unit IV
07 February 2022 ~~10:03~~

Correlation & Regression

Covariance:

If x and y are rv's then the Covariance b/w them is

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

Properties:

1. If x and y are independent, then $Cov(x,y) = 0$.
2. $Cov(ax, by) = ab\, Cov(x,y)$
3. $Cov(x+a, y+b) = Cov(x,y)$
4. $V(x_1+x_2) = V(x_1) + V(x_2) + 2\,Cov(x_1, x_2)$
5. $V(x_1-x_2) = V(x_1) + V(x_2) - 2\,Cov(x_1, x_2)$

Correlation:

Karl Pearson's Coefficient of Correlation:

Let x & y be given r.v. Then Karl Pearson's Coefficient of Correlation is denoted by r_{xy} (or) $r(x,y)$.

$$r_{xy} = r(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}} = \frac{Cov(x,y)}{\sigma_x\,\sigma_y}$$

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

$$= \frac{1}{n}\sum xy - \bar{x}\,\bar{y}, \quad \bar{x} = \frac{\sum x}{n} \quad \bar{y} = \frac{\sum y}{n}$$

'n' - no. of items.

$$\sigma_x^2 = Var(x) = \frac{1}{n}\sum x^2 - (\bar{x})^2$$

$$\sigma_y^2 = Var(y) = \frac{1}{n}\sum y^2 - (\bar{y})^2$$

Note: Correlation Coefficient always lies b/w -1 to $+1$.

Problem 1: Calculate the Correlation Coefficient for the following height (in inches) of father (x) and their son 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
544	552	37560	37028	38132

$\bar{x} = \frac{544}{8} = 68$ $\quad \bar{y} = \frac{552}{8} = 69$ $\quad \bar{x}\bar{y} = 68 \cdot 69 = 4692$

$\sigma_x = \sqrt{\frac{1}{n}\Sigma x^2 - (\bar{x})^2}$ $\quad \sigma_y = \sqrt{\frac{1}{n}\Sigma y^2 - (\bar{y})^2}$

$= \sqrt{\frac{1}{8}(37028) - (68)^2}$ $\quad = \sqrt{\frac{1}{8}(38132) - (69)^2}$

$\sigma_x = 2.121$ $\quad \sigma_y = 2.345$

$Cov(x,y) = \frac{1}{n}\Sigma xy - \bar{x}\bar{y}$

$= \frac{1}{8}(37560) - 68 \cdot 69$

$= 4695 - 4692$

$= 3.$

The Correlation Coefficient of x & y is given by

$r_{xy} = r(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{3}{(2.121)(2.345)} = \frac{3}{4.973}$

$= 0.6032$

$r_{xy} = 0.6032.$

Rank Correlation:

If (x_i, y_i) $i = 1, 2, \ldots n$ are the rank of the individuals in two characteristics A & B respectively, then the rank Correlation Coefficient is given by

$$\rho = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2$$

$d_i = x_i - y_i$ $\quad n \to$ no of items.

Karl Pearson's formula for rank Correlation Coefficient.

Problem Find the rank Correlation coefficient from the following data.

Rank in X	1	2	3	4	5	6	7
Rank in Y	4	3	1	2	6	5	7

X	Y	$d_i = x_i - y_i$	d_i^2
1	4	−3	9
2	3	−1	1
3	1	2	4
4	2	2	4
5	6	−1	1
6	5	1	1
7	7	0	0
			20

$n = 7$

∴ rank Correlation Coefficient is

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6(20)}{7(49-1)}$$

$$= 1 - \frac{120}{7\times 48}$$

$$= 1 - \frac{5}{14} = 9/14 = 0.6428$$

Problem: 10 Participants were ranked according to their performance in a musical test by 3 judges in the following order.

Rank by X:	1	6	5	10	3	2	4	9	7	8
" " Y:	3	5	8	4	7	10	2	1	6	9
Z:	6	4	9	8	1	2	3	10	5	7

Use rank Correlation method & discuss which pair of Judges has the nearest approach to common likings of music.

Soln

$$\rho(X,Y) = 1 - \frac{6\sum d_1^2}{n(n^2-1)} \qquad \rho(Z,X) = 1 - \frac{6\sum d_3^2}{n(n^2-1)}$$

$$\rho(Y,Z) = 1 - \frac{6\sum d_2^2}{n(n^2-1)} \qquad \rho(Z,X) = 1 - \frac{6\sum d_3^2}{n(n^2-1)}$$

X	Y	Z	$d_1 = x-y$	d_1^2	$d_2 = y-z$	d_2^2	$d_3 = z-x$	d_3^2
1	3	6	-2	4	-3	9	5	25
6	5	4	1	1	1	1	-2	4
5	8	9	-3	9	-1	1	4	16
10	4	8	6	36	-4	16	-2	4
3	7	1	-4	16	6	36	-2	4
2	10	2	-8	64	8	64	0	0
4	2	3	2	4	-1	1	-1	1
9	1	10	8	64	-9	81	1	1
7	6	5	1	1	1	1	-2	4
8	9	7	-1	1	2	4	-1	1
				$\sum d_1^2 = 200$		$\sum d_2^2 = 214$		$\sum d_3^2 = 60$

Here $n = 10$.

$$\rho(X,Y) = 1 - \frac{\cancel{6}^{2}(200)}{\cancel{10}\times\cancel{99}_{33}} = 1 - \frac{40}{33} = \frac{33-40}{33} = \frac{-7}{33} = -0.212$$

$$\rho(Y,Z) = 1 - \frac{\cancel{6}^{1}(214)}{\cancel{10}_{5}\times\cancel{99}_{33}} = 1 - \frac{214}{165} = \frac{165-214}{165} = \frac{-49}{165} = -0.296$$

$$\rho(Z,X) = 1 - \frac{\cancel{6}^{2}(\cancel{60}^{\cancel{30}^{\cancel{2}}})}{\cancel{10}\times\cancel{99}_{\cancel{33}_{11}}} = 1 - \frac{4}{11} = \frac{7}{11} = 0.6363$$

Since $\rho(X,Y)$ & $\rho(Y,Z)$ are negative & $\rho(Z,X)$ is positive

Z, X Judges has the nearest approach to common likings & music.