UNIT-5 Correlation Coefficient

Dinesh Kumar Baghel

November 27, 2025

Table of contents

Correlation Coefficient

- Correlation is a standardized measure of the *strength* and *direction* of a linear relationship between two variables.
- Covariance is *unstandardized* and *scale-dependent*
- Overall, correlation is just a rescaled version of covariance and always lies between -1 and 1.
- Scatter diagram gives a vague idea about whether the variables are corrleated or not. Not suitable when the number of observations are large.

Karl Pearson's Coefficient of Correlation

Definition 1

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

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

where σ_X and σ_Y are the standard deviations of X and Y respectively.

Correlation Coefficient

Limits of correlation coefficient

$$-1 \leq r(X, Y) \leq 1$$

$$r(X, Y) = \begin{cases} -1, & \text{if } X = -aY (a > 0), \\ 0, & \text{if } X \text{ and } Y \text{ are uncorrelated}, \\ +1, & \text{if } X = aY (a > 0) \end{cases}$$

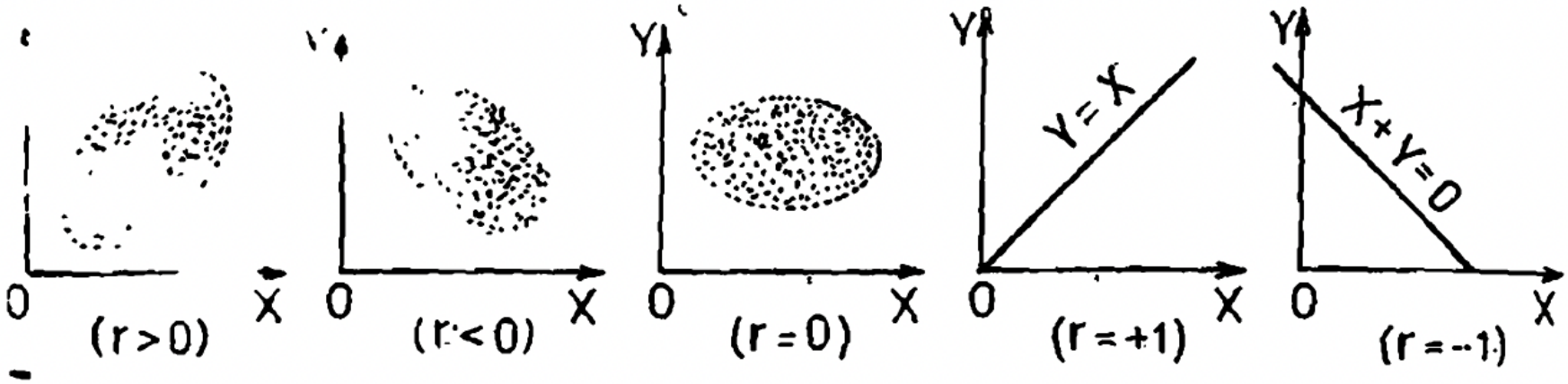

Figure 1: Figures of the standard data for r

- The correlation coefficient is independent of change of origin and scale.

- Two independent random variables are uncorrelated.

- $r(X, Y) = 1 \Rightarrow$ perfect positive linear relationship; $r(X, Y) = -1 \Rightarrow$ perfect negative linear relationship; $r(X, Y) = 0 \Rightarrow$ no linear relationship.
  NOTE: A perfect correlation means that the deviation in one variable is followed by a corresponding and proportional deviation in the other.

- Correlation can be compared across different variable pairs and is unaffected by the linear scaling of X and Y.

Covariance vs Correlation

Conceptual

- Covariance tells direction of joint linear variation
- Correlation tells both direction and strength

Scale and Units

- Covariance has units (units of X and Y; their product)
- Magnitude of covariance depends on scale
- Correlation has no units
- It is invariant under linear scaling of either variable.

Covariance vs Correlation

Range

- Covariance has range from $-\infty$ to ∞
- Correlation has range from -1 to $+1$.

– Correlation is more interpretable and comparable across different datasets.

Example

Q1: 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

Example

Q1: 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

$Cov(X, Y) = \frac{1}{n}\sum XY - \bar{X}\bar{Y}$

$\sigma_X = \frac{1}{n}\sum X^2 - \bar{X}^2$

References I