Unit 2 Correlation and Linear regression

Part 1: CORRELATION

Correlation Analysis

Correlation analysis is a statistical tool used to measure the strength of the linear relationship between two variables and compute their association.

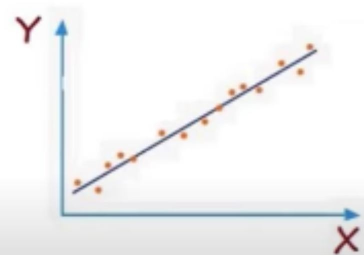
Kinds of Correlations:

can be either a positive, a negative, or no correlation.

Positive Correlation: If the values of the two variables deviate (moves) in the same direction, i.e., if the increase (decrease) in the values of one variable results, on an average, in a corresponding increase (decrease) in the values of the other variables

Examples:

1) Height and weights.
2) Family income and expenditure on luxury items.
3) Price and supply of a commodity



X	2	5	8	10
Y	18	25	34	51

X increases; Y also increases ------- Positive Correlation

Negative or Inverse Correlation: If the values of the two variables deviate (moves) in the opposite direction, i.e., if the increase (decrease) in the values of one variable results, on an average, in a corresponding decrease (increase) in the values of the other variables.

Example:

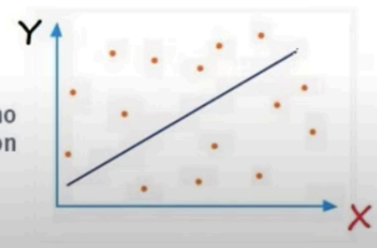
1) Price and demand of a commodity.
2) Volume and pressure of a perfect gas. ($PV = constant$)

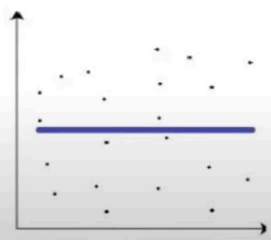
X	8	4	3	1
Y	8	10	15	25

X decreases; Y increases; ----- Negative Correlation

Weak/Zero correlation: It exists when one variable does not affect the other. For example, there is no correlation between the number of years of school a person has attended and the letters in his/her name.

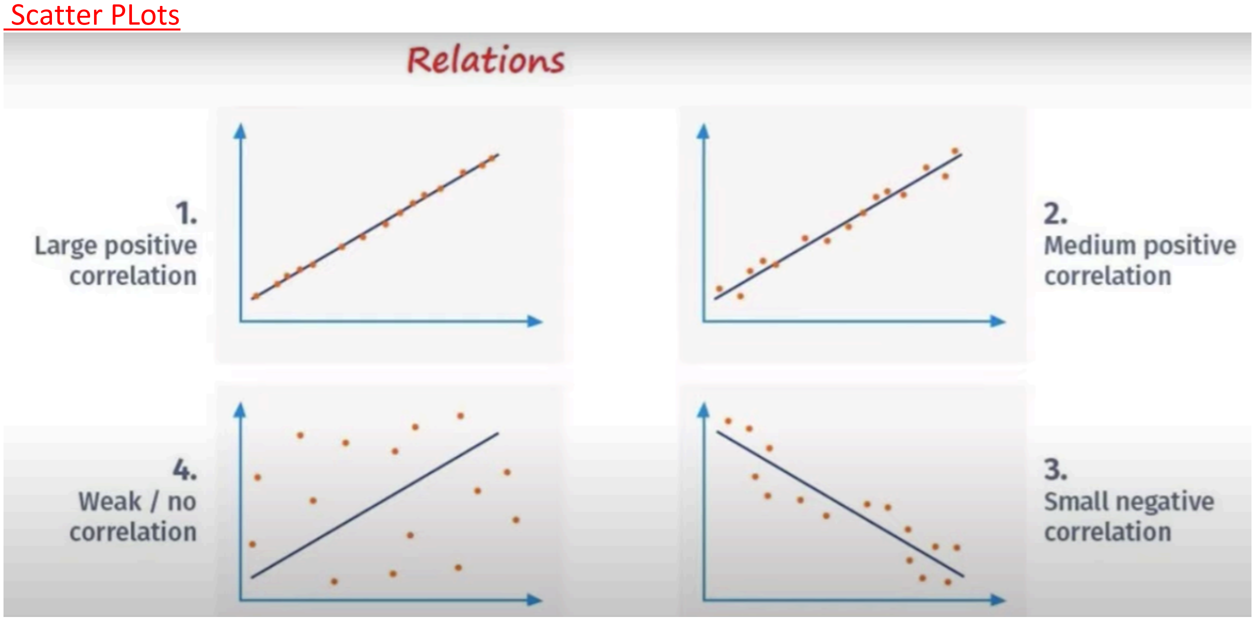
Weak / no correlation





No correlation

Scatter PLots



Remark:

- If the correlation coefficient of two variables is zero, there is no linear relationship between them.
- However, this is only for a linear relationship. It is possible that the variables have a strong curvilinear relationship (non-linear relationship).

Methods of Studying Correlation

1) Pearson product-moment correlation
OR
Karl Pearson's coefficient of correlation (Covariance method).

2) Rank Correlation coefficient method

Karl Pearson's coefficient of correlation (Covariance method).

Karl Pearson's correlation coefficient between X and Y.

It is denoted by **r**(**X**, **Y**) or r_{XY} or simply **r**.

Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y is

$$r_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

or

Correlation coefficient is defined as

$$r = \frac{cov(X,Y)}{\sqrt{Var(X) \times Var(Y)}}$$

OR

$$r_{XY} = \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sqrt{E\{X - E(X)\}^2 E\{Y - E(Y)\}^2}}$$

Rules for finding Covariance

1st method

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

$$Var(X) = E(X^2) - \left(E(X)\right)^2$$

$$Var(Y) = E(Y^2) - \left(E(Y)\right)^2$$

2nd method

$$cov(X,Y) = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{n}$$

$$Var(X) = \frac{\sum(X - \bar{X})^2}{n}$$

$$Var(Y) = \frac{\sum(Y - \bar{Y})^2}{n}$$

OR

$$r_{XY} = \frac{\frac{1}{n}\Sigma x_i\, y_i - \frac{1}{n}\Sigma x_i \cdot \frac{1}{n}\Sigma y_i}{\sqrt{\left\{\frac{1}{n}\Sigma x_i^2 - \left(\frac{1}{n}\Sigma x_i\right)^2\right\}\left\{\frac{1}{n}\Sigma y_i^2 - \left(\frac{1}{n}\Sigma y_i\right)^2\right\}}}$$

$$r_{XY} = \frac{n\,\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{\left\{n\,\Sigma x^2 - (\Sigma x)^2\right\}\left\{n\,\Sigma y^2 - (\Sigma y)^2\right\}}}$$

Example: 1 Calculate the coefficient of correlation between X and Y from the following data:

Summation of product deviation of X and Y from their respective mean is 122.

	Series	
	X	Y
No. of pairs	15	15
Mean	25	18
Sum of Squares of deviations from mean	$\sum(X-\bar{X})^2 = 136$	$\sum(Y-\bar{Y})^2 = 138$

Solution: Given $\sum(X-\bar{X})(Y-\bar{Y}) = 122$

$$Cov(X,Y) = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{n} = \frac{122}{15}$$

$$Var(X) = \frac{\sum(X-\bar{X})^2}{n} = \frac{136}{15}$$

$$Var(Y) = \frac{\sum(Y-\bar{Y})^2}{n} = \frac{138}{15}$$

$$r = \frac{Cov(X,Y)}{\sqrt{Var(X) \times Var(Y)}}$$

$$= \frac{122/15}{\sqrt{\frac{136}{15} \times \frac{138}{15}}}$$

$= 0.8918$

Example 2 *Compute the coefficient of correlation between X and Y, using the following data:*

X	:	1	3	5	7	8	10
Y	:	8	12	15	17	18	20

Solution

x_i	y_i	x_i^2	y_i^2	$x_i\,y_i$
1	8	1	64	8
3	12	9	144	36
5	15	25	225	75
7	17	49	289	119
8	18	64	324	144
10	20	100	400	200
34	90	248	1446	582

Thus $n = 6$

$$\Sigma x_i = 34,\ \Sigma y_i = 90$$

$$\Sigma x_i^2 = 248,\ \Sigma y_i^2 = 1446$$

$$\Sigma x_i\, y_i = 582$$

$$r_{XY} = \frac{n\,\Sigma\, xy - \Sigma\, x \cdot \Sigma\, y}{\sqrt{\left\{n\,\Sigma\, x^2 - (\Sigma\, x)^2\right\}\left\{n\,\Sigma\, y^2 - (\Sigma\, y)^2\right\}}}$$

$$= \frac{6 \times 582 - 34 \times 90}{\sqrt{\left\{6 \times 248 - (34)^2\right\}\left\{6 \times 1446 - (90)^2\right\}}}$$

$$= \frac{432}{\sqrt{332 \times 576}} = 0.9879$$

Properties of Correlation Coefficient

1. $-1 \leq r_{XY} \leq 1$

or $|\text{Cov}(X, Y)| \leq \sigma_X \cdot \sigma_Y$.

Note *When $0 < r_{XY} \leq 1$, the correlation between X and Y is said to be positive or direct.*

When $-1 \leq r_{XY} \leq 0$, the correlation is said to be negative or inverse.
When $-1 \leq r_{XY} \leq -0.5$ or $0.5 \leq r_{XY} \leq 1$, the correlation is assumed to be high, otherwise the correlation is assumed to be poor.

2. Correlation coefficient is independent of change of origin and scale.

   i.e., If $U = \dfrac{X-a}{h}$ and $V = \dfrac{Y-b}{k}$, where $h, k > 0$, then $r_{XY} = r_{UV}$.

Note *If X and Y take considerably large values, computation of r_{XY} will become difficult. In such problems, we may introduce change of origin and scale and compute r using the above property.]*

3. Two independent RV's X and Y are uncorrelated, but two uncorrelated RV's need not be independent.

Two independent variables are uncorrelated.

If X and Y are independent variables, then

$$\text{Cov}(X, Y) = 0$$

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = 0$$

NOTE: If $r = 0$, there is no linear relationship between them. However, this is only for a linear relationship. It is possible that the variables have a strong curvilinear relationship (non-linear relationship).

3 Example: Consider the series

X	-2	-1	0	1	2
Y	4	1	0	1	4
XY	-8	-1	0	1	8

Then E(X)=0 ; E(Y) = 2 ; E(XY) = 0

$$\boxed{r = \frac{cov(X,Y)}{\sqrt{Var(X) \times Var(Y)}}}$$

Cov(XY) = E(XY)-E(X)E(Y)

$r = 0$,

(no linear relation between X & Y)

We easily found a relation which is NON-LINEAR $Y = X^2$

Example 4 *Compute the coefficients of correlation between X and Y using the following data:*

X	:	65	67	66	71	67	70	68	69
Y	:	67	68	68	70	64	67	72	70

Solution We effect change of origin in respect of both X and Y. The new origins are chosen at or near the average of extreme values. Thus we take $\frac{65+71}{2} = 68$ as the new origin for X and $\frac{64+72}{2} = 68$ as the new origin for Y. viz., we put $u_i = (x_i - 68)$ and $v_i = y_i - 68$ and find r_{UV}.

$X = x_i$	$Y = y_i$	$u_i = x_i - 68$	$v_i = y_i - 68$	u_i^2	v_i^2	$u_i v_i$
65	67	−3	−1	9	1	3
67	68	−1	0	1	0	0
66	68	−2	0	4	0	0
71	70	3	2	9	4	6
67	64	−1	−4	1	16	4
70	67	2	−1	4	1	−2
68	72	0	4	0	16	0
69	70	1	2	1	1	2
	Total	−1	2	29	39	13

$$r_{XY} = r_{UV} = \frac{n\sum uv - \sum u \cdot \sum v}{\sqrt{\left\{n\sum u^2 - (\sum u)^2\right\}\left\{n\sum v^2 - (\sum v)^2\right\}}}$$

$$= \frac{8 \times 13 - (-1) \times 2}{\sqrt{(8 \times 29 - 1)(8 \times 39 - 4)}} = \frac{106}{\sqrt{231 \times 308}} = 0.3974$$

REMARK It is advised to calculate the correlation coefficient by \arbitrary origin method rather than by the direct method; since the latter leads to much simpler arithmetical calculations

EXAMPLE 5

A computer while calculating correlation coefficient between two variables X and Y from 25 pairs of observations obtained the following results :

$n = 25$, $\sum X = 125$, $\sum X^2 = 650$, $\sum Y = 100$, $\sum Y^2 = 460$, $\sum XY = 508$

It was, however, later discovered at the time of checking that he had copied down two pairs as

X	Y
6	14
8	6

while the correct values were

X	Y
8	12
6	8

Obtain the correct value of correlation coefficient.

Solution.

Corrected $\sum X = 125 - 6 - 8 + 8 + 6 = 125$

Corrected $\sum Y = 100 - 14 - 6 + 12 + 8 = 100$

Corrected $\sum X^2 = 650 - 6^2 - 8^2 + 8^2 + 6^2 = 650$

Corrected $\sum Y^2 = 460 - 14^2 - 6^2 + 12^2 + 8^2 = 436$

Corrected $\sum XY = 508 - 6\times14 - 8\times6 + 8\times12 + 6\times8 = 520$

$$\bar{X} = \frac{1}{n}\sum X = \frac{1}{25}\times 125 = 5, \quad \bar{Y} = \frac{1}{n}\sum Y = \frac{1}{25}\times 100 = 4$$

$$\text{Cov}(X, Y) = \frac{1}{n}\sum XY - \bar{X}\bar{Y} = \frac{1}{25}\times 520 - 5\times 4 = \frac{4}{5}$$

$$\sigma_X^2 = \frac{1}{n}\sum X^2 - \bar{X}^2 = \frac{1}{25}\times 650 - (5)^2 = 1$$

$$\sigma_Y^2 = \frac{1}{n}\sum Y^2 - \bar{Y}^2 = \frac{1}{25}\times 436 - 16 = \frac{36}{25}$$

$$\therefore \quad \text{Corrected } r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X\,\sigma_Y} = \frac{\frac{4}{5}}{1\times\frac{6}{5}} = \frac{2}{3} = 0{\cdot}67$$

EXAMPLE 6

If the independent random variables X and Y have the variances 36 and 16 respectively, find the correlation coefficient between (X + Y) and (X – Y).

Solution Let $U = X + Y$ and $V = X - Y$

$$E(U) = E(X) + E(Y); E(V) = E(X) - E(Y)$$

$$E(UV) = E(X^2 - Y^2) = E(X^2) - E(Y^2)$$

$$E(U^2) = E\{(X + Y)^2\} = E(X^2) + E(Y^2) + 2E(XY)$$

$$E(V^2) = E(X^2) + E(Y^2) - 2E(XY)$$

$$\begin{aligned} C_{UV} &= E(UV) - E(U) \cdot E(V) \\ &= E(X^2) - E(Y^2) - \{E^2(X) - E^2(Y)\} \\ &= [E(X^2) - E^2(X)] - [E(Y^2) - E^2(Y)] \\ &= \sigma^2_X - \sigma^2_Y = 36 - 16 = 20 \end{aligned}$$

$$\begin{aligned} \sigma^2_U &= E(U^2) - E^2(U) \\ &= \{E(X^2) + E(Y^2) + 2E(XY)\} - \{E^2(X) + E^2(Y) + 2E(X) \cdot E(Y)\} \\ &= [E(X^2) - E^2(X)] + [E(Y^2) - E^2(Y)] + 2[E(XY) - E(X) \cdot E(Y)] \\ &= 36 + 16 + 2 \times 0 \\ &\qquad [\because X \text{ and } Y \text{ are independent and hence uncorrelated}] \\ &= 52 \end{aligned}$$

Similarly, $\sigma^2_V = 52$

Now $r_{UV} = \dfrac{C_{UV}}{\sigma_U \cdot \sigma_V} = \dfrac{20}{52} = \dfrac{5}{13}$

Some extra questions based on definition and properties of correlation coefficient

7

Example: The covariance of two perfectly correlated variable X and Y is 0.96. Find the S.D. of X and Y if it is known that variance of X and Y are in the ratio of 4:9.

Solution: Cov(X,Y) = 0.96 ; $\sigma_X = ?$; $\sigma_Y = ?$ $\frac{\sigma_X^2}{\sigma_Y^2} = \frac{4}{9}$

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

For perfect correlation; $r = 1 \; or -1$

Since Cov(X,Y) is positive, thus $r = 1$

$$\frac{\sigma_X^2}{\sigma_Y^2} = \frac{4}{9} \quad \Rightarrow \sigma_X = \frac{2}{3}\sigma_Y$$

$$1 = \frac{0.96}{\sigma_X \sigma_Y}$$

$$\Rightarrow \sigma_X \sigma_Y = 0.96$$

$$\Rightarrow \frac{2}{3}\sigma_Y^2 = 0.96$$

$$\Rightarrow \sigma_Y = 1.2$$

Therefore, $\sigma_X = 0.8$

8

Example: Calculate the coefficient of correlation between X and Y from the following data:

Summation of product deviation of X and Y from their respective mean is 122.

	Series	
	X	Y
No. of pairs	15	15
Mean	25	18
Sum of Squares of deviations from mean	$\sum(X-\bar{X})^2 = 136$	$\sum(Y-\bar{Y})^2 = 138$

Solution: Given $\sum(X-\bar{X})(Y-\bar{Y}) = 122$

$$Cov(X,Y) = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{n} = \frac{122}{15}$$

$$Var(X) = \frac{\sum(X-\bar{X})^2}{n} = \frac{136}{15}$$

$$Var(Y) = \frac{\sum(Y-\bar{Y})^2}{n} = \frac{138}{15}$$

$$r = \frac{Cov(X,Y)}{\sqrt{Var(X) \times Var(Y)}}$$

$$= \frac{122/15}{\sqrt{\frac{136}{15} \times \frac{138}{15}}}$$

= 0.8918

9

Example: If X and Y are two random variables then show that the correlation coefficient between them is $\frac{\sigma_X^2+\sigma_Y^2-\sigma_{X-Y}^2}{2\sigma_X\sigma_Y}$

Solution:

$$\sigma_{X-Y}^2 = Var(X-Y)$$

$$= Var(X) + Var(Y) - 2Cov(X,Y)$$

$$= \sigma_X^2 + \sigma_Y^2 - 2r\sigma_X\sigma_Y$$

Use $Var(aX+bY) = a^2V(X) + b^2V(Y) + 2abCov(X,Y)$

$$\Rightarrow \quad r = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X\sigma_Y}$$

10

Example: If X and Y are two correlated random variables with the same variance and if r is the correlation coefficient between X and Y, find the correlation coefficient between X and X+Y.

Target

$$r(U,V) = \frac{Cov(U,V)}{\sqrt{Var(U)Var(V)}}$$

Solution: Given that $Var(X) = Var(Y) = k$

and $r = \frac{Cov(X,Y)}{\sqrt{Var(X)\times Var(Y)}} = \frac{Cov(X,Y)}{k}$

Let $U = X$; $V = X + Y$

Cov(U,V) = Cov(X, X+Y)
= E[X(X+Y)] - E(X) E(X+Y)
= E(X^2 + XY) - E(X)[E(X)+E(Y)]
= E(X^2) + E(XY) - $[E(X)]^2$ -E(X) E(Y)
= Var(X) + E(XY) - E(X) E(Y)
= k + Cov(X,Y)
= k + rk
= k (1+r)

Var(U) = Var(X) = k

Var(V) = Var(X+ Y)
= Var(X)+Var(Y)+2Cov(X,Y)
= k+k+2rk
= 2k(1+r)

Thus,

$$r(U,V) = \frac{k(1+r)}{\sqrt{k \times 2k(1+r)}}$$

$$= \sqrt{\frac{1+r}{2}}$$

11

Example: If X, Y and Z are uncorrelated random variables with zero means and Standard deviations 5, 12 and 9 respectively; $U = X + Y$ and $V = Y + Z$. Find the correlation coefficient between U and V.

Target

$$r(U,V) = \frac{Cov(U,V)}{\sqrt{Var(U)\times Var(V)}}$$

Solution: Given that X, Y and Z are uncorrelated random variables

$$\Rightarrow Cov(X,Y) = 0 \; ; \; Cov(X,Z) = 0 \; ; \; Cov(Y,Z) = 0$$

$$E(X) = E(Y) = E(Z) = 0 \, ; \; Var(X) = 25; \; Var(Y) = 144; \; Var(Z) = 81$$

Var(U) = Var(X+Y)
= Var(X) + Var(Y)
= 25 + 144
= 169

Var(V) = Var(Y+Z)
= Var(Y) + Var(Z)
= 144 + 81
= 225

Cov(U,V) = E(UV) - E(U) E(V)
= E($XY + XZ + Y^2 + YZ$)
-[E(X)+E(Y)][E(Y)+E(Z)]
= E(XY) + E(XZ) + E(Y^2)
+ E(YZ) - 0
= 0 + 0 +144 + 0
= 144

Hence, $$r(U,V) = \frac{144}{13 \times 25}$$

12.

Example: If X and Y are two independent random variables with means 5 and 10 and standard deviations 2 and 3 respectively. Find the correlation coefficient between 3X+4Y and 3X - Y.

Target

$$r(U,V) = \frac{Cov(U,V)}{\sqrt{Var(U)\times Var(V)}}$$

Solution: Given that E(X)=5; E(Y) = 10; Var(X) = 4; Var(Y) = 9

Let U = 3X+4Y and V = 3X - Y

Var(U) = Var(3X+4Y)
= 9Var(X)+16Var(Y)
= 9(4) + 16(9)
= 180

Var(V) = Var(3X-Y)
= 9Var(X)+Var(Y)
= 9(4) + 16
= 52

Cov(UV) = E(UV) - E(U) E(V)
= 247

E(UV) = E($9X^2$+9XY-$4Y^2$)
= 9E(X^2) + 9E(XY) -4E(Y^2)
= 9[4+25] + 9(5)(10) - 4[16+100]
= 247

Cov(UV) = E(UV) - E(U) E(V)

E(U) = 3E(X) + 4E(Y)

= 15 + 40 = 55

E(V) = 3E(X) - E(Y)

= 15 - 10 = 5

$\boxed{\therefore\ \text{Cov(UV)} = 247 - (55)(5) = -28}$

Hence,

$$r(U,V) = \frac{-28}{\sqrt{180\times 52}}$$

13

Example: Let X and Y be jointly distributed with the correlation coefficient $\frac{1}{2}$. The variance of X and Y is 4 and 9. Find Var(2X - 4Y + 3).

Solution: Given that $r(X,Y) = \frac{1}{2}$; $Var(X) = 4$; $Var(Y) = 9$

Var(2X - 4Y + 3)

= Var (2X - 4Y) + Var(3)

= Var (2X - 4Y) + 0 since Var(3) = 0

= 4Var(X) + 16 Var(Y) + 2(2)(-4) Cov(X,Y)

= 4(4) + 16 (9) -16 (1/2)(6)

= 112

Use $Var(aX + bY)$
$= a^2V(X) + b^2V(Y)$
$+ 2abCov(X,Y)$

14

Example: X and Y are two random variables with variances σ_X^2 and σ_Y^2 respectively and "r" is the correlation coefficient between them. If U = X+kY and $V = X + \frac{\sigma_X}{\sigma_Y}Y$, find the value of k so that U and V are uncorrelated.

Solution: For U and V are uncorrelated

$\Rightarrow r(U,V) = 0$

$\Rightarrow Cov(U,V) = 0$

$\Rightarrow E(UV) - E(U)\,E(V) = 0$

$$E(UV) = E\left(X^2 + kXY + \frac{\sigma_X}{\sigma_Y}XY + \frac{k\sigma_X}{\sigma_Y}Y^2\right)$$

$$= E(X^2) + \left(k + \frac{\sigma_X}{\sigma_Y}\right)E(XY) + \frac{k\sigma_X}{\sigma_Y}E(Y^2)$$

$$= \left[\sigma_X^2 + (E(X))^2\right] + \left(k + \frac{\sigma_X}{\sigma_Y}\right)E(XY) + \frac{k\sigma_X}{\sigma_Y}\left[\sigma_Y^2 + (E(Y))^2\right]$$

$$E(U) = E(X) + kE(Y)$$

$$E(V) = E(X) + \frac{\sigma_X}{\sigma_Y}E(Y)$$

$$E(U)E(V) = (E(X))^2 + \left(k + \frac{\sigma_X}{\sigma_Y}\right)E(X)E(Y) + \frac{k\sigma_X}{\sigma_Y}(E(Y))^2$$

Now $E(UV) = E(U)\,E(V)$

$$\Rightarrow \left[\sigma_X^2 + \cancel{(E(X))^2}\right] + \left(k + \frac{\sigma_X}{\sigma_Y}\right)E(XY) + \frac{k\sigma_X}{\sigma_Y}\left[\sigma_Y^2 + \cancel{(E(Y))^2}\right] = \cancel{(E(X))^2} + \left(k + \frac{\sigma_X}{\sigma_Y}\right)E(X)E(Y) + \cancel{\frac{k\sigma_X}{\sigma_Y}(E(Y))^2}$$

$$\Rightarrow \sigma_X^2 + \frac{k\sigma_X}{\sigma_Y}\sigma_Y^2 + \left(k + \frac{\sigma_X}{\sigma_Y}\right)Cov(X,Y) = 0$$

$\boxed{\text{Since } Cov(X,Y) = E(XY) - E(X)E(Y)}$

Use $r = \dfrac{Cov(X,Y)}{\sqrt{Var(X)\times Var(Y)}}$

$$\Rightarrow \sigma_X^2 + k\sigma_X\sigma_Y + \left(k + \frac{\sigma_X}{\sigma_Y}\right)r\sigma_X\sigma_Y = 0$$

$$\Rightarrow (1+r)\sigma_X^2 + k(1+r)\sigma_X\sigma_Y = 0$$

$$\Rightarrow (1+r)\sigma_X(\underbrace{\sigma_X + k\sigma_Y}) = 0$$

perf. Neg

Since $r \neq -1, \sigma_X, \sigma_Y > 0$, thus

$$\sigma_X + k\,\sigma_Y = 0$$

$$\Rightarrow k = -\frac{\sigma_X}{\sigma_Y}$$

15

Example: The covariance of two perfectly correlated variable X and Y is 0.96. Find the S.D. of X and Y if it is known that variance of X and Y are in the ratio of 4:9.

Solution: Cov(X,Y) = 0.96 ; $\sigma_X = ?$; $\sigma_Y = ?$ $\frac{\sigma_X^2}{\sigma_Y^2} = \frac{4}{9}$

For perfect correlation; $r = 1$ or -1

Since Cov(X,Y) is positive, thus $r = 1$

$$\boxed{\frac{\sigma_X^2}{\sigma_Y^2} = \frac{4}{9} \quad \Rightarrow \sigma_X = \frac{2}{3}\sigma_Y}$$

$$\boxed{r = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}}$$

$$1 = \frac{0.96}{\sigma_X\sigma_Y}$$

$$\Rightarrow \sigma_X\sigma_Y = 0.96$$

$$\Rightarrow \frac{2}{3}\sigma_Y^2 = 0.96$$

$$\Rightarrow \sigma_Y = 1.2$$

Therefore, $\sigma_X = 0.8$

16.

Example: Calculate the coefficient of correlation between X and Y from the following data:

Summation of product deviation of X and Y from their respective mean is 122.

	Series	
	X	Y
No. of pairs	15	15
Mean	25	18
Sum of Squares of deviations from mean	$\sum(X-\bar{X})^2 = 136$	$\sum(Y-\bar{Y})^2 = 138$

Solution: Given $\sum(X-\bar{X})(Y-\bar{Y}) = 122$

$$Cov(X,Y) = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{n} = \frac{122}{15}$$

$$Var(X) = \frac{\sum(X-\bar{X})^2}{n} = \frac{136}{15}$$

$$Var(Y) = \frac{\sum(Y-\bar{Y})^2}{n} = \frac{138}{15}$$

$$r = \frac{Cov(X,Y)}{\sqrt{Var(X) \times Var(Y)}}$$

$$= \frac{122/15}{\sqrt{\frac{136}{15} \times \frac{138}{15}}}$$

$= 0.8918$

Rank Correlation

A group of n individuals is arranged in order of merit or proficiency in possession of two characteristics A and B. These ranks in the two characteristics will in general be different.

For example, if we consider the relation between intelligence and beauty it is not necessary that a beautiful individual is intelligent also

Spearman's formula for the rank correlation coefficient.

or

Spearman's rank correlation coefficient

To compute the Spearman's rank correlation coefficient (ρ) between the two variables by

(i)

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

When there is a no tie during the rank

when there is a tie during the rank

(ii)

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

To $\sum d^2$ we add $\frac{m(m^2 - 1)}{12}$ for each value repeated, where m is the number of times a value occurs.

Tied Ranks. If some of the individuals recei.ve the same rank in a ranking or merit, they are said to be tied. Let us suppose that **m**

of the individuals, say, **(k** + 1)th, **(k** + 2)th, ... , **(k** + m)th are tied. Then each of these **m** individuals is assigned a common rank, which is the arithmetic mean of the ranks **k+1,** k +2, **..... k+m.**

i.e common rank value= [(k+1) + (k +2)+ **..... +(k+m)]/ m**

for example

two terms 4^{th} and 5^{th} have same value then

common rank value will be (4+5)/2 = 4.5

Lets say three entries 7^{th} , 8^{th} and 9^{th} have same value then Common rank assigned to them is (7+8+9)/3 =8

Example: The following data gives the HDL and LDL cholesterol levels of 7 adults in a locality

HDL(X)	36	39	23	31	33	51	45
HDL(Y)	80	72	101	90	98	70	50

Compute the rank correlation coefficient (ρ)

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

$$= 1 - 6 \times \frac{108}{7(7^2-1)}$$

$$= -0.928$$

HDL (X)	HDL (Y)	Rank of X (R_X)	Rank of Y (R_Y)	$d = R_X - R_Y$	d^2
36	80	4	4	0	0
39	72	5	3	2	4
23	101	1	7	-6	36
31	90	2	5	-3	9
33	98	3	6	-3	9
51	70	7	2	5	25
45	50	6	1	5	25
Total					108

Example 10·18. *Obtain the rank correlation coefficient for the following data:*

X	:	*68*	*64*	*75*	*50*	*64*	*80*	*75*	*40*	*55*	*64*
Y	:	*62*	*58*	*68*	*45*	*81*	*60*	*68*	*48*	*50*	*70*

X	Y	Rank X (x)	Rank Y (y)	$d = x - y$	d^2
68	62	4	5	−1	1
64	58	6	7	−1	1
75	68	2·5	3·5	−1	1
50	45	9	10	−1	1
64	81	6	1	5	25
80	60	1	6	−5	25
75	68	2·5	3·5	−1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
				$\Sigma d = 0$	$\Sigma d^2 = 72$

$$\rho = 1 - \frac{6\left[\Sigma d^2 + \frac{5}{2} + \frac{1}{2}\right]}{n(n^2 - 1)} =$$

$$1 - \frac{6(72 + 3)}{10 \times 99} = 0{\cdot}545$$

Example: Compute the rank correlation coefficient for the following grades of 12 students selected at random

Mathematics grade	85	83	87	84	88	82	90	86	88	87	89	90
Economic Grade	88	86	88	86	90	86	88	85	92	86	88	91

Solution:

(X)	(Y)	Rank of X (R_X)	Rank of Y (R_Y)	$d = R_X - R_Y$	d^2
85	88	4	7.5	-3.5	12.25
83	86	2	3.5	-1.5	2.25
87	88	6.5	7.5	-1.0	1
84	86	3	3.5	-0.5	0.25
88	90	8.5	10	-1.5	2.25
82	86	1	3.5	-2.5	6.25
90	88	11.5	7.5	4.0	16
86	85	5	1	4.0	16
88	92	8.5	12	-3.5	12.25
87	86	6.5	3.5	3.0	9
89	88	10	7.5	2.5	6.25
90	91	11.5	11	0.5	0.25
Total					84

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

$$= 1 - \frac{6}{12(12^2-1)}\left[\begin{array}{c} 84 + \frac{2(2^2-1)}{12} + \frac{2(2^2-1)}{12} \\ + \frac{2(2^2-1)}{12} + \frac{4(4^2-1)}{12} + \frac{4(4^2-1)}{12} \end{array}\right]$$

$$= -0.666$$

Example 12 *Ten competitors in a beauty contest were ranked by three judges as follows:*

Judges	*Competitors*									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
A:	*6*	*5*	*3*	*10*	*2*	*4*	*9*	*7*	*8*	*1*
B:	*5*	*8*	*4*	*7*	*10*	*2*	*1*	*6*	*9*	*3*
C:	*4*	*9*	*8*	*1*	*2*	*3*	*10*	*5*	*7*	*6*

Discuss which pair of judges have the nearest approach to common taste of beauty.

Solution

Rank by A (U)	*Rank by B* (V)	*Rank by C* (W)	$d_1 = U - V$	$d_2 = V - W$	$d_3 = U - W$	d_1^2	d_2^2	d_3^2
6	5	4	1	1	2	1	1	4
5	8	9	– 3	– 1	– 4	9	1	16
3	4	8	– 1	– 4	– 5	1	16	25
10	7	1	3	6	9	9	36	81
2	10	2	– 8	8	0	64	64	0
4	2	3	2	– 1	1	4	1	1
9	1	10	8	– 9	– 1	64	81	1
7	6	5	1	1	2	1	1	4
8	9	7	– 1	2	1	1	4	1
1	3	6	– 2	– 3	– 5	4	9	25
					Total:	157	214	158

$$r_{UV} = 1 - \frac{6\,\Sigma d_1^2}{n(n^2-1)} = 1 - \frac{6 \times 157}{10 \times 99} = 0.0485$$

$$r_{VW} = 1 - \frac{6\,\Sigma d_2^2}{n(n^2-1)} = 1 - \frac{6 \times 214}{10 \times 99} = -\,0.2970$$

$$r_{UW} = 1 - \frac{6\,\Sigma d_3^2}{n(n^2-1)} = 1 - \frac{6 \times 158}{10 \times 99} = 0.0424$$

Since r_{UV} is maximum, the judges A and B may be considered to have common taste of beauty to some extent compared to other pairs of judges.