4 CORRELATION & REGRESSION

- Correlation refers to the relationship b/w two or more variables.
- Regression establishes a func^{n} relationship b/w the variables
- correlation is classified into 4 types:-
  1. +ve & -ve correlation
  2. Simple & Multiple correlation
  3. Partial & Total correlation
  4. Linear & Nonlinear correlation
- There are two methods of studying correlation
  - → Graphical
  - → Mathematical
- correlation coefficient: $-1 \le r \le 1$
- Karl Pearson's coefficient of correlation.

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

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

[or]

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

- correlation co-efficient is independent of change of origin & change of scale.
- Two independent variables are uncorrelated. (Converse of the property is not true).

Q] Calc. correlation coefficient.

x	5	9	13	17	21
y	12	20	25	33	35

Given: $n = 5$, $\bar{x} = 13$, $\bar{y} = 25$

$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
-8	-13	64	169	104
-4	-5	16	25	20
0	0	0	0	0
4	8	16	64	32
8	10	64	100	80
Σ 0	0	160	358	236

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

$$r = \frac{236}{\sqrt{160}\sqrt{358}}$$

$$\therefore r = 0.986$$

Q] Calc. correlation coefficient

x	2	4	5	6	8	11
y	18	12	10	8	7	5

x	y	x^2	y^2	xy
2	18	4	324	36
4	12	16	144	48
5	10	25	100	50
6	8	36	64	48
8	7	64	49	56
11	5	121	25	55
Σ 36	60	266	706	293

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

$$r = \frac{293 - 2660/6}{\sqrt{266 - \frac{(86)^2}{6}}\sqrt{186 - \frac{(60)^2}{6}}}$$

$\therefore r = 0.9203$

Q] of x is 16. Find S.D. of y

$\because r = \frac{cov(x,y)}{\sigma_x \sigma_y}$

Given $r = 0.48$

$cov(x,y) = 36$

$var_x = 16$

$SD\ x = \sqrt{16} = 4$

$0.48 = \frac{36}{4 \cdot \sigma_y}$

$\sigma_y = \frac{36}{0.48 \cdot 4}$

$\sigma_y = \frac{36}{0.48 \cdot 4}$

$= 18.75$

Q] Calc. coefficient of correlation

x	17	19	21	26	20	28	26	29
y	23	27	25	26	27	25	30	33

Given: $n = 8$, $a = 23$, $b = 27$

$dx = x - a = x - 23$

$dy = y - 27$

x	y	dx	dy	dx^2	dy^2	dxdy
17	23	-6	-4	36	16	24
19	27	-4	0	16	0	0
21	25	-2	-2	4	4	4
26	26	3	-1	9	1	-3
20	27	-3	0	9	0	0
28	25	5	-2	25	4	-10
26	30	3	3	9	9	9
29	33	6	6	36	36	36
Σ		0	0	124	70	48

$r = \frac{\Sigma dxdy - \frac{\Sigma dx \Sigma dy}{n}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{n}}\sqrt{\Sigma dy^2 - \frac{\Sigma dy)^2}{n}}}$

$r = \frac{48 - 0}{\sqrt{124 - 0}\sqrt{70 - 0}}$

$\therefore r = 0.515$

Q] Given $n = 25$

$\Sigma x = 125$ corrected $\Sigma x = 125 - (6+8) + (8+6) = 125$

$\Sigma x^2 = 650$ $\Sigma x^2 = 650 - (6^2+8^2) + (8^2+6^2) = 650$

$\Sigma y = 100$ $\Sigma y = 100 - (14+6) + (12+8) = 100$

$\Sigma y^2 = 460$ $\Sigma y^2 = 100 - (14^2+6^2) + (6^2+8^2) = 436$

$\Sigma xy = 508$ $\Sigma xy = 508 - (6 \cdot 14 + 8 \cdot 6) + (8 \cdot 12 + 6 \cdot 8) = 520$

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

$r = \frac{520 - 125 \cdot 100/25}{\sqrt{650 - \frac{125^2}{25}}\sqrt{460 - \frac{(100)^2}{25}}}$

$r = 0.516$

Q] Coefficient of correlation b/w 2 variables x & y is 0.48. The covariance is 36. The variance

• Spearman's Rank correlation

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

$d = x - y$

Q] Calc. rank correlation coeff

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

x	y	d	d^2
1	3	-2	4
3	1	2	4
7	4	3	9
5	5	0	0
4	6	-2	4
6	9	-3	9
2	7	-5	25
10	8	2	4
9	10	-1	1
8	2	6	36
		Σ	96

$$\rho = 1 - \frac{6 \cdot 96}{10(100-1)}$$

$$\rho = 1 - \frac{576}{990}$$

$$\therefore \rho = 0.418$$

Q] Calc. rank correlation co-eff.

Math	84	56	98	25	75	82	92	65	35
Phy	86	57	91	60	68	62	86	58	49

* Given: n = 10

x	y	R_1	R_2	d	d^2
84	86	10	3	-7	49
56	57	7	8	-1	1
98	91	1	1	0	0
25	60	9	6	3	9
75	68	4	4	0	0
82	62	3	5	-2	4
92	86	2	2	0	0
62	58	6	7	-1	1
65	35	5	10	-5	25
35	49	8	9	-1	1
				Σ	890

$$\rho = 1 - \frac{6 \cdot 890}{10(100-1)}$$

$$= 1 - \frac{5340}{990}$$

$$\therefore \rho = 0.455$$

Q] Calc. rank correlation co-eff

x	10	12	18	18	15	40
y	12	18	25	25	50	25

x	y	R_1	R_2	d	d^2
10	12	1	1	0	0
12	18	2	2	0	0
18	25	4.5	4	0.5	0.25
18	25	4.5	4	0.5	0.25
15	50	3	6	-3	9
40	25	6	4	2	4
					13.5

$m_1 = 2$, $m_2 = 3$

$$\rho = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2)\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[13.5 + \frac{1}{2}(8-2) + \frac{1}{12}(27-3)\right]}{6(36-1)}$$

$$\therefore \rho = 0.543$$

Tutorial 4

Q1.

Day	Mon	Tue	Wed	Thu	Fri	Sat
No. of acc	14	18	12	11	15	14

A: Null Hypothesis (H_0): Accidents are equally distributed over the days.

Alternative Hypothesis (H_1): Accidents are not equally dist. over the days.

Test statistic

$E = \frac{84}{6} = 14$

O	E	O-E	$(O-E)^2$	χ^2
14	14	0	0	0
18	14	4	16	1.143
12	14	-2	4	0.285
11	14	-3	9	0.643
15	14	1	1	0.071
14	14	0	0	0
				2.142

$\Sigma = 2.142$

at $\alpha = 0.05$ i.e. 0.3 for $\gamma = 6$ d.o.f.

$\chi^2_{tab} = 1.635$

$\chi^2_{tab} < \chi^2_{cal}$

$\therefore$ H_0 is rejected. Thus, accidents are not equally dist. over the days.

Q2.