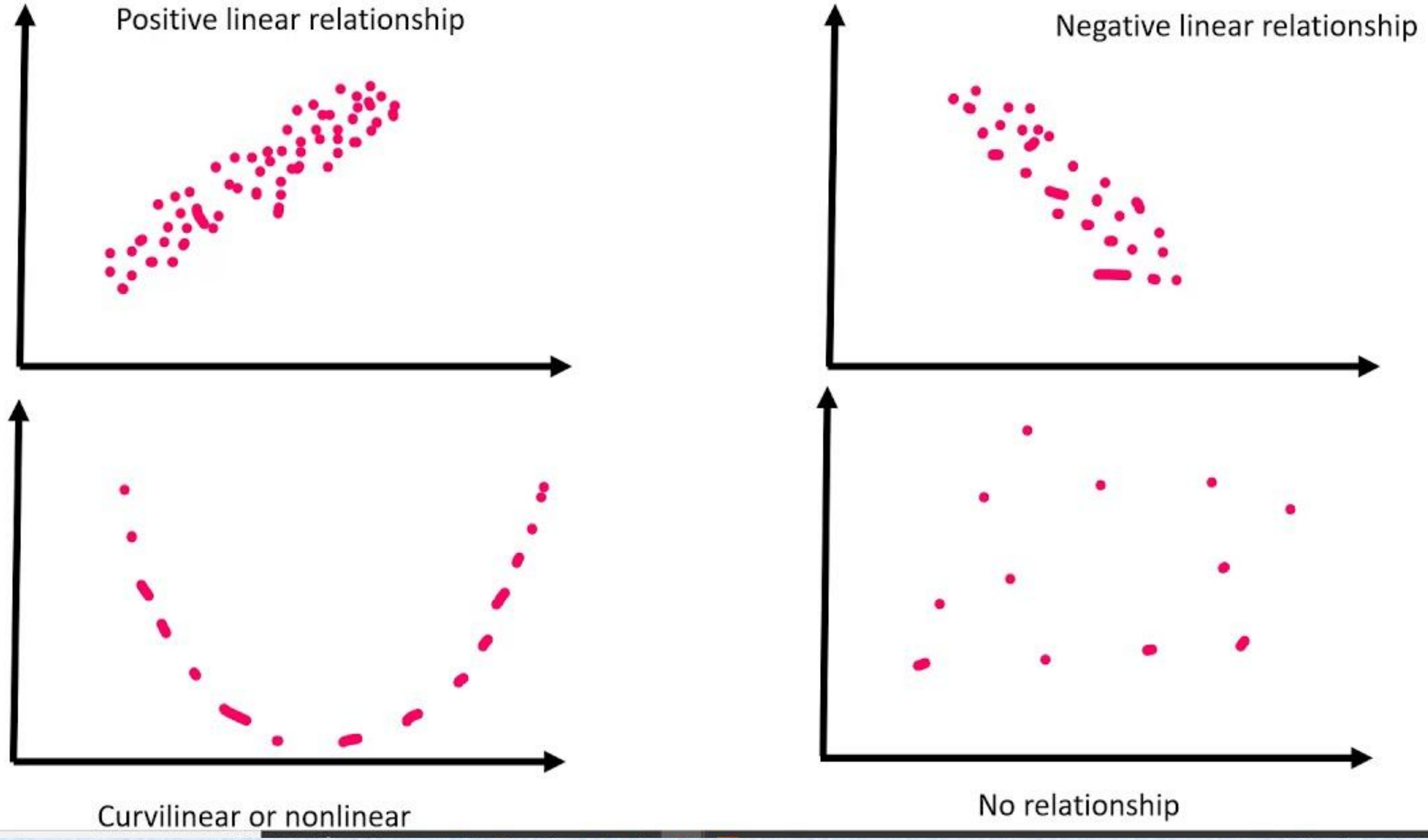
PRODUCT MOMENT CORRELATION

CORRELATION

- It measures the extent to which two or more than two variables are related.
- Correlation shows strength and direction of correlation.
- There are three possible outcome of a correlational study
  1. Positive Correlation (↑↑ or ↓↓)
  2. Negative Correlation (↑↓)
  3. Zero Correlation (no)
- Range : -1 to +1

Various patterns of scatter plots



STRENGTH OF CORRELATION

Negative Range	Strength of association	Positive Range
-1.00	**Perfect** Correlation	+1.00
-0.80 – 1.00	**Very High** Correlation	+0.80 – 1.00
-0.60 – 0.80	**High** Correlation	+0.60 – 0.80
-0.40 – 0.60	**Average** Correlation	+0.40 – 0.60
-0.20 – 0.40	**Low** Correlation	+0.20 – 0.40
-0.20 and below	**Very Low** Correlation	+0.20 and below

PRODUCT MOMENT CORRELATION (r)

- Devised by **Karl Pearson,** also known as Pearson Product Moment Correlation or Pearson Method
- It is an index of the degree of linear relationship between two variable.

- **Uses :-**

1. Prediction
2. Validity
3. Reliability
4. Theory verification

1. REAL MEAN METHOD

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

r = Pearson Correlation

x = X - $\bar{X}$ i.e. Deviation of a score of X-series from its mean

y = Y - $\bar{Y}$ i.e. Deviation of a score of Y-series from its mean

$\sum xy$ = Total of x multiplied by y

$\sum x^2$ = Sum of square deviation of x

$\sum y^2$ = Sum of square deviation of y

For example :-

0	2	-6	-2	36	4	12
10	6	4	2	16	4	8
4	2	-2	-2	4	4	4
8	4	2	0	4	0	0
8	6	2	2	4	4	4
30	**20**			**64**	**16**	**28**

$$\bar{X} = \frac{\sum X}{N} = \frac{30}{5} = 6$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{20}{5} = 4$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = \frac{28}{\sqrt{64 \times 16}} = \frac{28}{32} = 0.875$$

There is a very high positive correlation between X and Y.

2. DEVIATION FROM ZERO METHOD

$$r = \frac{\sum XY - NXY}{\sqrt{(\sum X^2 - N\bar{X}^2)(\sum Y^2 - N\bar{Y}^2)}}$$

r = Pearson Correlation

$\sum XY$ = The sum of the products of the scores X multiplied by score Y

N = Number of subjects

$\bar{X}$ = Mean of X-series

$\bar{Y}$ = Mean of Y-series

$\sum X^2$ = Sum of square deviation of x

$\sum Y^2$ = Sum of square deviation of y

$\bar{X}^2$ = The square of the mean of the X-Series.

$\bar{Y}^2$ = The square of the mean of the Y-Series.

For example :-

0	2	0	4	0
10	6	100	36	60
4	2	16	4	8
8	4	64	16	32
8	6	64	36	48
30	**20**	**244**	**96**	**148**

$$\bar{X} = \frac{\sum X}{N} = \frac{30}{5} = 6 \qquad \bar{Y} = \frac{\sum Y}{N} = \frac{20}{5} = 4$$

$$r = \frac{\sum XY - N\bar{X}\bar{Y}}{\sqrt{(\sum X^2 - N\bar{X}^2)(\sum Y^2 - N\bar{Y}^2)}} = \frac{148 - 5 \text{ x } 6 \text{ x } 4}{\sqrt{(244 - 5 \text{ x } 6^2)(96 - 5 \text{ x } 4^2)}} = \frac{148 - 120}{\sqrt{(244 - 180)(96 - 80)}} = \frac{28}{32}$$

$= \mathbf{0.875}$

There is a very high positive correlation between X and Y.

3. ASSUMED MEAN METHOD

$$r = \frac{\frac{\sum xy}{N} - C_X C_y}{\sigma_x \, \sigma_y}$$

$$C_X = \frac{\sum x}{N} \qquad C_y = \frac{\sum y}{N} \qquad \sigma_x = \sqrt{\frac{\sum x^2}{N} - c_x^2} \qquad \sigma_y = \sqrt{\frac{\sum y^2}{N} - c_y^2}$$

r = Pearson Correlation

x = X - A i.e. Deviation of a score of X-series from assumed mean of that series.

y = Y - A i.e. Deviation of a score of Y-series from assumed mean of that series.

$\sum xy$ = Total x multiplied by y

N = Number of subjects

C_X = Correction of X-series

C_y = Correction of Y-series

σ_x = Standard Deviation of X-series

σ_y = Standard Deviation of Y-series

For example :-

0	2	-6	-2	36	4	12
10	6	4	2	16	4	8
4	2	-2	-2	4	4	4
8	4	2	0	4	0	0
8	6	2	2	4	4	4
30	**20**	**0**	**0**	**64**	**16**	**28**

$$\bar{X} = \frac{\sum X}{N} = \frac{30}{5} = 6$$

$$A_x = 6$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{20}{5} = 4$$

$$A_y = 4$$

$$C_X = \frac{\sum x}{N} = \frac{0}{5} = 0$$

$$c_x^2 = 0$$

$$C_y = \frac{\sum y}{N} = \frac{0}{5} = 0$$

$$c_y^2 = 0$$

$$\sigma_x = \sqrt{\frac{\sum x^2}{N} - c_x^2} = \sqrt{\frac{64}{5} - 0} = 3.577$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{N} - c_y^2} = \sqrt{\frac{16}{5} - 0} = 1.788$$

$$r = \frac{\frac{\sum xy}{N} - C_X C_y}{\sigma_x \sigma_y} = \frac{\frac{28}{5} - 0 \text{ x } 0}{3.577 \text{ x } 1.788} = \frac{5.6}{6.395} = 0.875$$

There is a very high positive correlation between X and Y.

4. REDUCED SCORE METHOD

- $$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{[N\sum x^2 - (\sum x)^2]\,[N\sum y^2 - (\sum y)^2]}}$$

r = Pearson Correlation

x = X - A i.e. Deviation of a score of X-series from reduced mean of that series.

y = Y - A i.e. Deviation of a score of Y-series from reduced mean of that series.

$\sum xy$ = Total x multiplied by y

N = Number of subjects

$\sum x^2$ = Sum of squares of X scores

$\sum y^2$ = Sum of squares of Y scores

$\sum x$ = Sum of X scores

$\sum y$ = Sum of Y scores

For example :-

0	2	-6	-2	36	4	12
10	6	4	2	16	4	8
4	2	-2	-2	4	4	4
8	4	2	0	4	0	0
8	6	2	2	4	4	4
30	**20**	**0**	**0**	**64**	**16**	**28**

$$\bar{X} = \frac{\sum X}{N} = \frac{30}{5} = 6$$

Reduced Mean = 6

$$\bar{Y} = \frac{\sum Y}{N} = \frac{20}{5} = 4$$

Reduced Mean = 6

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}} = \frac{5 \times 28 - 0 \times 0}{\sqrt{(5 \times 64 - 0^2)(5 \times 16 - 0^2)}} = \frac{140 - 0}{\sqrt{320 \times 80}} = \frac{140}{160} = \mathbf{0.875}$$

There is a very high positive correlation between X and Y.

5. DIFFERENCE METHOD

$$r = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \ \sum y^2}}$$

r = Pearson Correlation

$\sum x^2$ = Sum of squares of deviation of scores of X-series from the mean

$\sum y^2$ = Sum of squares of deviation of scores of Y-series from the mean

$\sum d^2$ = Sum of squares of difference between x and y deviations

For example :-

0	2	-6	-2	36	4	-4	16
10	6	4	2	16	4	2	4
4	2	-2	-2	4	4	0	0
8	4	2	0	4	0	2	4
8	6	2	2	4	4	0	0
30	**20**	**0**	**0**	**64**	**16**		**24**

$$\bar{X} = \frac{\sum X}{N} = \frac{30}{5} = 6$$

$$\bar{Y} = \frac{\sum Y}{N} = \frac{20}{5} = 4$$

$$r = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \ \sum y^2}} = \frac{64+16-24}{2\sqrt{64 \times 16}} = \frac{56}{2 \times 32} = 0.875$$

There is a very high positive correlation between X and Y.

THANK YOU