bar{u}\bar{v}}{\sqrt{\frac{1}{n}\Sigma u^2 - (\bar{u})^2}\sqrt{\frac{1}{n}\Sigma v^2 - (\bar{v})^2}}$$

$$= \frac{\frac{26}{7} - 0\times 0}{2\times 2} = \frac{3.7}{4} = 0.85$$

The regrsn lines of u & v are

① $v - \bar{v} = \frac{r\sigma_v}{\sigma_u}(u - \bar{u})$

ie; $v - 0 = 0.85 \times \frac{2}{2}(u - 0)$

$v = y - 11$
$u = x - 4$

$v = 0.85 \times u$

$y - 11 = 0.85(x - 4)$

ie; $0.85x - y = 11 - 0.85 \times 4$

$0.85x - y = 7.6$

② $u - \bar{u} = \frac{r\sigma_u}{\sigma_v}(v - \bar{v})$

ie; $u - 0 = 0.85 \times \frac{2}{2}(v - 0)$

$x - 4 = 0.85(y - 11)$

$x - 0.85y = 4 + 0.85 \times 11$

$x - 0.85y = 13.35$.

? Obtain the eqn of the regrsn lines from the following data, using the method of least squares. Hence find the coefficient of corrln b/w x & y. Also estimate the value of

(i) y, when $x = 38$

(ii) x, when $y = 18$.

x	22	26	29	30	31	31	34	35
y	20	20	21	29	27	24	27	31

→ put $u = x - 29$; $v = y - 27$.

Let the eqn of the regrsn line of y on x be $y = Ax + B$.

or $v = Au + B$.

Normal eqns are:-

$\Sigma v = A\Sigma u + nB$ — ①

$\Sigma uv = A\Sigma u^2 + B\Sigma u$ — ②

① & ② ⇒ $6A + 8B = -17$ — ③

$128A + 6B = 90$ — ④

x	y	$u=x-29$	$v=y-27$	u^2	v^2	uv
22	20	-7	-7	49	49	49
26	20	-3	-7	9	49	21
29	21	0	-6	0	36	0
30	29	1	2	1	4	2
31	27	2	0	4	0	0
31	24	2	-3	4	9	-6
34	27	5	0	25	0	0
35	31	6	4	36	16	24
		6	-17	128	163	90

Solving (3) & (4) we get

$A = 0.83$; $B = -2.75$

Here regrs^n of y is

$y - 27 = 0.83(x-29) - 2.75$

$\therefore y = 0.83x + 0.28$ —— (5)

Let the eqn of the regrs^n line of x on y

$x = Cy + D$.

or $u = Cv + D$.

Normal eqns are

$\Sigma u = C\Sigma v + nD$ —— (6)

$\Sigma uv = C\Sigma v^2 + D\Sigma v$ —— (7)

Solving (6) & (7), we get

$-17C + 8D = 6$ —— (8)

$163C - 17D = 90$ —— (9)

Solving (8) & (9) we get

$C = 0.81$ and $D = 2.47$.

Hence regrs^n line of X on Y is

$x - 29 = 0.81(y-27) + 2.47$.

ie, $x = 0.81y + 9.60$ —— (10)

From (5), $b_{yx} = 0.83$

From (10), $b_{xy} = 0.81$

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

$r = \pm 0.81$ [$\because$ b_{yx} & b_{xy} are +ve $\therefore r$ is +ve

$\therefore r = 0.81$

We use (5) to estimate value of Y when $x = 38$.

$y = 0.83 \times 38 + 0.18 = 31.72$.

Using eqn (10) we estimate value of x when $y = 18$.

$x = 0.81 \times 18 + 9.60 = 24.18$

$\Rightarrow$ Identif^n of Regression Lines:-

Let $a_1x + b_1y + c_1 = 0$ —— (1)

$a_2x + b_2y + c_2 = 0$ —— (2)

are eqns of 2 regrs^n eqn of 2 lines (given)

Let us assume that one is regrs^n

line of y on x and the other is regr^n line of x on y.

One can written as:

$$b_1 y = -a_1 x - c_1$$

$$y = -\frac{a_1}{b_1} x - \frac{c_1}{b_1}$$

$$\therefore b_{yx} = -\frac{a_1}{b_1}$$

2nd can be written as:

$$a_2 x = -b_2 y - c_2$$

$$x = -\frac{b_2}{a_2} y - \frac{c_2}{a_2}$$

$$\therefore b_{xy} = -\frac{b_2}{a_2}$$

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

$$= \sqrt{-\frac{b_2}{a_2} \times -\frac{a_1}{b_1}}$$

If $r < 1$, our assumption is correct [bwz $-1 \leq r \leq 1$].

If $r > 1$, then take the line of the reverse order and the actual value of the corr^n = reciprocal of the number obtained in above case.

Q. 2 lines of regr^n are $x + 2y - 5 = 0$ and $2x + 3y - 8 = 0$ and variance is $x = 12$. find $\bar{x}, \bar{y}, \bar{r}$ and regr^n coefficients and σ_y^2.

→ The given lines are

$$x + 2y - 5 = 0 \quad \text{—} (1)$$

$$2x + 3y - 8 = 0 \quad \text{—} (2)$$

(1) & (2) passes through $(\bar{x}, \bar{y})$

$$\bar{x} + 2\bar{y} - 5 = 0$$

$$2\bar{x} + 3\bar{y} - 8 = 0$$

Solving these eqns, we get

$$\bar{x} = 1 ; \ \bar{y} = 2.$$

Assuming that $x + 2y - 5 = 0$ is the regr^n line at y on x and $2x + 3y - 8 = 0$ is the regr^n line at x on y.

Rewriting the eqns as

$$y = \frac{-1}{2} x + \frac{5}{2} \text{ and}$$

$$x = \frac{-3}{2} y + 8.$$

Regr^n coefficients are:-

b_{yx} = coeff of x = $-1/2$

b_{xy} = coeff of y = $-3/2$

$\therefore r = \sqrt{b_{yx} \times b_{xy}} = \sqrt{-1/2 \times -3/2} = \sqrt{3/4}$

$r^2 = 0.75 \Rightarrow r = \pm 0.866$

Since b_{yx} & b_{xy} are $-ve$

$\therefore r = -0.866$

Since $b_{yx} = \frac{r\sigma_y}{\sigma_x}$ $\sigma_x^2 = 12$

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

$= \frac{\sqrt{12} \times -1/2}{-0.866} = 2.$

Q. From the following data, find the most likely value of y when $x = 24$.

	y	x
Mean	985.8	18.1
SD	36.4	2.0

$r = 0.58$

→ The regrsn eqn of y on x is

$y - \bar{y} = \frac{r\sigma_y}{\sigma_x}(x - \bar{x})$ — (1)

Given $\bar{x} = 18.1$ and $\bar{y} = 985.5$

$\sigma_x = 2$, $\sigma_y = 36.4$

$\therefore$ from (1), we get

$y - 985.5 = \frac{0.5 \times 36.4}{2}(x - 18.1)$

$\Rightarrow y = 10.556x + 794.74$

When $x = 24$;

$y = 10.556 x \times 24 + 794.74$

$y = 1048.084$

Q. From the following data find y when $x = 45$,

	x	y
Mean	53	142
SD	130	165

$\Sigma(x-\bar{x})(y-\bar{y}) = 1220$ and $n = 10$.

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

$r = \frac{1220/10}{130 \times 165}$

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

$\bar{x} = 53$
$\bar{y} = 142$
$\sigma_x = 130$
$\sigma_y = 165$
$x = 24$

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

$= 142 + \frac{122}{130 \times 165} \times \frac{165}{130}(24 - 53)$

$y = 134.5$

HW Q. Find the two regression lines of the following dates.