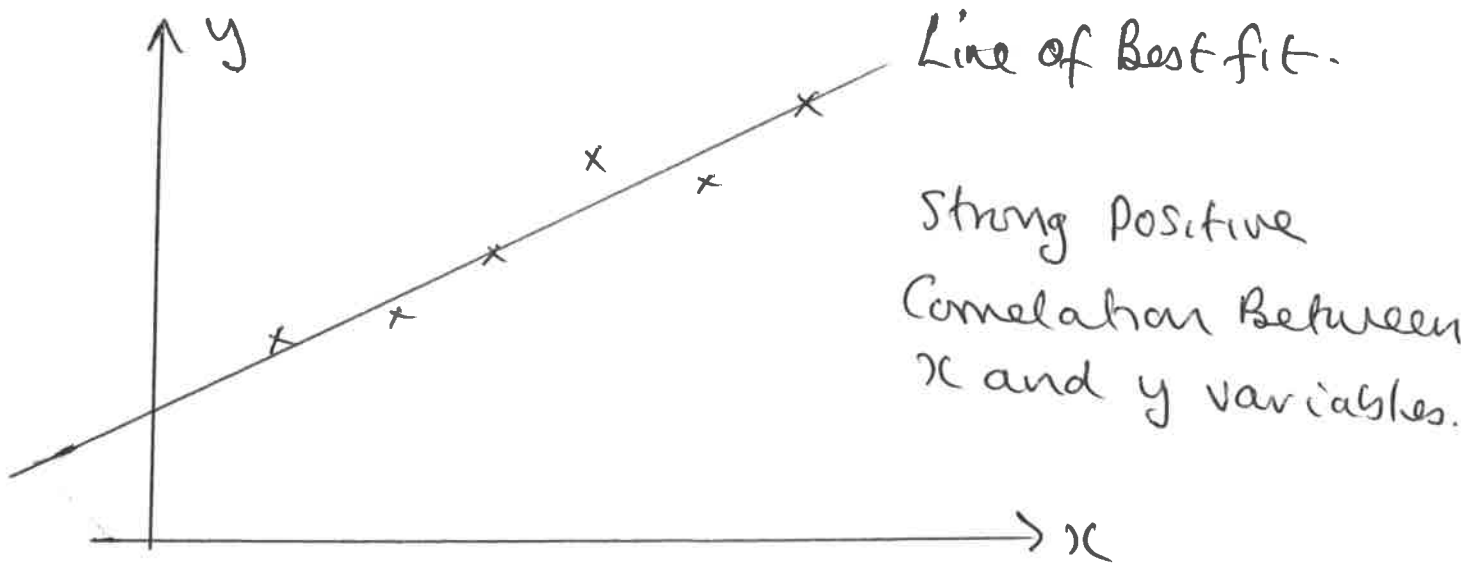


Least Squares: Regression Lines.



The Line of Best fit by Calculation is given by:

$$y - \bar{y} = \frac{S_{xy}}{S_{xx}} (x - \bar{x})$$

Where $\bar{x} = \frac{1}{n} \sum_1^1 x$ The Mean of the x values
 $\bar{y} = \frac{1}{n} \sum_1^1 y$ Mean of y values

$S_{xx} = \frac{1}{n} \sum_1^1 x^2 - (\bar{x})^2$ Variance of x

$S_{xy} = \frac{1}{n} \sum_1^1 xy - (\bar{x})(\bar{y})$ Covariance of x and y .

Lines can also show strong negative Correlation.
i.e. Slope Downwards.

Least Squares Regression Lines

1

x	1	2	3	4
y	1	3	5	7

$$y = 2x - 1$$

2

x	1	2	3	4
y	4.9	7.1	8.8	11.1

$$y = 2.03x - 2.9$$

3

x	1	2	3	4
y	3.9	7.2	9.9	12.8

$$y = 2.94x - 1.1$$

4

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

$$y = -2x + 1$$

x	1	2	3	4
y	1	3	5	7

x	y	x^2	xy	y^2
1	1	1	1	1
2	3	4	6	9
3	5	9	15	25
4	7	16	28	49
$\sum x = 10$		$\sum y = 16$	$\sum x^2 = 30$	$\sum xy = 50$
				$\sum y^2 = 84$

Least Squares Regression line is

$$\hat{y} - \bar{y} = \frac{S_{xy}}{S_{xx}} (x - \bar{x})$$

$$\bar{y} = \frac{1}{n} \sum y$$

$$\bar{y} = \frac{1}{4} (16) = 4$$

$$\bar{x} = \frac{1}{n} \sum x$$

$$\bar{x} = \frac{1}{4} (10)$$

$$\bar{x} = 2.5$$

$$\begin{aligned}
 S_{xx} &= \frac{1}{n} \sum x^2 - (\bar{x})^2 \\
 &= \frac{1}{4} (30) - (2.5)^2 \\
 &= 1.25
 \end{aligned}$$

$$\begin{aligned}
 S_{xy} &= \frac{1}{n} \sum_i xy - (\bar{x})(\bar{y}) \\
 &= \frac{1}{4} (50) - (2.5)(4) \\
 &= 2.5
 \end{aligned}$$

∴ Regression line y on x is:

$$y - 4 = \frac{2.5}{1.25} (x - 2.5)$$

$$y - 4 = 2(x - 2.5)$$

$$y - 4 = 2x - 5$$

$$\underline{\underline{y = 2x - 1}}$$

$$r_p = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

$$\begin{aligned}
 S_{yy} &= \frac{1}{n} \sum_i y^2 - (\bar{y})^2 \\
 &= \frac{1}{4} (84) - (4)^2 \\
 &= 21 - 16 \\
 &= 5
 \end{aligned}$$

$$r_p = \frac{2.5}{\sqrt{1.25 \times 5}}$$

$$= \frac{2.5}{2.5}$$

$$= +1$$

Perfect 'true' Correlation

2/

x	1	2	3	4
y	4.9	7.1	8.8	11.1

x	y	x^2	y^2	xy
1	4.9	1	24.01	4.9
2	7.1	4	50.41	14.2
3	8.8	9	77.44	26.4
4	11.1	16	123.21	44.4

$$\sum x = 10 \quad \sum y = 31.9 \quad \sum x^2 = 30 \quad \sum y^2 = 275.07 \quad \sum xy = 89.9$$

$$\bar{x} = \frac{1}{n} \sum x \quad \frac{1}{n} \sum y = \bar{y}$$

$$\bar{x} = \frac{1}{4} (10) \quad \frac{1}{4} (31.9) = \bar{y}$$

$$\bar{x} = 2.5 \quad 7.975 = \bar{y}$$

$$\begin{aligned} S_{xx} &= \frac{1}{n} \sum x^2 - (\bar{x})^2 \\ &= \frac{1}{4} (30) - (2.5)^2 \\ &= 1.25 \end{aligned}$$

$$\begin{aligned} S_{yy} &= \frac{1}{n} \sum y^2 - (\bar{y})^2 \\ &= \frac{1}{4} (275.07) - (7.975)^2 \\ &= 5.1669 \text{ (4 d.p's)} \end{aligned}$$

$$\begin{aligned}
 S_{xy} &= \frac{1}{n} \sum_i xy - (\bar{x})(\bar{y}) \\
 &= \frac{1}{4} (89.9) - (2.5)(7.975) \\
 &= \underline{2.5375}
 \end{aligned}$$

least squares Regression line is:-

$$y - \bar{y} = \frac{S_{xy}}{S_{xx}} (x - \bar{x})$$

$$y - 7.975 = \frac{2.5375}{1.25} (x - 2.5)$$

$$y - 7.975 = 2.03 (x - 2.5)$$

$$y - 7.975 = 2.03x - 5.075$$

$$y = 2.03x - 5.075 + 7.975$$

$$\underline{y = 2.03x - 2.9}$$

$$r_p = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{2.5375}{\sqrt{1.25 \times 5.1669}}$$

$$= \underline{0.9985}$$

Almost Perfect Positive Correlation.

3/

x	1	2	3	4
y	3.9	7.2	9.9	12.8

x	y	x^2	y^2	xy
1	3.9	1	15.21	3.9
2	7.2	4	51.84	14.4
3	9.9	9	98.01	29.7
4	12.8	16	163.84	51.2

$$\sum x = 10 \quad \sum y = 33.8 \quad \sum x^2 = 30 \quad \sum y^2 = 328.9 \quad \sum xy = 99.2$$

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum x \\ &= \frac{1}{4} (10) \\ &= \underline{2.5}\end{aligned}$$

$$\begin{aligned}\bar{y} &= \frac{1}{n} \sum y \\ &= \frac{1}{4} (33.8) \\ &= \underline{8.45}\end{aligned}$$

$$\begin{aligned}S_{xx} &= \frac{1}{n} \sum x^2 - (\bar{x})^2 \\ &= \frac{1}{4} (30) - (2.5)^2 \\ &= \underline{1.25}\end{aligned}$$

$$\begin{aligned}S_{yy} &= \frac{1}{n} \sum y^2 - (\bar{y})^2 \\ &= \frac{1}{4} (328.9) - (8.45)^2 \\ &= \underline{10.8225}\end{aligned}$$

$$\begin{aligned}
 S_{xy} &= \frac{1}{n} \sum_i xy - (\bar{x})(\bar{y}) \\
 &= \frac{1}{4} (99.2) - (2.5)(8.45) \\
 &= \underline{3.675}
 \end{aligned}$$

Then least squares regression line is:

$$y - \bar{y} = \frac{S_{xy}}{S_{xx}} (x - \bar{x})$$

$$y - 8.45 = \frac{3.675}{1.25} (x - 2.5)$$

$$y - 8.45 = 2.94(x - 2.5)$$

$$y - 8.45 = 2.94x - 7.35$$

$$y = 2.94x - 7.35 + 8.45$$

$$\underline{y = 2.94x - 1.1}$$

$$r_p = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{3.675}{\sqrt{1.25 \times 10.8225}} = +0.9992$$

Strong the Correlation almost Perfect.

$$y = -2x + 1$$

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

x	y	x^2	y^2	xy
1	-1	1	1	-1
2	-3	4	9	-6
3	-5	9	25	-15
4	-7	16	49	-28
$\sum x = 10$ $\sum y = -16$ $\sum x^2 = 30$ $\sum y^2 = 84$ $\sum xy = -50$				

$$\bar{x} = \frac{1}{n} \sum x$$

$$\bar{x} = \frac{1}{4} (10)$$

$$\bar{x} = 2.5$$

$$\bar{y} = \frac{1}{n} \sum y$$

$$\bar{y} = \frac{1}{4} (-16)$$

$$\bar{y} = -4$$

$$\begin{aligned}
 S_{xx} &= \frac{1}{n} \sum x^2 - (\bar{x})^2 \\
 &= \frac{1}{4} (30) - (2.5)^2 \\
 &= 1.25
 \end{aligned}$$

$$\begin{aligned}
 S_{yy} &= \frac{1}{n} \sum y^2 - (\bar{y})^2 \\
 &= \frac{1}{4} (84) - (-4)^2 \\
 &= 5
 \end{aligned}$$

$$\begin{aligned}
 S_{xy} &= \frac{1}{n} \sum_1^1 xy - (\bar{x})(\bar{y}) \\
 &= \frac{1}{4} (-50) - (2.5)(-4) \\
 &= -12.5 + 10 \\
 &= \underline{-2.5}
 \end{aligned}$$

Least Squares Regression line is -

$$y - \bar{y} = \frac{S_{xy}}{S_{xx}} (x - \bar{x})$$

$$y - (-4) = \frac{-2.5}{1.25} (x - 2.5)$$

$$y + 4 = -2(x - 2.5)$$

$$y + 4 = -2x + 5$$

$$\underline{y = -2x + 1}$$

$$r_p = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

$$r_p = \frac{-2.5}{\sqrt{1.25 \times 5}}$$

$$\underline{r_p = -1}$$

Perfect 've' Correlation.