Correlation & Regression

Date: 9/9/25

Tutorial-5

1. Compute Karl Pearson's coefficient of correlation betn x & y for following data:

x	10	14	18	22	26	30
y	18	12	24	6	30	36

Soln

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}} \longrightarrow (1)$$

$n = 6$

	x	y	xy	x^2	y^2
	10	18	180	100	324
	14	12	168	196	144
	18	24	432	324	576
	22	6	132	484	36
	26	30	780	676	900
	30	36	1080	900	1296
Σ	120	126	2772	2680	3276

From eqn (1)

$$r = \frac{2772 - \frac{120(126)}{6}}{\sqrt{(2680)^2 - \frac{(120)^2}{6}}\sqrt{3276 - \frac{(126)^2}{6}}}$$

Page No.:
Date:

$$= \frac{2772 - 2520}{\sqrt{2680 - 2400}\sqrt{3276 - \cancel{7938}2646}}$$

$$= \frac{252}{\sqrt{16.7332}\sqrt{25.099}}$$

$$= \frac{252}{16.7332 \times 25.0998}$$

$$= \frac{252}{419.999}$$

$$\boxed{r = 0.6}$$

2 Find coefficient of correlation betn x & y if regression lines are: $x + 6y = 6$ & $3x + 2y = 10$.

Solⁿ Let regression lines are $x + 6y = 6$ and $3x + 2y = 10$

Regression of y on x

$$6y = 6 - x$$

$$y = \frac{1}{6}[6 - x]$$

$$\boxed{b_{yx} = -\frac{1}{6}}$$

Regression of x on y

$$3x + 2y = 10$$

$$x = \frac{1}{3}[10 - 2y]$$

Page No.:
Date:

$$\boxed{b_{xy} = -\frac{2}{3}}$$

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

$$= \sqrt{-\frac{1}{6} \times -\frac{2}{3}}$$

$$= \sqrt{\frac{2}{18}}$$

$$= \sqrt{\frac{1}{9}}$$

$$\boxed{r = -\frac{1}{3}}$$

Since b_{yx} & b_{xy} are negative, so r is also negative.

3. In a partially destroyed record on analysis of correlation data, only the following are legible: Variance of x, $\sigma_x^2 = 9$, regression eq^n $8x - 10y + 66 = 0$, $40x - 18y = 214$. Find (i) mean values of x & y (ii) the standard deviation of y

Sol^n Let regression lines

$$8x - 10y = -66 \quad \times 5$$
$$40x - 18y = 214$$

The regression curve lines passing through $(\bar{x}, \bar{y})$

$$40\bar{x} - 50\bar{y} = -330$$
$$40\bar{x} - 18\bar{y} = +214$$
$$- \qquad + \qquad -$$
$$-32\bar{y} \cancel{68\bar{y}} = -544$$

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

$$4 \quad \bar{x} = \frac{-330 + 50\bar{y}}{40} = \frac{-330 + 50(17)}{40}$$

$$\boxed{\bar{x} = 13}$$

The sample means $\bar{x} = 13$, $\bar{y} = 17$

Let line $8x - 10y = -66$ be regression of y on x

$$\cancel{6y =} \quad -10y = -66 - 8x$$

$$y = \frac{-(66 + 8x)}{-10}$$

$$= \frac{-[66 + 8(13)]}{-10}$$

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

ii) $\sigma_x^2 = 9 \Rightarrow \sigma_x = 3$

$$y = \frac{-66 - 8x}{10}$$

$$b_{yx} = \frac{8}{10}$$

$$\therefore \boxed{b_{yx} = 0.8}$$

$$x = \frac{1}{40}(214 - 18y)$$

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

$$\therefore \boxed{b_{xy} = 0.45}$$

$$r = \sqrt{byx \cdot bxy}$$

$$= \sqrt{0.8 \times 0.45}$$

$$= \sqrt{0.36}$$

$$\boxed{r = 0.6}$$

For standard deviation Y,

$$byx = r \frac{\sigma_y}{\sigma_x}$$

$$\sigma_y = \frac{byx \cdot \sigma_x}{r}$$

$$= \frac{(0.8)(3)}{0.6}$$

$$\therefore \boxed{\sigma_y = 4}$$

4

4. Ten competitors in a contest are ranked by 3 judges in the following order. Use the correlation coefficient to determine which pair of judges has the nearest approach.

1st Judge	1	6	5	10	3	2	4	9	7	8
2nd	3	5	8	4	7	10	2	1	6	9
3rd	6	4	9	8	1	2	3	10	5	7

A. Let 1st judge = x, 2nd judge = y, 3rd judge = z

x	y	z	$dx = x - y$	d^2x	$dy = y - z$	d^2y	$dz - x$	d^2z
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
55	55	55	0	200	0	214	0	60

$$r_1(x,y) = 1 - \frac{6 \sum d^2x}{10((10)^2 - 1)}$$

$$= 1 - \frac{6[200]}{10[99]}$$

Page No.:

Date:

$$\frac{990 - 1200}{990}$$

$$\boxed{r(x,y) = -0.2121}$$

$$r(y,z) = 1 - \frac{6 \Sigma d_y^2}{10[(10)^2 - 1]}$$

$$= 1 - \frac{6[214]}{990}$$

$$= \frac{990 - 1284}{990}$$

$$\boxed{r(y,z) = -0.2969}$$

$$r(z,x) = 1 - \frac{6 \Sigma d_z^2}{10((10)^2 - 1)}$$

$$= \frac{990 - 6[60]}{990}$$

$$\boxed{r(z,x) = 0.6363}$$

Since $r(z,x)$ is maximum the pair of judge A & C has nearest common approach.

Page No.:

Date:

5 Calculate Pearson's coefficient of skewness for the following data:

x	12	17	22	27	32
f	28	42	54	108	129

Soln Mean

$\mu = E(x)$

$= \sum x P(x)$

$= 12(28) + 17(42) + 22(54) + 27(108) + 32(129)$

$= 336 + 714 + 1188 + 2916 + 4128$

$\therefore \boxed{\mu = 9282}$

$$\sigma = \sqrt{\frac{2670}{5} - \left(\frac{110}{5}\right)^2}$$

$$= \sqrt{534 - 484}$$

$$= \sqrt{50}$$

$$\boxed{\sigma = 7.07106}$$

mode = 32

$$S_K = \frac{9282 - 32}{7.07106}$$

$$\boxed{S_K = 1308.148}$$

Page No.:
Date.

6. Calculate the coefficient of correlation betn x & y for following data:

x	65	66	67	67	68	69	70	72
y	67	68	65	68	72	72	69	71

Solⁿ $r = \dfrac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}$

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	5712	5184	5041
544	552	37560	~~32539~~ 37028	38134 38134

$r = \dfrac{37560 - (544(552))/8}{\sqrt{\cancel{37028} - \frac{(544)^2}{8}}\sqrt{\cancel{38134} - \frac{(552)^2}{8}}}$

$= \dfrac{262728}{(-66.\cancel{730}) \times (-9.53939)}$

$= \dfrac{\cancel{262728}}{(\cancel{36.537})} \quad \dfrac{262728}{(-6)\times(6.7823)}$

Page No.:
Date:

$$= \frac{37560 - 37536}{\sqrt{37028 - \frac{(544)^2}{8}} \cdot \sqrt{38134 - \frac{(552)^2}{8}}}$$

$$= \frac{+24}{\sqrt{36}\ \sqrt{46}}$$

$$= \frac{24}{6 \times 6.7823}$$

$$= \frac{24}{40.6938}$$

$$\boxed{r = 0.5897}$$

7.) The following are the lines of regression $9y = x + 288$ and $4y = x + 38$. Estimate y when $x = 99$ & x when $y = 30$. Also find the means of x & y.

→ Let regression lines $9y = x + 288$ and $4y = x + 38$

$9y - x = 288$

$4y - x = 38$

Regression curve passing through $(\bar{x}, \bar{y})$

$$9\bar{y} - \cancel{\bar{x}} = 288$$
$$4\bar{y} - \cancel{\bar{x}} = 38$$

$$5\bar{y} = 250$$

$$\boxed{\bar{y} = 50}$$

Page No.:

Date:

$$\bar{x} = 9\bar{y} - \bar{x} = 288$$

$$9(50) - \bar{x} = 288$$

$$450 - 288 = \bar{x}$$

$$\boxed{162 = \bar{x}}$$

Mean, $\bar{x} = 162$ & $\bar{y} = 50$

When $x = 99$

$$4y = 99 + 38$$

$$4y = 137$$

$$\boxed{y = 34.25}$$

$y = 30$

$$x = 9y - 288$$

$$= 9(30) - 288$$

$$\boxed{x = -18}$$

8) For a group of 10 items, $\Sigma x = 452$, $\Sigma x^2 = 24270$ & mode $= 43.7$. Find Karl Pearson's coefficient of skewness

$n = 10$

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}$$

$\Sigma x = 452 \implies$ mean $= \frac{452}{10} = 45.2$

$\Sigma x^2 = 24270$

mode $= 43.7$

Page No.:
Date:

$$S_k = \frac{\text{mean} - \text{mode}}{\sigma}$$

$$\sigma^2 = \frac{\Sigma x^2}{n} - x^2$$

$$= \frac{24270}{10} - (45.2)^2$$

$$\sigma^2 = 383.96$$

$$\boxed{\sigma = 19.594}$$

$$S_k = \frac{45.2 - 43.7}{19.594}$$

$$\boxed{S_k = 0.07655}$$

Ex: 9) Find the correlation coefficient for the following data:

X	-3	-2	-1	0	1	2	3
Y	9	4	1	0.5	1	4	9

→

$$r = \frac{\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}$$

x	y	xy	x^2	y^2
-3	9	-27	9	81
-2	4	-8	4	16
-1	1	-1	1	1
0	0.5	0	0	0.25
1	1	1	1	1
2	4	8	4	16
3	9	27	9	81
Σx 0	28.5	0	28	196.25

$\boxed{r = 0}$

10 Calculate regression coefficients & find the two lines of regression for following data: Find value of y when $x = 65$

x	57	58	59	59	60	61	62	64
y	67	68	65	68	72	72	69	71

$$b_{yx} = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2} \rightarrow (1)$$

$$b_{xy} = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (y - \bar{y})^2} \rightarrow (2)$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{480}{8} = 60$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{552}{8} = 69$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
57	67	−3	−2	9	4	6
58	68	−2	−1	4	1	2
59	65	−1	−4	1	16	4
59	68	−1	−1	1	1	1
60	72	0	3	0	9	0
61	72	1	3	1	9	3
62	69	2	0	4	0	0
64	71	4	2	16	4	8
480	552	0	0	36	44	24

$$b_{yx} = \frac{24}{36}$$
$$= 0.667$$

$$b_{xy} = \frac{24}{44}$$
$$= 0.5454$$

$$r = \sqrt{b_{yx} \cdot b_{xy}}$$
$$= \sqrt{0.66 \times 0.545}$$

$$\boxed{r = 0.6029}$$

Line of regression y on x

Eq^n of regression of line

$$y-\bar{y} = byx\,[x-\bar{x}]$$
$$y-69 = 0.667x - 40.02$$
$$\boxed{y = 0.667x + 28.98} \rightarrow ③$$

Eq^n of regression

$$x-\bar{x} = bxy\,[y-\bar{y}]$$
$$x-60 = 0.545\,[y-69]$$
$$x = 0.545y + 22.395$$

From ③

$$y = 0.667(65) + 28.98$$
$$\boxed{y = 72.335}$$

11. Two judges in a beauty contest rank the 12 contestants as follows

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

Calculate rank correlation coefficient

Page No.:

Date:

x	y	d	d^2
1	12	-11	121
2	9	-7	49
3	6	-3	9
4	10	-6	36
5	3	2	4
6	5	1	1
7	4	3	9
8	7	1	1
9	6	~~8~~ 3	9
10	2	8	64
11	11	0	0
12	1	11	121
			424

$$r = 1 - \frac{6\,\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6[424]}{12[144-1]}$$

$$= 1 - \frac{2544}{1716}$$

$$= \frac{1716 - 2544}{1716}$$

$$\boxed{r = -0.482517}$$

(12)

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	y^2	xy
23	5	-33.3	-2.5	83.25	25	115
43	6	-13.33	-1.5	19.995	36	258
53	7	-3.33	0	1.665	49	371
63	8	6.67	0.5	3.335	64	504
73	9	16.67	1.5	25.005	81	657
83	10	26.67	2.5	66.675	100	830
338	45			199.925	355	2735

$n = 6$

$$\bar{x} = \frac{338}{6} = 56.33$$

$$\bar{y} = \frac{45}{6} = 7.5$$

$$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}$$

$$= \frac{2735 - \frac{338(45)}{6}}{355 - \frac{(45)^2}{6}}$$

$$= \frac{1200}{105}$$

$$= 11.4287$$

Page No.:
Date:

Eqⁿ of regression line of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 56.33 = 11.429(y - 7.5)$$

$$x = 11.429y - 29.387$$

Estimated performance y = 11 is

$$x = 11.429(11) - 29.387$$

$$\boxed{x = 96.332}$$

25/9/25