Correlation

→ Corelation is a statistical measure for finding out degree of association b/w 2 or more variables

(OR)

→ The relationship b/w 2 variables such that a change in one variable results in the positive or negetive change in the other is called corelation

(OR)

→ A greater change in one variable resulting in a corresponding greater or smaller change in other variable is called corelation

Types of corelation

→ ① Positive & negetive

→ ② Simple and multiple

→ ③ partial & Totale

□ Positive & negetive

When 2 variables tend to move together in the same direction then the corelation is called positive corelation

→ When 2 variables tend to move together in opposite dirn then corelation is called -ve corelation

eg:- Demand and supply
height & weight
age & IQ
income & expenditure.

☐ Simple & Multiple

→ When we study only two variables say x & y then the relationship is described as simple correlation

→ If we study more than two variables simultaneously then the corelation is called multiple corelation.

Partial & Total

→ If the study of the variables excludes some other variables then it is called partial correlation

→ If all the facts are taken into account then the correlation is called total corelation.

Co-efficient of Co-rrelation

→ The extent or degree of relationship b/w two variables measured in terms of another parameter is called coefficient of correlation
(or)

→ The numerical measure of linear relationship b/w 2 variables is called coefficient of correlation.

→ It is denoted by 'r' & is defined as *

$$\boxed{r = \frac{\Sigma dx\,dy}{\sqrt{\Sigma dx^2 \, \Sigma dy^2}}}$$

where $dx = x - \bar{x}$ | $dy = y - \bar{y}$ | $\bar{x} = \frac{\Sigma x}{n}$

$dx^2 = (x - \bar{x})^2$ | $dy^2 = (y - \bar{y})^2$ | $\bar{y} = \frac{\Sigma y}{n}$

*

Q → Find the co-efficient of correlation b/w x and y for the given data

x	1	2	3	4	5	6	7	8	9
y	10	11	12	14	13	15	16	17	18

Sol. $\bar{x} = \frac{\sum x}{n} = 5$

$\bar{y} = \frac{\sum y}{n} = 14$

Computation table

x	$dx = x-\bar{x}$	$dx^2 = (x-\bar{x})^2$	y	$dy = y-\bar{y}$	$dy^2 = (y-\bar{y})^2$	$dx\,dy = (x-\bar{x})(y-\bar{y})$
1	−4	16	10	−4	16	16
2	−3	9	11	−3	9	9
3	−2	4	12	−2	4	4
4	−1	1	14	0	0	0
5	0	0	13	−1	1	0
6	1	1	15	1	1	1
7	2	4	16	2	4	4
8	3	9	17	3	9	9
9	4	16	18	4	16	16
45	0	60	126	0	60	59

$$r = \frac{59}{\sqrt{(60)(60)}} = \frac{59}{60} = 0.9833$$

Note :- The coefficient of correlation lies between -1 and $1 \Rightarrow \boxed{-1 \le r \le 1}$

→ if $r = 1$ we say that variables are +vely perfectly correlated

→ if $r = -1$ we say that variables are -vely perfectly correlated

→ If $r = 0$ the variables are not correlated or there is no correlation b/w x and y.

→ If r lies 0 & 1 then variables are partially +vely correlated

→ If $-1 < r < 0$ then the variables are partially negetively correlated.

Q → find the coefficient of correlation for the following data

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

Sol $\bar{x} = \frac{\Sigma x}{n} = 68$

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

Computation table

x	$dx = x - \bar{x}$	$dx^2 = (x-\bar{x})^2$	y	$dy = y - \bar{y}$	$dy^2 = (y-\bar{y})^2$	$dxdy = (x-\bar{x})(y-\bar{y})$
65	-3	9	67	-2	4	6
66	-2	4	68	-1	1	2
67	-1	1	65	-4	16	4
67	-1	1	68	-1	1	1
68	0	0	72	3	9	0
69	1	1	72	3	9	3
70	2	4	69	0	0	0
72	4	16	71	2	4	8
Σ 544	$\Sigma = 0$	$\Sigma = 36$	Σ 552	$\Sigma = 0$	$\Sigma = 44$	$\Sigma = 24$

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \, \Sigma dy^2}} \Rightarrow \frac{24}{\sqrt{(36)(44)}}$$

$$r = 0.6030$$

r lies between 0 & 1 it partially +ve correlated

Q Find the coefficient of correlation for the given data

x	64	65	66	67	68	69	70
y	66	67	68	69	70	71	72

Sol $\bar{x} = \frac{\Sigma x}{n} = \frac{469}{7} = 67$

$\bar{y} = \frac{\Sigma y}{n} = \frac{483}{7} = 69$

x	$dx = x-\bar{x}$	$dx^2 = (x-\bar{x})^2$	y	$dy = y-\bar{y}$	$dy^2 = (y-\bar{y})^2$	$dxdy = (x-\bar{x})(y-\bar{y})$
64	-3	9	66	-3	9	9
65	-2	4	67	-2	4	4
66	-1	1	68	-1	1	1
67	0	0	69	0	0	0
68	1	1	70	1	1	1
69	2	4	71	2	4	4
70	3	9	72	3	9	9
$\Sigma x = 469$	$\Sigma dx = 0$	$\Sigma dx^2 = 28$	$\Sigma y = 483$	$\Sigma dy = 0$	$\Sigma dy^2 = 28$	$\Sigma dxdy = 28$

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \Sigma dy^2}} = \frac{28}{\sqrt{28 \times 28}} = 1$$

$r = 1$ variables are perfectly truly corellated

Regression

→ Regression is the measure of avg relationship between 2 or more variables in terms of original units of data

→ In regression analysis the nature of actual relationship if it exist b/w 2 or more variables is studied by determining the mathematic eqn b/w the variables.

→ It is mainly used to predict or estimate one variable in terms of the other variables

→ The Functional relationship of a dependent variable with one or more independent variables is called regression eqn. It is also called prediction eqn or estimating eqn.

→ If the curve is a st. line then it is called line of regression.

→ If there are 2 variables under consideration then the regression is called simple regression

→ If there are more than 2 variables under consideration then the regression is called multiple regression.

Scanned with CamScanner

Note: The regression line y on x passing through the point $(\bar{x}, \bar{y})$ is given by ~~$y-\bar{y}=b(x-\bar{x})$~~

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

where $b_{yx} = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

Note: The regression line x on y passing through the point $(\bar{x}, \bar{y})$ is given by

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

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

→ regression lines are also known as regression eq^n.

Regression Coefficient

$$r^2 = b_{xy} \cdot b_{yx}$$

Note: In regression we can estimate the variable with the value of another variable which is known

→ The statistical method which helps us to estimate the unknown value of one variable from the know value of related variable is called regression.

Properties of Coefficient of correlation

1) The value of coefficient of correlation lies b/w -1 & +1

2) If $r = 0$ the variables are said to be independ

3) If $r = \pm 1$ then there is prerfect coff correlation coefficient.

Q → Find the coefficient of corellation and eq^n of regression line for the following data

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

$$\bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{108}{9} = 12$$

x	$dx = (x-\bar{x})$	$dx^2 = (x-\bar{x})^2$	y	$dy = y-\bar{y}$	$dy^2 = (y-\bar{y})^2$	$dxdy = (x-\bar{x})(y-\bar{y})$
1	-4	16	9	-3	9	12
2	-3	9	8	-4	16	12
3	-2	4	10	-2	4	4
4	-1	1	12	0	0	0
5	0	0	11	-1	1	0
6	1	1	13	1	1	1
7	2	4	14	2	4	4
8	3	9	16	4	16	12
9	4	16	15	3	9	12
$\Sigma x = 45$	$\Sigma dx = 0$	$\Sigma dx^2 = 60$	$\Sigma y = 108$	$\Sigma dy = 0$	$\Sigma dy^2 = 60$	57

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \Sigma dy^2}} = \frac{57}{\sqrt{60 \times 60}} = 0.95$$

partially truly correlated

regression eq^n for y on x

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

where $b_{yx} = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

$$b_{yx} = \frac{57}{60}$$

$$b_{yx} = 0.95$$

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

$$y-12 = 0.95(x-5)$$

$$y = 12 + 0.95x - 4.75 \Rightarrow y = 0.95x + 7.25$$

regression eq^n for x on y

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

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

$$= 0.95$$

$$x-5 = 0.95(y-12)$$

$$x = 5 + 0.95y - 11.4$$

$$x = 0.95y - 6.4$$

Scanned with CamScanner

Q → Find COC and regression eqⁿ for the following data

x	1	2	3	4	5	6	7	8	9	10
y	10	12	16	28	25	36	41	49	40	50

Sol

$$\bar{x} = \frac{\Sigma x}{n} = \frac{55}{10} = 5.5$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{307}{10} = 30.7$$

x	$dx = (x-\bar{x})$	$dx^2 = (x-\bar{x})^2$	y	$dy = (y-\bar{y})$	$dy^2 = (y-\bar{y})^2$	$dxdy = (x-\bar{x})(y-\bar{y})$
1	−4.5	20.25	10	−20.7	428.49	93.15
2	−3.5	12.25	12	−18.7	349.69	65.45
3	−2.5	6.25	16	−14.7	216.09	36.75
4	−1.5	2.25	28	−2.7	7.29	4.05
5	−0.5	0.25	25	−5.7	32.49	2.85
6	0.5	0.25	36	5.3	28.09	2.65
7	1.5	2.25	41	10.3	106.09	15.45
8	2.5	6.25	49	18.3	334.89	45.75
9	3.5	12.25	40	9.3	86.49	32.55
10	4.5	20.25	50	19.3	372.49	86.85
$\Sigma x = 55$	$\Sigma dx = 0$	$\Sigma dx^2 = 82.5$	$\Sigma y = 307$	$\Sigma dy = 0$	$\Sigma dy^2 = 1962.1$	$\Sigma dxdy = 385.5$

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \, \Sigma dy^2}} = \frac{385.5}{\sqrt{(82.5)(1962.1)}}$$

$$r = 0.95815785$$

y on x

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

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

$$y - 30.7 = 4.6727(x-5.5)$$

$$y = 30.7 + 4.6727x - 285.7$$

$$y = 4.6727x + 5$$

x on y

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

$$x - 5.5 = b_{xy}(y - 30.7)$$

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

$$b_{xy} = \frac{385.5}{1962.1}$$

$$b_{xy} = 0.1964$$

$$x - 5.5 = 0.1964(y - 30.7$$

$$x = 5.5 + 0.1964y - 6.02948$$

$$x = 0.1964y - 0.52948$$

14/10/2022

$Q_1 \rightarrow$ Find the coefficient of co-relation for the following data

x	68	64	75	50	64	80	75	40	55	6
y	62	58	68	45	81	60	68	48	50	7

$r = 0.62$

Q_2 Find the corelation ~~coefficient~~ coefficient and also the regression lines for the following data $\rightarrow r = 0.5197$

x	1	2	3	4	5	6	7
y	2	4	7	6	5	6	5

$x = 0.6875y + 0.562$

$y = 0.3928x + 3.428$

Q_3 Coefficient correlation & regration lines

x	1	2	3	4	5
y	2	5	3	8	7

$r = 0.8$

$y = 0.393x + 3.428$

$x = 0.688y + 0.56$

Q_4: The regretion eq^ns calculated from a given set of observations are
$x = -0.4y + 6.4$ and $y = -0.6x + 4.6$
calculate $\bar{x}$ and $\bar{y}$.

Q Two random variables have the regression lines
$3x + 2y = 26$ and $6\cancel{y}^{x} + y = 31$ find the mean value and regression co-efficient

ans $\bar{x} = 4$, $\bar{y} = 7$ $r^2 = 0.2499$

→ regression co-efficient ⇒ $r^2 = b_{xy} \times b_{yx}$
$x = a + by$ ~~represents~~ represents regression eq^n x on y
$y = a + bx$ represents regression line y on x

$3x + 2y = 26$	$6x + y = 31$
$3x = 26 - 2y$	$y = 31 - 6x$
$x = \frac{26}{3} - \frac{2}{3}y$	$y = a + bx$
$x = a + by$	$b_{yx} = 6$
$b_{xy} = \frac{2}{3}$	

$2y = 26 - 3x$	$6x = 31 - \cancel{6}y$
$y = \frac{26}{2} - \frac{3}{2}x$	$x = \frac{31}{6} - \frac{1}{6}y$
$b_{xy} = \frac{3}{2}$	$b_{yx} = 1/6$

$$r^2 = \frac{3}{2} \times \frac{1}{6} = \frac{\cancel{3}1}{\cancel{12}4} = \cancel{0.25}\ 0.25$$

Q. The regression eq^n of variable x & y are $x = 0.7y + 5.2$

$y = 0.3x + 2.8$

Sol. $\bar{x} = \frac{\Sigma x}{n} = \frac{635}{10} = 63.5$

$\bar{y} = \frac{\Sigma y}{n} = \frac{610}{10} = 61$

x	$dx = (x-\bar{x})$	$dx^2 = (x-\bar{x})^2$	y	$dy = (y-\bar{y})$	$dy^2 = (y-\bar{y})^2$	$dxdy = (x-\bar{x})(y-\bar{y})$
68	4.5	20.25	62	1	1	4.5
64	0.5	0.25	58	-3	9	-1.5
75	11.5	132.25	68	7	49	80.5
50	-13.5	182.25	45	-16	256	216
64	0.5	0.25	81	20	400	10
80	16.5	272.25	60	-1	1	-16.5
75	11.5	132.25	68	7	49	80.5
40	-23.5	552.25	48	-13	169	305.5
55	-8.5	72.25	50	-11	121	93.5
64	0.5	0.25	70	9	81	4.5

$\Sigma x = 63.5$ $\Sigma dx = 0$ $\Sigma dx^2 = 1364.5$ $\Sigma y = 61$ $\Sigma dy = 0$ $\Sigma dy^2 = 1136$ $\Sigma dxdy = 777$

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \Sigma dy^2}}$$

$$= \frac{777}{\sqrt{(1364.5)^2 (1136)^2}}$$

$$= 0.62$$

Scanned with CamScanner

Ans 2

$\bar{x} = \frac{\Sigma x}{n} = \frac{28}{7} = 4$

$\bar{y} = \frac{\Sigma y}{n} = \frac{35}{7} = 5$

x	$dx = (x-\bar{x})$	$dx^2 = (x-\bar{x})^2$	y	$dy = (y-\bar{y})$	$dy^2 = (y-\bar{y})^2$	$dxdy = (x-\bar{x})(y-\bar{y})$
1	-3	9	2	-3	9	9
2	-2	4	4	-1	1	2
3	-1	1	7	2	4	-2
4	0	0	6	1	1	0
5	1	1	5	0	0	0
6	2	4	6	1	1	2
7	3	9	5	0	0	0
$\Sigma x = 28$	$\Sigma dx = 0$	$\Sigma dx^2 = 28$	$\Sigma y = 35$	$\Sigma dy = 0$	$\Sigma dy^2 = 16$	$\Sigma dxdy = 11$

$$r = \frac{\Sigma dxdy}{\sqrt{\Sigma dx^2 \Sigma dy^2}} = \frac{11}{\sqrt{28 \times 16}} = 0.5917$$

x on y

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

$x - 4 = 0.6875y - 3.4376$ $\qquad = \frac{11}{16} = 0.6875$

$x = 0.6875y + 0.5624$

y on x

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

$y - 5 = 0.3928(x - 4)$ $\qquad = \frac{11}{28} = 0.3928$

$y = 0.3928x - 1.5712 + 5$

$y = 0.3298x + 3.4288$

Ans 3

$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5} = 3$

$\bar{y} = \frac{\Sigma y}{n} = \frac{25}{5} = 5$

x	$dx = (x-\bar{x})$	$dx^2 = (x-\bar{x})^2$	y	$dy = (y-\bar{y})$	$dy^2 = (y-\bar{y})^2$	$dxdy = \Sigma(x-\bar{x})(y-\bar{y})$
1	-2	4	2	-3	9	6
2	-1	1	5	0	0	0
3	0	0	3	-2	4	0
4	1	1	8	3	9	3
5	2	4	7	2	4	4

$\Sigma x = 15$ $\Sigma dx = 0$ $\Sigma dx^2 = 10$ $\Sigma y = 25$ $\Sigma dy = 0$ $\Sigma dy^2 = 26$ $\Sigma dxdy = 13$

$$r = \frac{\Sigma dx\,dy}{\sqrt{dx^2 dy^2}} = \frac{13}{\sqrt{10 \times 26}} = 0.8062$$

<u>x on y</u>

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

$x - 3 = 0.5(y-5)$

$x = 0.5y - 2.5 + 3$

$x = 0.5y + 0.5$

$$b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} = \frac{13}{26} = \frac{1}{2} = 0.5$$

Scanned with CamScanner

Sol 4 given $eq^n s$ $x = -0.4y + 6.4$ $y = -0.6x + 4.6$

$x + 0.4y = 6.4$

$y + 0.6x = 4.6$

$\bar{x} = 6$

$\bar{y} = 1$

Sol 5 Given

$x = 0.7y + 5.2$

$y = 0.3x + 2.8$

Solving ① & ②

$\bar{x} = 9.0632$ $\bar{y} = 5.5189$

$r^2 = b_{xy} \times b_{yx}$

$b_{xy} = 0.7$

$b_{yx} = 0.3$

$r^2 = b_{xy} \times b_{yx} = (0.7)(0.3) = 0.21$

Angle between two lines of Regression.

→ Consider the regression lines x on y i.e

$$x - \bar{x} = b_{xy}(y - \bar{y}) \quad \text{— (1)}$$

y on x → $$y - \bar{y} = b_{yx}(x - \bar{x}) \quad \text{— (2)}$$

Note

$\bar{x}$ is mean values of x

$\bar{y}$ is mean values of y

r = coefficient of correlation

σ_x = Standard deviation of values of x from $\bar{x}$

σ_y = standard deviation of values of y from $\bar{y}$

$$(1) \rightarrow y - \bar{y} = \frac{1}{b_{xy}}(x - \bar{x})$$

$$\Rightarrow \cancel{y - \bar{y} = \frac{1}{r\sigma_x}} \qquad b_{xy} = \frac{r\sigma_x}{\sigma_y}$$

$$b_{yx} = \frac{r\sigma_y}{\sigma_x}$$

$$(1) \rightarrow y - \bar{y} = \frac{1}{\frac{r\sigma_x}{\sigma_y}}(x - \bar{x})$$

$$y - \bar{y} = \frac{\sigma_y}{r\sigma_x}(x - \bar{x}) \quad \text{— (3)}$$

$$(2) \quad y - \bar{y} = \frac{r\sigma_y}{\sigma_x}(x - \bar{x}) \quad \text{— (4)}$$

$$m_1 = \frac{\sigma_y}{r\sigma_x} \qquad m_2 = \frac{r\sigma_y}{\sigma_x}$$

let θ be the angle b/w 2 regression lines ③ & ④

we have $\tan\theta = \pm \dfrac{m_1 - m_2}{1 + m_1 m_2}$

$$\tan\theta = \pm \frac{\dfrac{\sigma_y}{r\sigma_x} - \dfrac{r\sigma_y}{\sigma_x}}{1 + \dfrac{\sigma_y}{\cancel{r}\sigma_x} \cdot \dfrac{\cancel{r}\sigma_y}{\sigma_x}}$$

$$\theta = \pm \frac{\dfrac{\sigma_y}{\sigma_x}\left[\dfrac{1}{r} - r\right]}{1 + \dfrac{\sigma^2 y}{\sigma^2 x}}$$

$$\tan\theta = \pm \frac{\left(\dfrac{1-r^2}{r}\right)\dfrac{\sigma_y}{\sigma_x}}{\dfrac{\sigma^2 x + \sigma^2 y}{\sigma^2 x}}$$

$$\boxed{\tan\theta = \pm \frac{\left(\dfrac{1-r^2}{r}\right)\sigma_x \sigma_y}{\sigma^2 x + \sigma^2 y}}$$

Note :- Positive sign gives acute value of lines, −ve sign gives obtuse value for lines.

Spearman's Rank correlation coefficient

Spearman's Rank correlation coefficient is denoted by ρ and defined as $\boxed{\rho = 1 - \frac{6\sum D^2}{n(n^2-1)}}$

where n is no. of paired observations

D^2 is sum of squares of difference of 2 ranks

and ρ is rank correlation coefficient.

Q → A random sample of 5 college students is selected and there grades in maths & statistics are found to be.

Mathematics	85	60	73	40	90
Statistics	93	75	65	50	80

Calculate Spearman's Rank correlation coefficient (S.R.C.C)

Sol S.R.CC is given by $1 - \frac{6\sum D^2}{n(n^2-1)}$

Computation table

Mathematics x	Rank of x	Statistics y	Rank of y	$D = x - y$	$D^2 = (x-y)^2$
85	2	93	1	1	1
60	4	75	3	1	1
73	3	65	4	-1	1
40	5	50	5	0	0
90	1	80	2	-1	1

$\Sigma D^2 = 4$

$$\rho = 1 - \frac{6(4)}{5(24)} \Rightarrow 1 - \frac{24}{5(24)}$$

$$\boxed{\rho = 0.8}$$

②

x	40	46	54	60	70	80	82	85	86	90	95
y	45	46	50	43	40	75	55	72	65	42	70

S.R.CC is given by $1 - \frac{6\Sigma D^2}{n(n^2-1)}$

P.T.O

Computation table.

x	Rank of x	y	Rank of y	$D = x - y$	$D^2 = (x-y)^2$
40	11	45	8	3	9
46	10	46	7	3	9
54	9	50	6	3	9
60	8	43	9	-1	1
70	7	40	11	-4	16
80	6	75	1	5	25
82	5	55	5	0	0
85	4	72	2	2	4
86	3	65	4	-1	1
90	2	42	10	-8	64
95	1	70	3	-2	4

$\Sigma D^2 = 142$

$$\rho = 1 - \frac{6(142)}{11(11^2 - 1)}$$

$$\boxed{\rho = 0.3545}$$

(3)

x	10	15	12	17	13	16	24	14	22
y	30	42	45	46	33	34	40	35	39

Scanned with CamScanner

x	Rank of x	y	Rank of y	$D = x - y$	$D^2 = (x-y)^2$
10	9	30	9	0	0
15	5	42	3	2	4
12	8	45	2	6	36
17	3	46	1	2	4
13	~~6~~ 7	33	8	-1	1
16	4	34	7	-3	9
24	1	40	4	-3	9
14	6	35	6	0	0
22	2	39	5	-3	9

$\Sigma D^2 = 72$

$$\rho = 1 - \frac{6 \Sigma D^2}{n(n^2-1)}$$

$$\rho = 1 - \frac{6(72)}{9(80)}$$

$$\boxed{\rho = 0.4}$$

10/10/2022

Spearman's rank correlation coefficient for eq

ranks

→ If any two or more persons are bracketed eq in any classification of if there is more th one item with the same value in the series then the spearman's formula for calculation the rank correlation coefficient breaks down

$$\rho = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - n) \ldots\right)}{n(n^2-1)}$$

where n = number of paired observations

D^2 = Sum of squares of diffrences of 2 ranks

ρ = Rank correlation coefficient

m = ~~#~~ No. of item whose rank are common

1) From the following data calculate the rank correllation coefficient

x	48	33	40	9	16	16	65	24	16	57
y	13	13	24	6	15	4	20	9	6	19

x	Rank of x	y	Rank of y	$D = x - y$	$D^2 = (x-y)^2$
48	3	13	5.5	-2.5	6.25
33	5	13	5.5	-0.5	0.25
40	4	24	1	3	9
9	10	6	8.5	1.5	2.25
16	8	15	4	4	16
16	8	4	10	-2	4
65	1	20	2	-1	1
24	6	9	7	-1	1
16	8	6	8.5	-0.5	0.25
57	2	19	3	-1	1

$\Sigma D^2 = 41$

→ 16 is repeated 3 times in x hence m=3

→ Since 13 & 6 are repeated twice in y

hence m=2

$$\rho = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m)\right)}{n(n^2-1)}$$

$$\rho = 1 - \frac{6\left(41 + \frac{1}{12}(3^3-3) + \frac{1}{12}(2^3-2) + \frac{1}{12}(2^3-2)\right)}{10(10^2-1)}$$

$\rho = 0.733$

2) Compute the rank correlation coefficient of 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	Rank of x	y	Rank of y	$D = x - y$	$D^2 = (x-y)^2$
68	4	62	5	-1	1
64	6	58	7	-1	1
75	2.5	68	3.5	-1	1
50	9	45	10	-1	1
64	6	81	1	5	25
80	1	60	6	-5	25
75	2.5	68	3.5	-1	1
40	10	48	9	1	1
55	~~7~~ 8	50	8	0	0
64	6	70	2	4	16

$\Sigma D^2 = 72$

$$\rho = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m)\right)}{n(n^2-1)}$$

$$\rho = 1 - \frac{6\left(72 + \frac{1}{12}(2^3-2) + \frac{1}{12}(3^3-3) + \frac{1}{12}(2^3-2)\right)}{10(99)}$$

$$\rho = 0.5454$$

Q Determine the rank correlation for the following data which shows the marks obtained in two quizzes in maths → $\rho = 0.5181$

x	6	5	8	8	7	6	10	4	9	7
y	8	7	7	10	5	8	10	6	8	6

x	Rank of X	Y	Rank of Y	$D = X - Y$	$D^2 = (X-Y)^2$
6	7.5	8	4	3.5	12.25
5	9	7	6.5	2.5	6.25
8	3.5	7	6.5	-3	9
8	3.5	10	1.5	2	4
7	5.5	5	10	-4.5	20.25
6	7.5	8	4	3.5	12.25
10	1	10	1.5	-0.5	0.25
4	10	6	8.5	1.5	2.25
9	2	8	4	-2	4
7	5.5	6	8.5	-3	9

$\Sigma D^2 = \cancel{40}\ 79.5$

$$\rho = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m)\right]}{n(n^2-1)}$$

$$\rho = 1 - \frac{6\left(79.5 + \frac{1}{12}(6) + \frac{1}{12}(6) + \frac{1}{12}(6) + \frac{1}{12}(6) + \frac{1}{12}(24) + \frac{1}{12}(6) + \frac{1}{12}(6)\right)}{10(99)}$$

$\rho = \cancel{0}\ 0.4878$

Q → A sample of 12 father and their elder sons gave the following data

→ $\rho = 0.722$

father	65	63	67	64	68	62	70	66	68	67	69	7
Son	68	66	68	65	69	66	68	65	71	67	68	7

father (x)	Rank of x	Son (y)	Rank of y	D = x − y	$D^2 = (x-y)^2$
65	9	68	5.5	3.5	12.25
63	11	66	9.5	1.5	2.25
67	6.5	68	5.5	1	1
64	10	65	11.5	−1.5	2.25
68	4.5	69	3	1.5	2.25
62	12	66	9.5	2.5	6.25
70	2	68	5.5	−3.5	12.25
66	8	65	11.5	−3.5	12.25
68	4.5	71	1	3.5	12.25
67	6.5	67	8	−1.5	2.25
69	3	68	5.5	−2.5	6.25
71	1	70	2	−1	1

$\Sigma D^2 = 72.5$

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

$$\rho = 1 - \frac{6\left(72.5 + \frac{1}{12}(6) + \frac{1}{12}(6) + \frac{1}{12}(4^3-4) + \frac{1}{12}(6) + \frac{1}{12}(6)\right)}{12(12^2-1)}$$

$$\rho = 0.722$$

2) The following table gives the score obtained by 11 students in physics & chemistry. Th find rank correlation coefficient

physics	40	46	54	60	70	80	82	85	86	90	95
chemistry	45	46	50	43	40	75	55	72	65	42	70

physics x	Rank of X	chemistry y	Rank of Y	$D = X - Y$	$D^2 = (X-Y)^2$
40	11	45	8	3	9
46	10	46	7	3	9
54	9	50	6	3	9
60	8	43	9	-1	1
70	7	40	11	-4	16
80	6	75	1	5	25
82	5	55	5	0	0
85	4	72	2	2	4
86	3	65	4	-1	1
90	2	42	10	-8	64
95	1	70	3	-2	4

$\Sigma D^2 = 142$

$$\rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)} \Rightarrow 1 - \frac{6(142)}{11(11^2-1)}$$

$$\rho = 0.3545$$

3) Ten Participants in a contest are ranked by two judges as follows

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

X	Rank of X	Y	Rank of Y	$D=(x-y)$	$D^2=(x-y)^2$
1	10	6	5	5	25
6	5	4	7	-2	4
5	6	9	2	4	16
10	1	8	3	-2	4
3	8	1	10	-2	4
2	9	2	9	0	0
4	7.	3	8	-1	1
9	2	10	1	1	1
7	4	5	6	-2	4
8	3	7	4	-1	1

$\Sigma D^2 = 60$

$$\rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)}$$

$$\rho = 1 - \frac{6(60)}{10(99)}$$

$$\rho = 0.63$$

Scanned with CamScanner

Q4 → find the rank correlation for the following data

x	56	42	72	36	63	47	55	49	38	42	68	60
y	147	125	160	118	149	128	150	145	115	140	152	155

X	Rank of X	Y	Rank of Y	D = X − Y	$D^2 = (X-Y)^2$
56	5	147	6	−1	1
42	9.5	125	10	−0.5	0.25
72	1	160	1	0	0
36	12	118	11	1	1
63	3	149	5	−2	4
47	8	128	9	−1	1
55	6	150	4	2	4
49	7	145	7	0	0
38	11	115	12	−1	1
42	9.5	140	8	1.5	2.25
68	2	152	3	−1	1
60	4	155	2	2	4

$\Sigma D^2 = 19.5$

$$\rho = 1 - \frac{6\left(\Sigma D^2 + \frac{1}{12}(m^3 - m)\right)}{n(n^2-1)}$$

$$\rho = 1 - \frac{6\left(19.5 + \frac{1}{12}(6)\right)}{12(143)}$$

$$\rho = 0.930$$

Scanned with CamScanner

21/10/2022

Ten competitors in a musical test were ran[illegible] by the three judges. A, B, C in the following order.

Rank A	1	6	5	10	3	2	4	9	7	8
rank B	3	5	8.	4	7	10	2	1	6	9
rank C	6	4	9	8	1	2	3	10	5	7

Using rank correlation method discuss which pair of judges has the nearest approach to common linkings in music.

Ranks by A x	Ranks by B y	Ranks by C (z)	D_1 $x-y$	D_2 $x-z$	D_3 $y-z$	D_1^2	D_2^2	D_3^2
1	3	6	-2	-5	-3	4	25	9
6	5	4	1	2	1	1	4	1
5	8	9	-3	-4	-1	9	16	1
10	4	8	6	2	-4	36	4	16
3	7	1	-4	2	6.	16	4	36
2	10	2	-8	0	8	64	0	64
4	2	3	2	1	-1	4	1	1
9	1	10	8	-1	-9	64	1	81
7	6	5	1	2	1	1	4	1
8	9	7	-1	1	2	1	1	4

$\Sigma D_1^2 = 200$ $\Sigma D_2^2 = 60$ $\Sigma D_3^2 = 21[illegible]$

$$\rho_k = 1 - \frac{6(\Sigma D_k^2)}{n(n^2-1)}$$

$$\rho_1 = 1 - \frac{6(200^2)}{10(99)} = -0.212$$

$$\rho_2 = 1 - \frac{6(60)}{990} = 0.6363$$

$$\rho_3 = 1 - \frac{6(214)}{990} = -0.2969$$

since $\rho_3(x,z)$ is max. we conclude that a pair of judge A and C has the nearest approach to common likings in music.

Curve Fitting

An approximate non mathematical relationship b/w the two variables can be established by a digram called scatter diagrame. The extra mathematical relationship b/w ~~too~~ the two variables is given by a simple algebraic expression called curve fitting

Straight line

The eq^n $y = a + bx$ is an eq^n of the first degree in x & y & it represents a st. line.

Fitting a St. line

Scanned with CamScanner

$y = a + bx$

→ where y = estimated trend value

x = time period

→ a & b are computed values and are obtained by solving the following two eqⁿs known as normal eqⁿ of straight lines.

$$\Sigma y = na + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

Q Fit a st. line to the following data

x	1	2	3	4	6	8
y	2.4	3	3.6	4	5	6

Sol $y = a + bx$ is st. line whose normal eqⁿ are

$$\Sigma y = na + b\Sigma x \quad \text{— (1)}$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{— (2)}$$

x	y	x^2	xy
1	2.4	1	2.4
2	3	4	6
3	3.6	9	10.8
4	4	16	16
6	5	36	30
8	6	64	48
$\Sigma x = 24$	$\Sigma y = 24$	$\Sigma x^2 = 130$	$\Sigma xy = 113.2$

Scanned with CamScanner

Here n = 6 & using all ~~[illegible]~~ values in normal eqⁿ we get

$24 = 6a + 24b$ —①

$113.2 = 24a + 130b$ —②

$a = 1.976$

$b = 0.505$

Therefore $y = 1.976 + 0.05058x$

Fit a st. line to the following data

x	4	6	8	10	12	14
y	14	24	30	31	42	60

$\Sigma y = na + b\Sigma x$ —①

$\Sigma xy = a\Sigma x + b\Sigma x^2$ —②

x	y	x^2	xy
4	14	16	56
6	24	36	144
8	30	64	240
10	31	100	310
12	42	144	504
14	60	196	840
$\Sigma x = 54$	$\Sigma y = 201$	$\Sigma x^2 = 556$	$\Sigma xy = 2094$

$n = 6$

$201 = 6a + 54b$

$2094 = 54a + 556b$

$a = -3.14$

$b = +4.07$

$y = -3.14 + 4.07x$

Q Fit the st. line to following data.

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

$y = 7.25 + 0.95x$

x	y	x^2	xy
1	9	1	9
2	8	4	16
3	10	9	30
4	12	16	48
5	11	25	55
6	13	36	78
7	14	49	98
8	16	64	128
9	15	81	135
$\Sigma x = 45$	$\Sigma y = 108$	$\Sigma x^2 = 285$	$\Sigma xy = 597$

$\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$108 = 9a + 45b$ — (1)

$597 = 45a + 285b$ — (2)

$a = 7.25$

$b = 0.95$

$y = a + bx$

$y = 7.25 + 0.95x$

Q Fit the st. line for the following data

x	-2	-1	0	1	2
y	1	2	3	3	4

Sol $\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

x	y	Σx^2	Σxy
-2	1	4	-2
-1	2	1	-2
0	3	0	0
1	3	1	3
2	4	4	8
$\Sigma x = 0$	$\Sigma y = 13$	$\Sigma x^2 = 10$	$\Sigma xy = 7$

$\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma bx + b\Sigma x^2$

$13 = 5a + 0$ — ①

$7 = 10b$ — ②

$a = 2.6$

$b = 0.7$

$y = a + bx$

$y = 2.6 + 0.7x.$

Scanned with CamScanner

Fitting of a Parabola

Write the normal eq^ns of 2° parabolic eq^n

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad -①$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad -②$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

→ Fit a parabola of 2nd degree to the following

x	0	1	2	3	4
y	1	1.8	1.3	2.5	6.3

Consider $y = a + bx + cx^2$ parabola of second degree whose normal eq^n are

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad -①$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \quad -②$$

$$\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad -③$$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	1.8	1	1	1	1.8	1.8
2	1.3	4	8	16	2.6	5.2
3	2.5	9	27	81	7.5	22.5
4	6.3	16	64	256	25.2	100.8

$\Sigma x = 10$ $\Sigma y = 12.9$ $\Sigma x^2 = 30$ $\Sigma x^3 = 100$ $\Sigma x^4 = 354$ $\Sigma xy = 37.1$ $\Sigma x^2 y = 130.3$

$$12.9 = 5a + 10b + 30c$$

$$37.1 = 10a + 30b + 100c$$

$$130.3 = 30a + 100b + 354c$$

$a = 1.42$

$b = -1.07$

$c = 0.55$

Hence $y = 1.42 - 1.07x + 0.55x^2$.

x	2	4	6	8	10
y	3.07	12.85	31.47	57.38	91.29

x	y	x^2	x^3	x^4	xy	x^2y
2	3.07	4	8	16	6.14	12.28
4	12.85	16	64	256	51.4	205.6
6	31.47	36	216	1296	188.82	1132.92
8	57.38	64	512	4096	459.04	3672.32
10	91.29	100	1000	10000	912.9	9129

$\Sigma x = 30$ $\Sigma y = 196.06$ $\Sigma x^2 = 220$ $\Sigma x^3 = 1800$ $\Sigma x^4 = 15664$ $\Sigma xy = 1618.3$

$\Sigma x^2 y = 14152$

$$196.06 = 5a + 30b + 220c$$

$$1618.3 = 30a + 220b + 1800c$$

$$14152.12 = 220a + 1800b + 15664c$$

$a = 0.696$

$b = -0.855$

$c = 0.991$

Hence $y = 0.696 - 0.855x + 0.991x^2$

QP - March/April 2022

Q → Fit the curve $y = a + bx + cx^2$ for following data

x	-1	0	1	2	3	7
y	9	7	7	9	13	49

x	y	x^2	x^3	x^4	xy	x^2y
-1	9	1	-1	1	-9	9
0	7	0	0	0	0	0
1	7	1	1	1	7	7
2	9	4	8	16	18	36
3	13	9	27	81	39	117
7	49	49	343	2401	343	2401

$\Sigma x = 12$ $\Sigma y = 94$ $\Sigma x^2 = 64$ $\Sigma x^3 = 378$ $\Sigma x^4 = 2500$ $\Sigma xy = 398$ $\Sigma x^2y = 2570$

$$94 = 6a + 12b + 64c$$

$$398 = 12a + 64b + 378c$$

$$2570 = 64a + 378b + 2500c$$

$$a = 7 \quad b = -1 \quad c = 1$$

$$y = 7 - x + x^2$$

Scanned with CamScanner

Q Fit a parabola $y = a + bx + cx^2$

x	1	2	3	4
y	1.7	1.8	2.3	3.2

x	y	x^2	x^3	x^4	xy	x^2y
1	1.7	1	1	1	1.7	1.7
2	1.8	4	8	16	3.6	7.2
3	2.3	9	27	81	6.9	20.7
4	3.2	16	64	256	12.8	51.2

$\Sigma x = 10$ $\Sigma y = 9.0$ $\Sigma x^2 = 30$ $\Sigma x^3 = 100$ $\Sigma x^4 = 354$ $\Sigma xy = 25$ $\Sigma x^2y = 80.$

$9 = 4a + 10b + 30c$

$25 = 10a + 30b + 100c$

$80.8 = 30a + 100b + 354c$

$a = 2$

$b = -0.5$

$c = 0.2$

$y = 2 - 0.5x + 0.2x^2$

Q Fit a Parabola $y = a + bx + cx^2$

x	-3	-1	1	3
y	15	5	1	5

x	y	x^2	x^3	x^4	xy	x^2y
-3	15	9	-27	81	-45	135
-1	5	1	-1	1	-5	5
1	1	1	1	1	1	1
3	5	9	27	81	15	45

$\Sigma x = 0$ $\Sigma y = 26$ $\Sigma x^2 = 20$ $\Sigma x^3 = 0$ $\Sigma x^4 = 164$ $\Sigma xy = -34$ $\Sigma x^2y = 186$

$\cancel{\Sigma y = na +}$

$26 = 4a + 20c$

$-34 = 20b$

$186 = 20a + 164c$

$a = 2.125$

$b = -1.7$

$c = 0.875$

$y = 2.125 - 1.7x + 0.875x^2$

Q → Fit the parabola $y = a + bx + cx^2$

x	-3	0	2	4
y	3	1	1	3

x	y	x^2	x^3	x^4	xy	x^2y
-3	3	9	-27	81	-9	27
0	1	0	0	0	0	0
2	1	4	8	16	2	4
4	3	16	64	256	12	48

$\Sigma x = 3$ $\Sigma y = 8$ $\Sigma x^2 = 29$ $\Sigma x^3 = 45$ $\Sigma x^4 = 353$ $\Sigma xy = 5$ $\Sigma x^2y = 79$

$8 = 4a + 3b + 29c$

$5 = 3a + 29b + 45c$

$79 = 29a + 45b + 353c$

$a = 0.85$

$b = -0.192$

$c = 0.178$

$y = 0.85 - 0.192x + 0.178x^2$

Fitting of exponential curves

(i) $y = ae^{bx}$

$y = ae^{bx}$

applying log on both sides

$\log y = \log_e ae^{bx}$

$\log y = \log_e a + \log_e e^{bx}$

$\log_e y = \log_e a + bx \log_e e$

$\log y = \log_e a + bx$

if we take

$Y = \log_e y$

$A = \log_e a \rightarrow a = e^{A}$

$BX = bx$

$Y = A + BX$ which represents st. line

normal eq^{n}s are $\Sigma Y = nA + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$A =$

$B =$

Q Fit an exponential curve of form $y = ae^{bx}$
to the following data

x	2	4	6	8
y	25	38	56	84

Sol computation table

x	y	$Y = \log y$	X^2	XY
2	25	3.2188	4	6.4376
4	38	3.6375	16	14.55
6	56	4.0253	36	26.1012
8	84	4.4308	64	35.4464

$\Sigma x = 20$ $\Sigma y = 203$ $\Sigma Y = 15.3124$ $\Sigma X^2 = 120$ $\Sigma XY = 80.5858$

Normal eq^{n}s = $\Sigma Y = nA + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$15.3124 = 4A + B20$

$80.5858 = 20A + 120B$

$A = 2.8229$

$B = 0.2010$

$b = B = 0.2010$

$a = e^A = 16.81211$

$$\boxed{y = 16.81211\, e^{0.2010x}}$$

Scanned with CamScanner

Q2 Fit an exponential curve of form $y = ae^{bx}$ to following data

x	1	2	3	4
y	4	11	35	100

x	y	$Y = \log_e y$	X^2	XY
1	4	1.3862	1	1.3862
2	11	2.3978	4	4.7956
3	35	3.5553	9	10.6659
4	100	4.6051	16	18.4204

$\Sigma x = 10$ $\quad \Sigma y = 150$ $\quad \Sigma Y = 11.9444$ $\quad \Sigma X^2 = 30$ $\quad \Sigma XY = 35.2681$

$$\Sigma Y = nA + B\Sigma X$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

$$11.9444 = 4A + 10B$$

$$35.2681 = 10A + 30B$$

$$A = 0.2825$$

$$B = 1.0814$$

$$b = B = 1.0814$$

$$a = e^A = 1.3264$$

$$\boxed{y = 1.3264\, e^{1.0814x}}$$

(3)

x	0.5	1.0	2	2.5	3
y	0.57	1.46	5.10	7.65	9.20

x	y	$Y = \log_e^y$	X^2	XY
0.5	0.57	-0.5621	0.25	-0.2810
1.0	1.46	0.3784	1	0.3784
2.0	5.10	1.6292	4	3.2584
2.5	7.65	2.0347	6.25	5.0867
3	9.20	2.2192	9	6.6576

$\Sigma x = 9$ $\Sigma y = 23.98$ $\Sigma Y = 5.6994$ $\Sigma X^2 = 20.5$ $\Sigma XY = 15.1001$

$$\Sigma Y = nA + B\Sigma X$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

$$5.6994 = 5A + 9B$$

$$15.1001 = 9A + 20.5B$$

$$A = -0.8866$$

$$B = 1.1258$$

$$B = b = 1.1258$$

$$a = e^A = 0.4120$$

$$y = 0.4120\, e^{1.1258x}$$

(ii) $y = ab^x$

Transforming the exponential eq^n $y = ab^x$ by taking ~~$\log_{10}$~~ logarithms to the base 10 on both sides. we get

$$\log_{10} y = \log_{10} a + \log_{10} b^x$$

$$\log_{10} y = \log_{10} a + x \log_{10} b$$

if we replace $\log_{10} y = Y$, $\log_{10} a = A$ & $x \log_{10} b = Bx$

then $Y = A + bx$ which defines st. lines whose ~~eq~~ normal eq^{n}s are $\Sigma Y = nA + B\Sigma x$

$$\Sigma xY = A\Sigma x + B\Sigma x^2$$

Q. Fit an exponential curve of the form $y = a$
for the following data.

x	1	2	3	4
y	4	11	35	100

x	y	$Y = \log_{10} y$	X^2	XY
1	4	0.6020	1	0.6020
2	11	1.0413	4	2.0826
3	35	1.5440	9	4.632
4	100	2	16	8

$\Sigma x = 10$ $\Sigma y = 150$ $\Sigma Y = 5.1873$ $\Sigma X^2 = 30$ $\Sigma XY = 15.3166$

$$\Sigma Y = nA + B\Sigma X$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2$$

$$5.1873 = 4A + 10B$$

$$15.3166 = 10A + 30B$$

$$A = 0.12265$$

$$B = 0.46967$$

$$a = 10^A = 1.3263$$

$$b = 10^B = 2.9489$$

$$y = ab^x$$

$$y = (1.3263)(2.9489)^x$$

(2)

x	1	2	3	4	5
y	1.6	4.5	13.8	40.2	125.0

x	y	$Y = \log_{10} y$	X^2	XY
1	1.6	0.2041	1	0.2041
2	4.5	0.6532	4	1.3064
3	13.8	1.1398	9	3.4494
4	40.2	1.6042	16	6.4168
5	125.0	2.0969	25	10.4845

$\Sigma x = 15$ $\Sigma y = 185.1$ $\Sigma Y = 5.6982$ $\Sigma X^2 = 55$ $\Sigma XY = 21.8312$

$$5.6982 = 5A + 15B$$

$$21.8312 = 15A + 55B$$

$$A = -0.2813$$

$$B = 0.4736$$

$$a = 10^A = 0.5232$$

$$b = 10^B = 2.9757$$

$$y = (0.5232)(2.9757)^x$$

(8) Fit an exponential curve of the form $y = ab^x$

x	61	26	7	26
y	350	400	500	600

x	y	$Y = \log_{10} y$	X^2	XY
61	350	2.5440	3721	155.184
26	400	2.6026	676	67.652
7	500	2.6989	49	18.8923
26	600	2.7781	676	72.2306

$\Sigma x = 120$ $\quad \Sigma y = 1850$ $\quad \Sigma Y = 10.623$ $\quad \Sigma X^2 = 5122$ $\quad \Sigma XY = 313.958$

$$10.623 = 4A + 120B$$

$$313.9589 = 120A + 5122B$$

$$A = 2.7490$$

$$B = -0.0031$$

$$a = 10^A = 561.04$$

$$b = 10^B = 0.9928$$

$$y = (561.04)(0.9928)^x$$

Test of Significance (or) Test of Hypothesis

→ A statistical hypothesis or simply hypothesis is an assumption about the parameters of the distributions and sometimes it is also concerns the type and nature of distribution.

→ Hypothesis testing enables us to make statement about population parameters.

→ Many experiments are carried out with the intention of testing a hypothesis.

→ Hypothesis is an assumption or statement which may or may not be true.

→ Hypothesis is a statement about a population parameter. It is an assumption made in order to arrive at a decision regarding population through a sample of population.

eg. 1) The avg. height of soldiers in the army is 165cm

2) The given machine has an effective life of 20 years

→ These entire hypothesis may be verified on the basis of certain sample test.

→ Procedures or tests which enable us to decide whether to accept or reject the hypothesis is called test of hypothesis or test of significance.

Scanned with CamScanner

→ To verify our assumption which is based on sample study we collect data & find out the difference between the sample value & the population value.

→ If there is a difference or the difference is very small then our hypothesized value is correct.

→ In general there are two types of hypothesis namely (i) Null hypothesis (H_0)
(ii) Alternative hypothesis (H_1)

☐ Null Hypothesis

→ The hypothesis formulated for the purpose of its rejection under the assumption that it is true is called Null Hypothesis. It is denoted by H_0.

☐ Alternative Hypothesis

→ The negetion of Null hypothesis is called Alternative hypothesis. It is denoted by H_1

☐ Degree of freedom (dof)

$$\boxed{V = n-1}$$

→ Degree of freedom is denoted by V and is defined by $\boxed{V = n-1}$. where n = no. of observations

☐ Chi - Square Test (χ^2)

→ Chi-Square test is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is observed frequency

E_i is expected frequency

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$$

→ Chi square test is used to check the significance of discrepancy between theory and experiment. It enables us to find the deviation of experiment to theory is just by chance or it is really due to inadequacy of the theory to fit the observed data.

→ Conditions under which Chi Square test is used

1) The total number of observation used in the test must be less than or equal to 30.

Scanned with CamScanner

2) The test is wholly dependent on degrees of freedom.

3) Each of the observations making up the Sample for chi square test should be independent of each other

4) The observation collected for chi square test must be based on method of random sampling.

Q Find the value of Chi-Square for the following data and test for goodness of fit (5% level of significance

observed frequency	14	15	18	20	15	10
expected frequency	17	10	15	25	10	15

$$\boxed{\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}}$$

$$\chi^2 = \frac{(14-17)^2}{17} + \frac{(15-10)^2}{10} + \frac{(18-5)^2}{15} + \frac{(20-25)^2}{25} + \frac{(15-10)^2}{10} + \frac{(10-15)^2}{15}$$

$$\chi^2 = \frac{9}{17} + \frac{25}{10} + \frac{9}{15} + \frac{25}{25} + \frac{25}{10} + \frac{25}{15}$$

$$\chi^2_{cal} = 8.7960$$

$$\boxed{\chi^2_{tab} = 11.070}$$

Since χ^2_{cal} ~~calculated~~ $< \chi^2_{tab}$ $\therefore$ we accept the null hypothesis and the fit is good

(2)

observed frequency	14	18	12	11	15	14
Expected frequency	14	14	14	14	14	14

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{(14-14)^2}{14} + \frac{(18-14)^2}{14} + \frac{(12-14)^2}{14} + \frac{(11-14)^2}{14} + \frac{(15-14)^2}{14} + \frac{(14-14)^2}{14}$$

$$= 0 + \frac{16}{14} + \frac{4}{14} + \frac{9}{14} + \frac{1}{14} + 0$$

$$= \frac{30}{14}$$

$$\chi^2_{cal} = 2.1428$$

$$\chi_{tab} = 11.070$$

$\therefore \chi^2_{cal} < \chi^2_{tab}$ we accept null hypothesis fit is good.

T-Test

When the population of std. deviation is not known and size of the sample is less than or equal to 30 we use t test

→ To test significance of mean sample the test statistics is given by

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

where $\bar{x} = \frac{\Sigma x_i}{n}$

$n =$ size of sample

$s =$ Standard deviation $= \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$

Q1 → Find the student t test for the follow variable values in a sample of eight −4, −2, −2, 0, 2, 2, 3, 3 using the mean universe to be zero ($\mu = 0$)

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{-4 - 2 - 2 + 0 + 2 + 2 + 3 + 3}{8}$$

$$= \frac{2}{8} = 0.25$$

x_i	$x_i - x_0$	$(x_i - x)^2$
-4	-4.25	18.0625
-2	-2.25	5.0625
-2	-2.25	5.0625
0	-0.25	0.0625
2	1.75	3.0625
2	1.75	3.0625
3	2.75	7.5625
3	2.75	7.5625

$\Sigma x = 2$ $\quad$ $\Sigma x - \bar{x} = 0$ $\quad$ $\Sigma (x-\bar{x})^2 = 49.5$

$$S = \sqrt{\frac{49.5}{7}}$$

$$S = 2.6592$$

$$t = \frac{\bar{x} - 0}{\left(\frac{S}{\sqrt{n}}\right)}$$

$$t = \frac{0.25 - 0}{\frac{2.6592}{\sqrt{8}}}$$

$$t_{cal} = 0.2659$$

$$t_{tab} = 1.895$$

$t_{cal} < t_{tab}$ we accept null hypothesis

② A Random sample of 10 boys had IQ
70, 120, 110, 101, 88, 83, 95, 98, 107, 100
Do these data support the assumption of population mean of 160 (5%)

$\bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2$

x	$x-\bar{x}$	$(x-\bar{x})^2$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84

$\Sigma x = 972$ $\quad \Sigma x - \bar{x} = 0$ $\quad \Sigma(x-\bar{x})^2 = 1833.6$

$s = \sqrt{\frac{1833.6}{9}}$

$s = 14.2735$

$t = \frac{97.2 - 160}{\left(\frac{14.2735}{\sqrt{10}}\right)}$

$t_{cal} = -13.91$

~~tcal~~ as IQ cannat be negetive

$t_{cal} = 13.91$

~~too~~ $t_{tab} = 1.833$

$t_{cal} > t_{tab}$. alternative hypothesis

Data does not support the assumption of population mean of 160. ~~(150)~~