# Module: 3 Correlation and Regression

- Correlation and Regression - Rank Correlation- Partial and Multiple correlation- Multiple regression.

## Correlation:

Correlation is a statistical technique to study the relation between two or more variables. It is denoted by ‘$r$’. It always lies between -1 to +1.

## Types of Correlation

1. Positive Correlation
2. Negative Correlation
3. No Correlation

## Positive Correlation:

If the value of one variable increases then the value of another variable increases automatically then there is positive correlation between that two variable. **(or)** If the value of two variable changes in the same direction then there is positive correlation between two variable and in this case $r \geq 0$.

## Negative Correlation:

If the value of one variable increases then the value of another variable decreases automatically then there is negative correlation between two variables. **(or)** If the value two variable changes but in opposite direction then there is negative correlation between two variable and in this case $r \leq 0$**.**

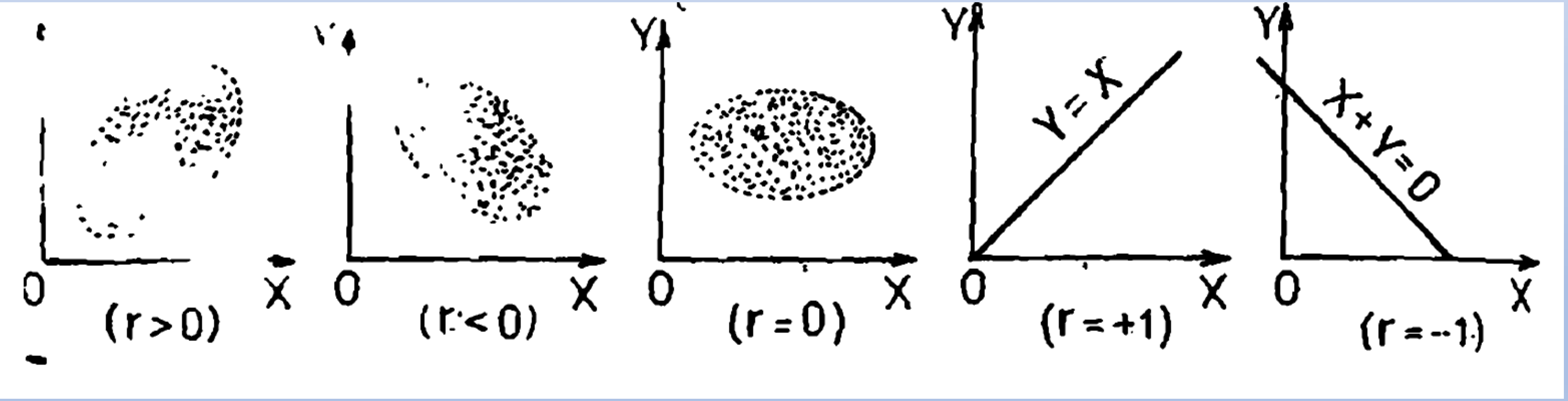
## No Correlation:

If the presence of one variable does not affect on the presence or absence of another variable then there is no correlation between two variables. In this case $r = 0$.

**Methods to Study The Correlation:**

1. Scatter Diagram
2. Karl Pearson's Correlation Coefficient (r)
3. Spearman Rank Correlation Coefficient (R)

1. **Scatter Diagram**



**Karl Pearson Coefficient of Correlation:**

Correlation coefficient between two random variables $X$ and $Y$, usually denoted by $r(X,Y)$ or simply $r_{XY}$. is a numerical measure of linear relationship between them and is defined as

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

$$Cov(X,Y)=E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]$$

$$Cov(X,Y)=\frac{1}{N}\sum_i (x_i-\bar{x})(y_i-\bar{y})$$

$$\sigma_X=\sqrt{\frac{1}{N}\sum_i (x_i-\bar{x})^2}$$

$$\sigma_Y=\sqrt{\frac{1}{N}\sum_i (y_i-\bar{y})^2}$$

or

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

or

$$r(X,Y)=\frac{N\sum XY-\sum X\sum Y}{\sqrt{N\sum X^2-\left(\sum X\right)^2}\sqrt{N\sum Y^2-\left(\sum Y\right)^2}}$$

**Properties of Karl Pearson’s Correlation Coefficient (r):**

1. It always lies between $-1$ to $+1$.
2. It is denoted by “$r$”.
3. It is independent of change of origin and scale.
4. If $r = +1$ then there is perfect positive correlation coefficient between two variable.
5. If $r = -1$ then there is perfect negative correlation coefficient between two variable.
6. If $r = 0$ then there is no correlation coefficient between two variable.
7. If $r > 0$ then there is positive correlation coefficient between two variable.
8. If $r < 0$ then there is negative correlation coefficient between two variable.

*Calculate the correlation coefficient for the following heights (in inches) of fathers (X) and their sons (Y) :*

| $X$ | 65 | 66 | 67 | 67 | 68 | 69 | 70 | 72 |
|---|---|---|---|---|---|---|---|---|
| $Y$ | 67 | 68 | 65 | 68 | 72 | 72 | 69 | 71 |

Solution:

| $X$ | $Y$ | $X^2$ | $Y^2$ | $XY$ |
|---|---|---|---|---|
| 65 | 67 | 4225 | 4489 | 4355 |
| 66 | 68 | 4356 | 4624 | 4488 |
| 67 | 65 | 4489 | 4225 | 4355 |
| 67 | 68 | 4489 | 4624 | 4556 |
| 68 | 72 | 4624 | 5184 | 4896 |
| 69 | 72 | 4761 | 5184 | 4968 |
| 70 | 69 | 4900 | 4761 | 4830 |
| 72 | 71 | 5184 | 5041 | 5112 |
| 544 | 552 | 37028 | 38132 | 37560 |

$$r(X,Y)=\frac{N\sum XY-\sum X\sum Y}{\sqrt{N\sum X^2-\left(\sum X\right)^2}\sqrt{N\sum Y^2-\left(\sum Y\right)^2}}$$

$$=\frac{8\times 37560-544\times 552}{\sqrt{8\times 37028-544^2}\sqrt{8\times 38132-552^2}}$$

$$=\frac{192}{16.9705\times 18.7616}$$

$$=0.6030$$

Compute the product moment coefficient of correlation for the following data: $N = 100, \bar{x}=62,\ \bar{y} = 53, \sigma_x = 10, \sigma_y = 12, \sum(x-\bar{x})(y-\bar{y}) = 8000.$

Solution:

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

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

$$r(X,Y) = \frac{\frac{\sum(X-\bar{X})(Y-\bar{Y})}{N}}{\sigma_X \sigma_Y}$$

$$r = \frac{8000/100}{10/12}$$

$$r = 80/120$$

$$r = 0.6667$$

**Spearman Rank Correlation Coefficient (R):**

The Psychologist Spearman suggested formula to calculate correlation for qualitative data by considering the ranks,

$$R = 1 - \frac{6D^2}{N\left(N^2 - 1\right)}$$

where

$$D = R_X - R_Y$$

Calculate Spearman Rank correlation coefficient for the following information:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Marks in Mathematics | 67 | 57 | 60 | 76 | 79 | 64 | 73 |
| Marks in Accountancy | 78 | 75 | 66 | 70 | 74 | 72 | 69 |

| X | Y | $R_1$ | $R_2$ | $D=R_1-R_2$ | $D^2$ |
|---|---|---|---|---|---|
| 67 | 78 | 4 | 1 | 3 | 9 |
| 57 | 75 | 7 | 2 | 5 | 25 |
| 60 | 66 | 6 | 7 | -1 | 1 |
| 76 | 70 | 2 | 5 | -3 | 9 |
| 79 | 74 | 1 | 3 | -2 | 4 |
| 64 | 72 | 5 | 4 | 1 | 1 |
| 73 | 69 | 3 | 6 | -3 | 9 |
| | | | | | $\sum D^2 = 57$ |

**Here N=7**

$$R = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

$$R = 1 - \frac{6\,X57}{7(7^2-1)}$$

$$R = 1 - \frac{342}{7(49-1)}$$

$$R = 1 - \frac{342}{7(48)}$$

$$R = 1 - \frac{342}{336}$$

$$R = 1 - 1.017857$$

$$R = -0.017855$$

**Comment:** There is negative correlation between marks getting in the subject Mathematics &

Accountancy