11 November 2020 09:03

Correlation & Regression (Analysis)

1 Correlation Analysis (Gupta & Kapoor)

1) Defn. – Two variables (x & y)

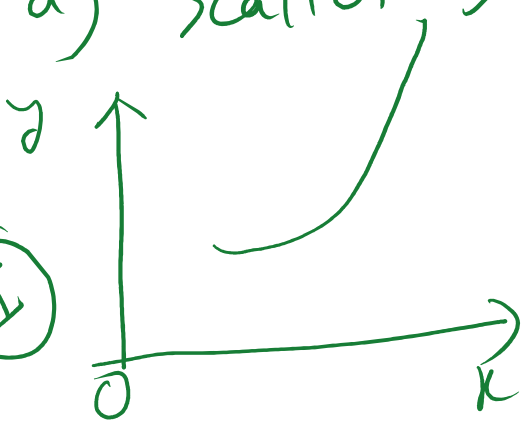
If two variables x & y are related to each other, we say they are correlated.

i) income & expenditure
ii) age & height etc.

a) Scatter Diagram method
b) Karl Pearson's Product Moment Correlation Coefficient
c) Spearman's Rank Corrl. Coeff.

a) Scatter Diagram Method

I



x	y

positive corrl.

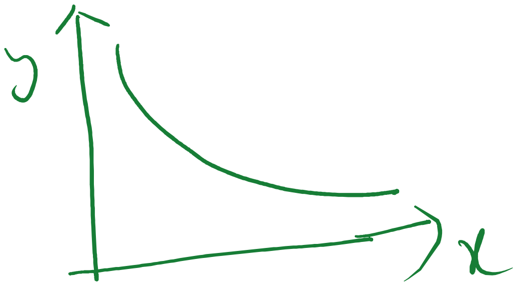
(both x & y increase or

decrease in same direction)

i) Income & exp.

ii) Age & height

II

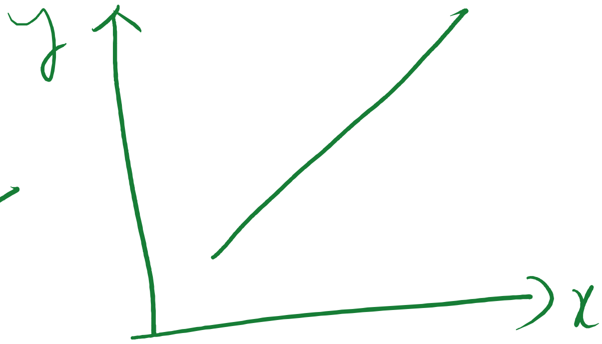


Negative Corel

x & y increase or decrease in opposite or inverse direction

Ex i) price & demand

III

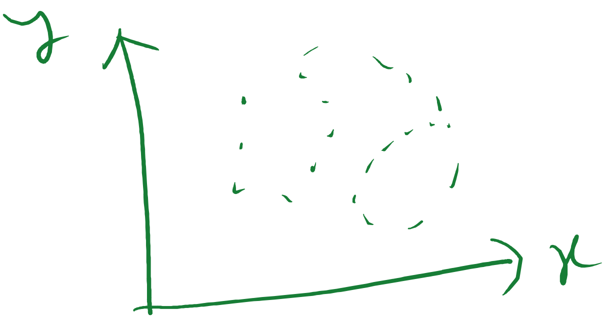


perfect positive correlation.

IV

perfect negative correlation

V



No or zero correlation

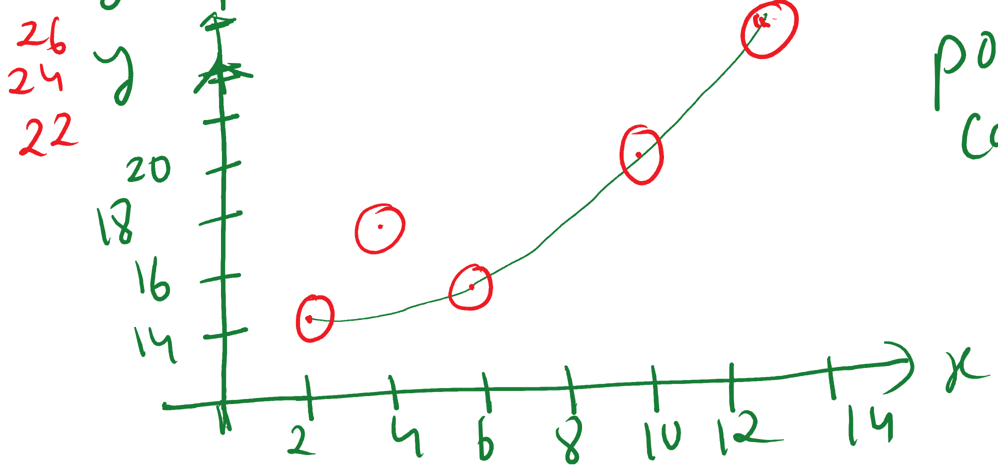
0

0.2

* Nonsense or spurious Correlation

(1) Draw scatter diagram & comment on the corrl.

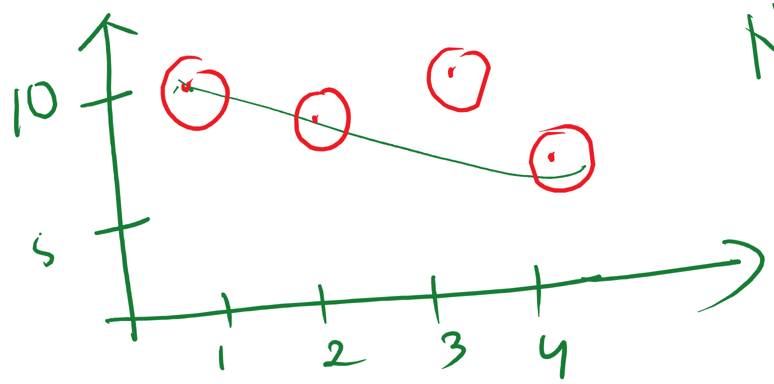
x	2	6	5	10	13
y	14	15	17	20	25



positive corrl.

(2)

x	1	2	3	4
y	10	8	9	5



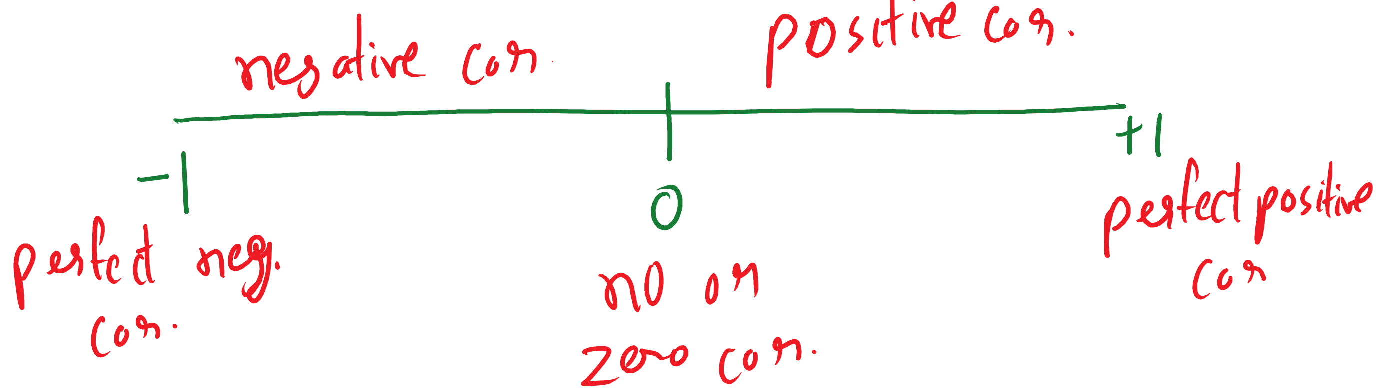
Negative corrl.

18.11.20 (2) Karl Pearson's Product Moment Correl. Coeff. (r)

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

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

Note: The range of r is -1 to $+1$



① Find Pearson's Coeff.

x	y	x^2	y^2	xy
4	12	16	144	48
6	13	36	169	78
7	10	49	100	70
8	15	64	225	120
10	18	100	324	180
$\sum x = 35$	$\sum y =$	$\sum x^2 =$	$\sum y^2 =$	$\sum xy$

$\Sigma x = 35$	$\Sigma y = 68$	$\Sigma x^2 = 265$	$\Sigma y^2 = 962$	Σxy 496

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

$$= \frac{5 \times 496 - 35 \times 68}{\sqrt{5 \times 265 - (35)^2}\sqrt{5 \times 962 - (68)^2}}$$

$$= \frac{2480 - 2380}{\sqrt{100}\sqrt{186}}$$

$$= \frac{100}{10 \times 13.64} = 0.73$$

(positive corel.)

2

x	y	x^2	y^2	xy
2	10			20
3	9			27
7	5			35
$\Sigma x = 12$	$\Sigma y = 24$	$\Sigma x^2 = 62$	$\Sigma y^2 = 206$	$\Sigma xy = 82$

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

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

$$= \frac{3 \times 82 - 12 \times 24}{\sqrt{3 \times 62 - (12)^2}\sqrt{3 \times 206 - (24)^2}}$$

$$= \frac{-42}{\sqrt{42}\sqrt{42}}$$

$$= -1 \quad \text{(perfect negative corr)}$$

3. If $\sum x^2 = 46$, $\sum y^2 = 50$, $\sum xy = 48$
$\sum x = 20$, $\sum y = 30$, $n = 5$
Find r.

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

$$= \frac{5 \times 48 - 20 \times 30}{\sqrt{5 \times 46 - (20)^2}\sqrt{5 \times 50 - (30)^2}}$$

19.11.20

Alternative Formula of r

$$\sum(x - \bar{x})(y - \bar{y})$$

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

Problem

① $\sum (x-\bar{x})(y-\bar{y}) = 120$

$\sum (x-\bar{x})^2 = 110$, $\sum (y-\bar{y})^2 = 72$

$$r = \frac{120}{\sqrt{110}\sqrt{72}} = 1.36$$

② $\sum (x-\bar{x})(y-\bar{y}) = 192$

$\sum (x-\bar{x})^2 = 119$, $\sum (y-\bar{y})^2 = 345$

$$r = \frac{192}{\sqrt{119}\sqrt{345}} = 0.939$$

③ Spearman's Rank Correl. Coeff. (R)

Note. Range of R is -1 to +1

Non Repeated

18	20	17	15	19	14
3	1	4	5	2	6

Repeated Rank

18	20	18	15	19	14	3✓ 4✓
3.5	1	3.5	5	2	6	

	7	8	10	6	7	4	7
3✓ 4✓ 5✓	4	2	1	6	4	7	4

(a) For non repeated observations (ranks)

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

n = no. of obs.

$d = R_1 - R_2$

R_1 = Rank of x

R_2 = " " y

Problem (1) Find R

(R_1) Rank of x	Rank of y (R_2)	$d = R_1 - R_2$	d^2
5	4	1	1
3	3	0	0
4	5	−1	1
1	2	−1	1
2	1	1	1

2	1		1	1
				4

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

$$= 1 - \frac{6 \times 4}{5(5^2-1)}$$

$$= 1 - \frac{6 \times 4}{5 \times 24} = 1 - 0.2 = 0.8$$

(2)

x	y	R_1	R_2	d	d^2
17	13	1	1	0	
8	7	4	4	0	
12	10	3	3	0	
13	11	2	2	0	
					0

$$R = 1 - \frac{6 \sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 0}{4(4^2-1)}$$

$$= 1$$

(perfect positive corr)

20.11.20

x	y	R_1	R_2	d	d^2
212	500	4	5	-1	1
214	515	3	3	0	0

212	[illegible]	[illegible]	[illegible]		
214	515	3	3	0	0
205	511	5	4	1	1
225	530	1	1	0	0
218	525	2	2	0	0
					2

$$R = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 2}{5(5^2-1)}$$

$$= 1 - \frac{12}{5\times 24}$$

$$= 0.9$$

* Repeated obs. (ranks)

$$R = 1 - \frac{6\left[\sum d^2 + \sum \frac{m(m^2-1)}{12}\right]}{n(n^2-1)}$$

m = no. of times a rank is repeated

*

x	y	R_1	R_2	d	d^2
60	18	2.5	1	1.5	2.25
65	17	1	2	−1	1
60	16	2.5	3	−0.5	0.25
57	14	5	5	0	0
58	15	4	4	0	0
					$\sum d^2 = 3.5$

$\Sigma d^2 = 3.5$

For rank 2.5, $m = 2$

$$\frac{m(m^2-1)}{12} = \frac{2(2^2-1)}{12} = \frac{2\times 3}{12} = 0.5$$

$$\Sigma \frac{m(m^2-1)}{12} = 0.5$$

$$R = 1 - \frac{6\left[\Sigma d^2 + \Sigma \frac{m(m^2-1)}{12}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[3.5 + 0.5\right]}{5(5^2-1)}$$

$$= 1 - \frac{6\times 4}{5\times 24} = 0.8$$

x	y	R_1	R_2	d	d^2
52	65	2.5	3	−0.5	0.25
33	59	5	5	0	0
52	72	2.5	1	1.5	2.25
65	65	1	3	−2	4
43	65	4	3	1	1
					$\Sigma d^2 = 7.5$

2, 3, 4

$\frac{2+3+4}{3} = 3$

For rank 2.5, $m=2$, $\frac{m(m^2-1)}{12} = 0.5$

For rank 3, $m=3$

$$\frac{m(m^2-1)}{12} = \frac{3(3^2-1)}{12}$$

$$= \frac{3 \times 8}{12} = 2$$

$$\sum \frac{m(m^2-1)}{12} = 0.5 + 2 = 2.5$$

$$R = 1 - \frac{6\left[\sum d^2 + \sum \frac{m(m^2-1)}{12}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6[7.5 + 2.5]}{5(5^2-1)}$$

$$= 1 - \frac{6 \times 10}{5 \times 24} = 0.5$$