UMA 2452

Probability and statistics

G. Balakumar
3122 21 5002 019

1

i) Marginal distribution of x and y

$P_X(1) = P[X=1] = 5/36$
$P_X(2) = P[X=2] = 19/36$
$P_X(3) = P[X=3] = 1/3$

Marginal distribution of Y

$P_Y(1) = P[Y=1] = 1/4$
$P_Y(2) = P[Y=2] = 14/45$
$P_Y(3) = P[Y=3] = 79/180$

ii) Conditional distribution of X given Y=2

$$P(X/Y) = \frac{P_{XY}[x,y]}{P[Y=2]}$$

$$P[X=1, Y=2] = \frac{P_{XY}[1,2]}{P_Y[2]} = \frac{0}{14/45} = 0$$

$$P[X=2, Y=2] = \frac{P_{XY}[2,2]}{P_Y(2)} = \frac{1/9}{14/45} = \frac{5}{14}$$

$$P[X=3, Y=2] = \frac{P_{XY}(3,2)}{P_Y(2)} = \frac{1/5}{14/45} = \frac{9}{14}$$

iii) Conditional distribution of Y given $x=3$

$$P(Y/x) = \frac{P_{XY}(x,y)}{P_X(x=3)}$$

$$P[x=3, Y=1] = \frac{P_{XY}(3,1)}{P_X(3)} = \frac{0}{1/3} = 0$$

$$P[x=3, Y=2] = \frac{P_{XY}(3,2)}{P_X(3)} = \frac{1/5}{1/3} = 3/5$$

$$P[x=3, Y=3] = \frac{P_{XY}[3,3]}{P_X(3)} = 2/15 / 1/3 = 2/5$$

iv) $P(x \leq 2, y = 3)$

$$P[x \leq 2, y=3] = P_{XY}(1,3) + P_{XY}(2,3)$$

$$= 1/18 + 1/4$$

$$P[x \leq 2, y = 3] = 11/36$$

v) Find $P(Y \leq 2)$, $P(x+y<4)$

$$= P[Y \leq 2] = P_Y(1) + P_Y(2) = 1/4 + 14/45 = \frac{101}{180}$$

$$P[y \leq 2] = \frac{101}{180}$$

$$P(x+y<4) = P_{XY}(1,2) + P_{XY}(1,1) + P_{XY}(2,1)$$

$$= 0 + 1/12 + 1/6 = 1/4$$

$$P(x+y<4) = 1/4$$

2) The joint pdf of a two dimensional random variable is given by

$$f(x,y) = \begin{cases} 8xy, & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find marginal and conditional probability density function

$= f_y(x) =$ The marginal distribution of X

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

$$= \int_x^1 8xy\,dy$$

$$= 8x\left[\frac{y^2}{2}\right]_x^1$$

$$= \frac{8x}{2} - \frac{8x^3}{2} = 4x\,[1-x^2]$$

$$f_Y[y] = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$

$$= \int_0^y 8xy\,dx$$

$$= \left[\frac{x^2}{2}\right]_0^y 8y$$

$$= \frac{8y^3}{2} = 4y^3$$

$$f_X(x) \times f_Y(y) = 4x(1-x^2) \times 4y^3$$

$$= (4x - 4x^3)(4y^3)$$

$$= 16xy^3 - 16x^3y^3$$

$$\neq 8xy$$

$\therefore$ They are independent to each other

Conditional probability

$$P(x/y) = \frac{P(x \cap y)}{P(y)}$$

$$f_{x/y}(x,y) = \frac{8xy}{4y^3} = \frac{2x}{y^2}$$

$$f_{y/x}(x,y) = \frac{8xy}{4x(1-x)^2} = \frac{2x}{(1-x^2)}$$

Marginal distribution

$f_x(x) = 4x(1-x^2)$

$f_y(y) = 4y^3$

Conditional distribution

$f_{x/y}(x,y) = \frac{2x}{y^2}$

$f_{y/x}(x,y) = \frac{2y}{(1-x^2)}$ $\therefore$ X & y are not independs

---

(3.) $f(x,y) = \begin{cases} kxy\, e^{-(x^2+y^2)} & x>0,\ y>0 \\ 0 & \text{otherwise} \end{cases}$

i) Find k

$= \int_{-\infty}^{\infty} f(x,y) = 1 = \int_0^{\infty}\int_0^{\infty} kxy\, e^{-(x^2+y^2)}\, dx\, dy = 1$

$= k\left[\int_0^{\infty}\int_0^{\infty} y\, [e^{-(x^2+y^2)} x]\, dx\, dy\right] = 1$

$= k\left[\int_0^{\infty} x e^{-x^2} dx\right]\left[\int_0^{\infty} y e^{-y^2} dy\right] = 1$

$= \int_0^\infty x e^{-y^2} dx = \int_0^\infty \frac{e^{-t}}{2} dt \quad (x^2 = t, \; 2x\,dx = dt)$

$= \frac{1}{2}\left[\frac{e^{-t}}{-1}\right]_0^\infty = \frac{1}{2}[1] = \frac{1}{2}$

$= K[\frac{1}{2}][\frac{1}{2}] = 1 \quad \boxed{K = 4}$

ii) Prove x & y are independent

$f_x(x) = \int_0^\infty f_{xy}(x,y)\,dy$

$= \int_0^\infty 4xy e^{-(x^2+y^2)} dy$

$= 4x e^{-x^2} \int_0^\infty y e^{-y^2} dy$

$= 2x e^{-x^2}$

$f_y(y) = \int_0^\infty 4xy e^{-(x^2+y^2)} dx$

$= 4y e^{-y^2} \int_0^\infty x e^{-x^2} dx$

$= 4y e^{-y^2} [\frac{1}{2}]$

$= 2y e^{-y^2}$

$f(x) \times f_y(y) = (2x e^{-x^2})(2y e^{-y^2})$

$= 4xy e^{-(x^2+y^2)}$

$f_x(x) \times f_y(y) = f_{xy}(x,y)$

$\therefore$ x & y are independent

hence proved

4. Find co-eff of corellation & and obtain the regression lines.

= Coefficient of correllation: $r(x,y) = \frac{Cov(x,y)}{\sigma x \sigma y}$

lines of regression $y - \bar{y} = r \frac{\sigma y}{\sigma x}(x - \bar{x})$

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

x	62	64	65	69	70	71	72	74	68.37
y	126	125	139	145	165	152	180	200	155
x^2	3844	4096	4255	4761	4900	5041	5184	5476	4690
y^2	15876	15625	19321	21025	27725	23104	32400	43624	24730
xy	7812	8000	9035	10005	11500	10792	12960	15393	10693.

$E(x) = 68.375$ $\quad E(y) = 155$ $\quad E(xy) = 10693.25$

$E(x^2) = 4690.875$ $\quad E(y^2) = 24730$

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

$= 4690.875 - 4675.14$

$Var(x) = 15.7344$ $\quad Var(y) = 708$

$\boxed{\sigma_x = 3.966}$ $\quad \boxed{\sigma_y = 26.552}$

$Cov(x,y) = E(xy) - E(x)E(y)$

$= 10693.25 - 10598.125$

$= 95.155$

$$r(x,y) = \frac{\text{Cov}(x,y)}{\sigma x \sigma y} = \frac{93.125}{(3.966)(26.552)}$$

$$= 0.903$$

lines of regression

Y on x

$$y - 155 = 0.903 \left(\frac{26.552}{3.966}\right)(x - 68.375)$$

$$y = 6.043x + 155 - 413.326$$

$$\boxed{y = 6.045\, x - 258.32}$$

lines of regression

x on Y

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

$$x - 68.375 = 0.903 \left(\frac{3.961}{26.552}\right)(y - 155)$$

$$\boxed{x = 0.13487y + 47.486}$$

$$\boxed{r = 0.88}$$

5

x	25	28	35	32	31	36	39	38	34	32	32
y	43	46	49	41	36	32	31	30	33	39	38
x^2	625	784	1225	1024	961	1296	841	1444	1156	1024	1030
y^2	1849	2116	2401	1681	1296	1024	961	900	1089	1521	1483.8
xy	1075	1288	1715	1312	116	1152	899	1140	1122	1248	1206 →

Coeff of correlation $\boxed{r = \frac{Cov(x,y)}{\sigma_x \sigma_y}}$

$E(x) = 32 \quad E(y) = 38 \quad E(x^2) = 1038 \quad E(y^2) = 1483.8$

$E(xy) = 1206.7$

$var(x) = E(x^2) - [E(x)]^2 = 1038 - (32)^2$
$= 1038 - 1024$
$= 14$

$\boxed{\sigma_x = 3.742}$

$var(y) = E(y^2) - [E(y)]^2$

$= 1483.8 - 1444$

$= 39.8$

$\boxed{\sigma_y = 6.31}$

$Cov(x,y) = E(xy) - E(x)E(y)$

$= 1206.7 - 1216$

$= -9.3$

co-efficient of correlation $= r = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{-9.3}{(3.742)(6.31)}$

$\boxed{r = -0.39}$

lines of regression of Y on X

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

$y - 38 = \frac{(-0.39)(6.31)}{3.742} (x - 32)$

$$\boxed{y = -0.6576x + 59.045}$$

lines of regression of x on y

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

$$x - 32 = \frac{-(0.59)(3.742)}{6.31}[y - 38]$$

$$\boxed{x = -0.2313y + 40.79}$$

At $x = 30$ in

$y = 39.317 \approx 40$

$$\boxed{r = -0.39}$$

line of regression Y on x $\quad y = -0.6576x + 59.045$

line of regression x on Y $\quad x = -0.2313y + 40.79$

when economics mark is 30, statistics mark is 40

---

6. $f(x,y) = \begin{cases} \frac{1}{3}(x+y) & : 0 \le x \le 1 ; 0 \le y \le 2 \end{cases}$

Find the correlation coefficient and lines of regression

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

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

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

$$E(x) = 5/9$$

$$E(y) = \int_0^1 \int_0^2 y\left[\tfrac{1}{3}(x+y)\right] dx\,dy$$

$$= \tfrac{1}{3}\int_0^1 \int_0^2 (xy+y^2)\,dx\,dy = \tfrac{1}{3}\int_0^1 \left[\frac{xy^2}{2} + \frac{y^3}{3}\right]_0^2 dx$$

$$= \tfrac{1}{3}\int_0^1 \left(2x + \frac{8}{3}\right)dx = \tfrac{1}{3}\left[\frac{2x^2}{2} + \frac{8}{3}x\right]_0^1$$

$$E(y) = \tfrac{1}{3}\left[1 + \frac{8}{3}\right] = \tfrac{11}{9}$$

$$E(x^2) = \int_0^1 \int_0^2 x^2 \left(\tfrac{1}{3}(x+y)\right) dx\,dy$$

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

$$= \tfrac{1}{3}\left[\frac{2x^4}{4} + \frac{2x^3}{3}\right]_0^1 = \tfrac{1}{3}\left[\tfrac{1}{2} + \tfrac{2}{3}\right] = \tfrac{1}{3}\left[\tfrac{7}{6}\right]$$

$$E(x^2) = \tfrac{7}{18}$$

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

$$= \frac{1}{3}\left[\frac{8}{3}\left(\frac{x^2}{2}\right) + 4x\right]_0^1$$

$$E[y^2] = \frac{1}{3}\left[\frac{\cancel{8}\,4}{3}\left(\frac{1}{\cancel{2}}\right) + 4\right] = \tfrac{16}{9}$$

$$E(xy) = \int_0^1 \int_0^2 xy\left(\tfrac{1}{3}\right)(x+y)\,dx\,dy$$

$$= \int_0^1 \tfrac{1}{3}\left[\frac{x^2y^2}{2} + \frac{xy^3}{3}\right]_0^2 dx = \frac{1}{3}\int_0^1 \left(2x^2 + \frac{8x}{3}\right)dx$$

$$E(xy) = \frac{1}{3}\left[\frac{2}{3} + \frac{4}{3}\right] = \frac{6}{9} = \frac{2}{3}$$

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

$$= \frac{7}{18} - \left(\frac{5}{9}\right)^2 = \frac{7}{18} - \frac{25}{81} = \frac{13}{162}$$

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

$$= \frac{16}{9} - \left(\frac{11}{4}\right)^2 = \frac{23}{81}$$

$$\sigma_x = \sqrt{\frac{13}{162}} = 0.283 \qquad \sigma_y = \sqrt{\frac{23}{81}} = 0.532$$

Co-eff of correlation

$$r(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{-1/81}{(0.283)(0.532)} = -0.082$$

The co-efficient of correlation $= -0.082$ //

Regression line

Y on x

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

$$y - \frac{11}{9} = -(0.082)\left(\frac{0.532}{0.283}\right)\left(x - \frac{5}{9}\right)$$

$$\boxed{y = -0.154x + 1.307}$$

x on y

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

$$x - \frac{5}{9} = -(0.082)\left(\frac{0.282}{0.534}\right)\left(y - \frac{11}{9}\right)$$

$$\boxed{x = -0.043y + 0.655}$$

7. Calculate correlation coefficient for given heights

X	65	66	67	68	69	70
Y	67	68	65	68	72	72
X^2	4225	4489	4462	4761	4900	4900
Y^2	4484	4514	4123	4514	4030	4930
XY	4355	4763	4143	4331	4451	4457

$E(x) = 68$ $E(y) = 69$ $E(x^2) = 4628.5$ $E(y^2) = 4766.7$

$E(xy) = 4695$

$var(x) = E(x^2) - [E(x)]^2 = 4.6$

$Var(y) = E(y^2) - [E(y)]^2 = 5.5$

$\sigma x = \sqrt{4.6} = 2.141$ $\sigma y = \sqrt{5.5} = 2.347$

$Cov(x,y) = E(xy) - E(x)E(y)$

$= 3$

The co-efficient of correllation $= 0.596$.

8

x on y | y on x

$x - 4y = 5$ | $y - kx = 4$

$\bar{x} = 4y + 5$ | $\bar{y} \div k\bar{x} = 4$

$bxy = 4$ | $byx = k$

when $k = 1/16$

$\bar{x} - 4\bar{y} = 5$

$\frac{\bar{x}}{16} + \bar{y} = 4$

$12\bar{y} = 69$

$\bar{y} = 69/12$

$\boxed{y = 5.75}$

$\bar{x} = 5 + 4(5.75) = 5 \pm 2.3 = 28$

$\boxed{\bar{x} = 28}$

here

$r^2 = 4k = 4\left[\frac{1}{16}\right] = 1/4$

$\boxed{r = \pm 1/2}$ since bxy and byx are +ve

$\boxed{r = 1/2 = 0.5}$

The co-efficient of correlation $= 0.5$

---

9) Given

$3\bar{x} + 12\bar{y} = 19$ → ①

$9\bar{x} + 3\bar{y} = 46$ → ②

① × 3 ⇒ $9\bar{x} + 36\bar{y} = 57$

② ⇒ $9\bar{x} + 3\bar{y} = 46$

$33\bar{y} = 11$

$\bar{y} = 11/33$

$\boxed{\bar{y} = 1/3}$

$3\bar{x} + 12\left[\frac{11}{33}\right] = 19$

$3\bar{x} = 15$

$\boxed{\bar{x} = 5}$

line of regression of Y on X

$12y = -3x + 19$

$y = \frac{-3}{12}x + \frac{19}{12} \Rightarrow b_{yx} = -3/2$

line of regression of X on Y

$9x + 3y = 46$

$x = \frac{-3}{9}y + 46 \Rightarrow b_{xy} = -1/3$

$r^2 = b_{yx} \times b_{xy} = (-1/4)(-1/3) = \frac{7}{12}$

$\boxed{r = \pm\ 0.289}$

since b_{yx} and b_{xy} are negative r is negative

$\boxed{r = -0.289}$

$Var(y) = 3 \Rightarrow \sigma_y = \sqrt{3} \Rightarrow \sigma_y^2 = 3$

Now

$\frac{b_{yx}}{b_{xy}} = \frac{\sigma_y^2}{\sigma_x^2} = \frac{-1/4}{-1/3} = \frac{3}{\sigma_x^2} = \frac{3}{4}$

$\sigma_x^2 = 4$

$\boxed{Var(x) = 4}$

$E(x) = 5$

$E(y) = 1/3$

$r(x,y) = -0.289 - 1/\sqrt{12}$ $\quad$ $Var(x) = 4$

---

(10)

$8x - 10y + 66 = 0$

$40x - 18y - 214 = 0$

=

Line of regression Y on X

$8x - 10y + 66 = 0$ $\quad$ $y = \frac{8}{10}x + \frac{66}{16}$

$b_{yx} = 4/5$

Line of regression X on Y

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

$b_{xy} = \frac{18}{40} = \frac{9}{20}$

$r^2 = b_{yx} * b_{xy} = \left(\frac{4}{5}\right)\left(\frac{9}{20}\right) = \frac{36}{100} = 0.36$

Since b_{xy} and b_{yx} are (+ve) r is also positive

$\boxed{r = 0.6}$

(1) ×5 $\quad 40\bar{x} - 50\bar{y} + 330 = 0$

(2) ×1 $\quad 40\bar{x} - 18\bar{y} - 214 = 0$

$-32y + 544 = 0$

$\boxed{\bar{y} = 17}$

Sub $\bar{y}$ in ①

$$8\bar{x} - 170 + 66 = 0$$

$$8\bar{x} = 104$$

$$\bar{x} = 13$$

Now

$$b_{yx} = \frac{\sigma_y^2}{\sigma_x^2} = \frac{4/5}{9/20} = \frac{\sigma_y^2}{9}$$

$$\sigma_y^2 = \frac{4}{5} \times \frac{20}{9} \times 9 = 16$$

Variance of $Y = 16$

$E(x) = 13$

$E(y) = 17$

$var(y) = 16$

$r(x,y) = 0.6$