SNS COLLEGE OF TECHNOLOGY
(An Autonomous Institution)
Coimbatore – 641 035
DEPARTMENT OF MATHEMATICS
Covariance,Correlation and Regression

Covariance:

* If x and y are two dimensional random variable, then covariance of x and y is defined as

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

* If x and y are independent, then

$$E(xy) = E(x)\cdot E(y)$$

$$\Rightarrow \mathrm{cov}(x, y) = E(x)E(y) - E(x)E(y)$$
$$= 0$$

$\therefore$ It is uncorrelated.

Result:

1. $\mathrm{cov}(ax, by) = ab\,\mathrm{cov}(x, y)$
2. $\mathrm{cov}(ax+b, cy+d) = ac\,\mathrm{cov}(x, y)$

Correlation:

$$r_{xy} = \frac{\mathrm{cov}(x, y)}{\sigma_x\,\sigma_y}$$

$\sigma \rightarrow$ Standard Deviation
$\sigma \rightarrow \sqrt{\text{variance}}$

(or)

$$r_{xy} = \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}\,x}\,\sqrt{\mathrm{var}\,y}}$$

Regression:

Regression is the average relationship between two or more variables

Regression line:

x on y	y on x
$x-\bar{x} = b_{xy}(y-\bar{y})$	$y-\bar{y} = b_{yx}(x-\bar{x})$
$b_{xy} = r\,\frac{\sigma_x}{\sigma_y}$	$b_{yx} = r\,\frac{\sigma_y}{\sigma_x}$

Properties :

* $\bar{x} = \frac{\Sigma x}{n}$ and $\bar{y} = \frac{\Sigma y}{n}$
* correlation coefficient : $r = \pm\sqrt{b_{xy} \cdot b_{yx}}$
* Regression coefficient : b_{xy} and b_{yx}.
* ~~Correlation coefficient~~ —

1. Let X and Y be discrete random variable with probability mass function $P(x, y) = \frac{x+y}{21}$, find correlation coefficient. $x = 1, 2, 3 ; y = 1, 2$

Soln:

X \ Y	1	2	P(X)
1	2/21	3/21	5/21
2	3/21	4/21	7/21
3	4/21	5/21	9/21
P(Y)	9/21	12/21	1

$$r = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{E(xy) - E(x)E(y)}{\sigma_x \sigma_y}$$

$$E(X) = \Sigma x P(x)$$

$$= 1\left(\frac{5}{21}\right) + 2\left(\frac{7}{21}\right) + 3\left(\frac{9}{21}\right)$$

$$= \frac{5+14+27}{21}$$

$$E(X) = \frac{46}{21}$$

$$E(Y) = \Sigma y P(y)$$

$$= 1\left(\frac{9}{21}\right) + 2\left(\frac{12}{21}\right) = \frac{9+24}{21}$$

$$= \frac{33}{21}$$

$$E(xy) = \sum\sum xy\, P(x,y)$$

$$= 1(1)\left(\frac{2}{21}\right) + 1(2)\left(\frac{3}{21}\right) + 2(1)\left(\frac{3}{21}\right) + 2(2)\left(\frac{4}{21}\right) + 3(1)\left(\frac{4}{21}\right) + 3(2)\left(\frac{5}{21}\right)$$

$$= \frac{2+6+6+16+12+30}{21}$$

$$E(xy) = \frac{72}{21}$$

$$E(x^2) = \sum x^2 P(x)$$

$$= 1^2\frac{5}{21} + 2^2\frac{7}{21} + 3^2\frac{9}{21}$$

$$= \frac{5+28+81}{21}$$

$$E(x^2) = \frac{114}{21}$$

$$\operatorname{var}(x) = E(x^2) - [E(x)]^2$$

$$= \frac{114}{21} - \left(\frac{46}{21}\right)^2$$

$$= \frac{114}{21} - \frac{2116}{441}$$

$$\operatorname{Var}(x) = \frac{278}{441}$$

$$E(y^2) = \sum y^2 P(y)$$

$$= 1\left(\frac{9}{21}\right) + 4\left(\frac{12}{21}\right)$$

$$= \frac{9+48}{21}$$

$$= \frac{57}{21}$$

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

$$= \frac{57}{21} - \left(\frac{33}{21}\right)^2$$

$$= \frac{57}{21} - \frac{1089}{441}$$

$$Var(Y) = \frac{108}{441}$$

$$cov(X, Y) = E(XY) - E(X)E(Y)$$

$$= \frac{72}{21} - \frac{46}{21}\left(\frac{33}{21}\right)$$

$$= \frac{72}{21} - \frac{1518}{441}$$

$$= \frac{1512 - 1518}{441}$$

$$cov(X, Y) = \frac{-6}{441}$$

$$r = \frac{cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

$$= \frac{-6/441}{\sqrt{\frac{278}{441}}\sqrt{\frac{108}{441}}}$$

$$= \frac{-0.014}{0.393}$$

$$r = -0.036$$

2]. Suppose that the Two Dimensional R. Vs. (X, Y) has the joint Pdf

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

i). Obtain the correlation coefficient b/w X & Y.
ii). check whether X & Y are independent.

Soln.

MDF of X:

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$= \int_0^1 (x+y)\, dy$$

$$= \left(xy + \frac{y^2}{2}\right)_0^1$$

$$f(x) = x + \frac{1}{2}, \quad 0<x<1$$

MDF of Y:

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

$$= \int_0^1 (x+y)\, dx$$

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

$$f(y) = y + \frac{1}{2}, \quad 0<y<1$$

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

$$= \int_0^1 x\left(x+\frac{1}{2}\right)dx$$

$$= \int_0^1 \left(x^2+\frac{1}{2}x\right)dx$$

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

$$= \frac{1}{3}+\frac{1}{4}-0$$

$$E(x) = \frac{7}{12}$$

$$E(y) = \int_{-\infty}^{\infty} y\,f(y)\,dy$$

$$= \int_0^1 y\left(y+\frac{1}{2}\right)dy$$

$$= \int_0^1 \left(y^2+\frac{1}{2}y\right)_0^1 dy$$

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

$$= \frac{1}{3}+\frac{1}{4}$$

$$E(y) = \frac{7}{12}$$

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

$$= \int_0^1\int_0^1 xy\,(x+1)\,dx\,dy$$

$$= \int_0^1\int_0^1 (x^2y+xy)\,dx\,dy$$

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

$$= \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2} \right) dy$$

$$= \left(\frac{y^2}{6} + \frac{y^3}{6} \right)_0^1$$

$$= \frac{2}{6}$$

$$E(xy) = \frac{1}{3}$$

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

$$= \int_0^1 x^2 \left(x + \frac{1}{2} \right) dx$$

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

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

$$= \frac{1}{4} + \frac{1}{6} - 0$$

$$E(x^2) = \frac{10}{24}$$

$$E(y^2) = \int_{-\infty}^{\infty} y^2 F(y)\, dy = \int_0^1 y^2 \left(y + \frac{1}{2} \right) dy$$

$$= \int_0^1 \left(y^3 + \frac{1}{2} y^2 \right) dy$$

$$= \left(\frac{y^4}{4} + \frac{1}{2} \frac{y^3}{3} \right)_0^1$$

$$= \frac{1}{4} + \frac{1}{6} - 0$$

$$= \frac{10}{24}$$

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$= \frac{1}{3} - \frac{7}{12}\left(\frac{7}{12}\right)$$

$$= \frac{-1}{144} \neq 0$$

$\therefore$ x and y are dependent.

$$\sigma_x^2 = \text{Var}(X) = E(X^2) - \{E(X)\}^2$$

$$= \frac{10}{24} - \left(\frac{7}{12}\right)^2$$

$$= \frac{10}{24} - \frac{49}{144}$$

$$\sigma_x^2 = \frac{11}{144}$$

$$\sigma_x = \frac{\sqrt{11}}{12}$$

$$\sigma_y^2 = E(y^2) - [E(y)]^2$$

$$= \frac{5}{12} - \left(\frac{7}{12}\right)^2$$

$$\sigma_y^2 = \frac{11}{144}$$

$$\sigma_y = \frac{\sqrt{11}}{12}$$

$$\therefore r = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{-1/144}{\frac{\sqrt{11}}{12} \frac{\sqrt{11}}{12}} = \frac{-1}{11}$$

coefficient of correlation [If the interval is given]

$$r(x,y) = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$$

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

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} \qquad \bar{y} = \frac{\Sigma y}{n}$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2}$$

1. Calculate the correlation coefficient for the following heights (in inches) of father's x and their son's y.

x : 65 66 67 67 68 69 70 72

y : 67 68 65 68 72 72 69 71

Soln.

x	y	xy	x^2	y^2
65	67	4355	4225	4489
66	68	4488	4356	4624
67	65	4355	4489	4225
67	68	4556	4489	4624
68	72	4896	4624	5184
69	72	4968	4761	5184
70	69	4830	4900	4761
72	71	5112	5184	5041
$\Sigma x = 544$	$\Sigma y = 552$	$\Sigma xy = 37560$	$\Sigma x^2 = 37028$	$\Sigma y^2 = 38132$

Here $n=8$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{544}{8} = 68$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{552}{8} = 69 \quad \text{and} \quad \bar{x}\,\bar{y} = 68(69) = 4692$$

$$\operatorname{cov}(x,y) = \frac{\Sigma xy}{n} - \bar{x}\,\bar{y} = \frac{37560}{8} - 4692$$

$$\operatorname{cov}(x,y) = 3$$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\frac{37028}{8} - (68)^2}$$

$$= \sqrt{4628.5 - 462.4}$$

$$\sigma_x = 2.121$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2} = \sqrt{\frac{38132}{8} - (69)^2}$$

$$= \sqrt{4766.5 - 4761}$$

$$\sigma_y = 2.345$$

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

$$r = 0.6032$$