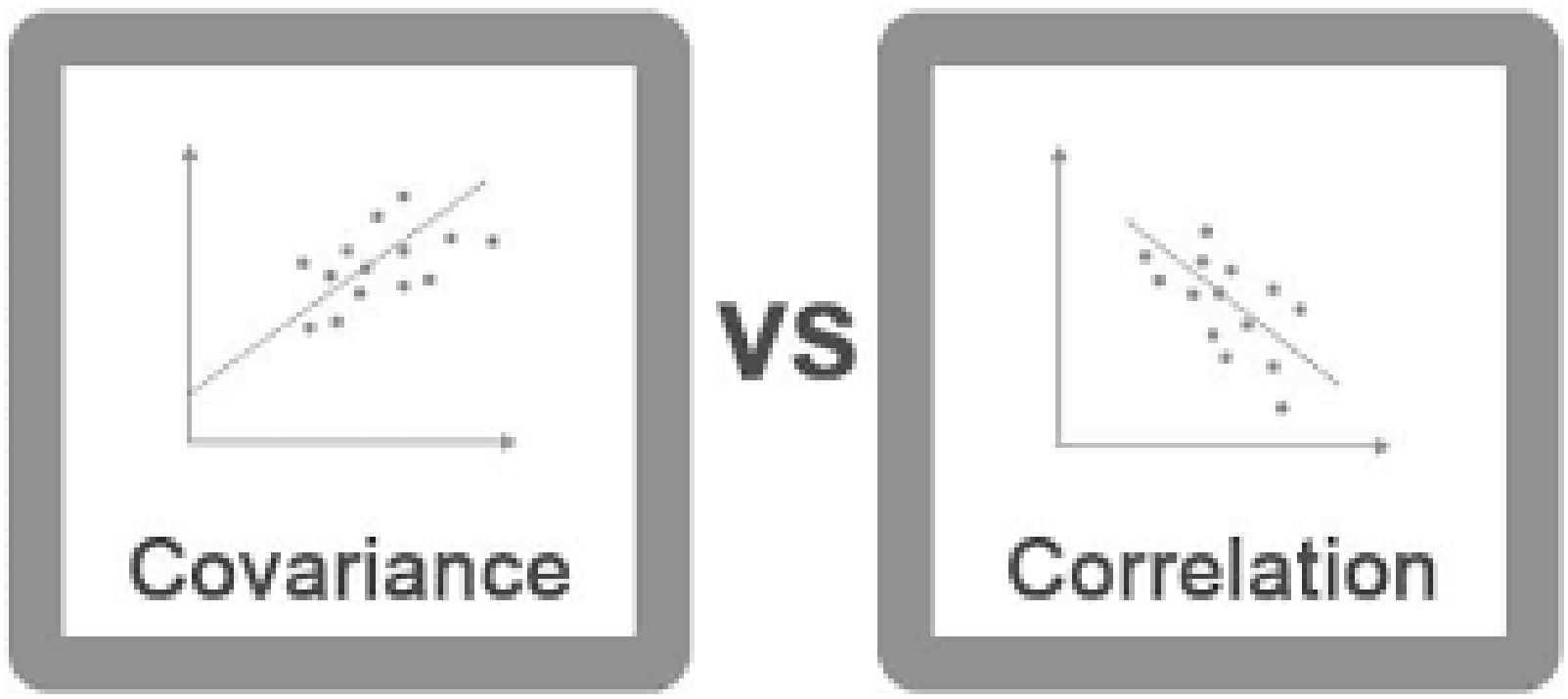


What is the Difference Between Covariance and Correlation?

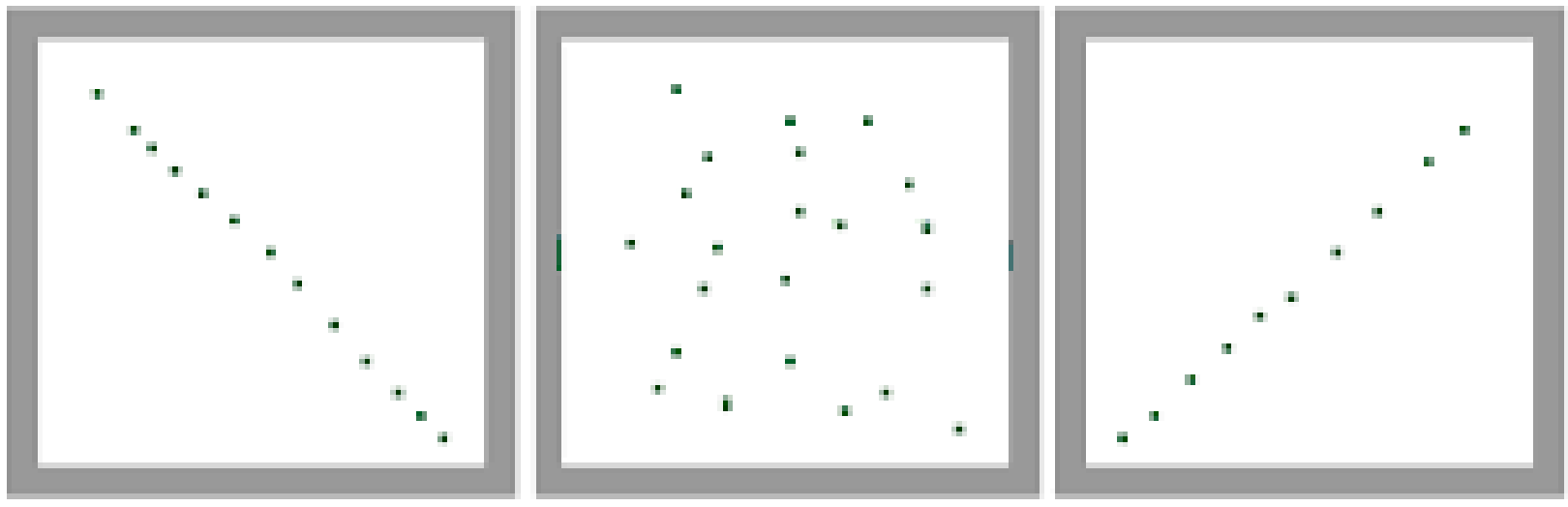
Analytics Vidhya

Have you ever wondered about the subtle differences between covariance and correlation in statistical analysis?

These two variables are often used interchangeably, but their nuances can greatly affect our interpretation and utilization of statistical relationships.

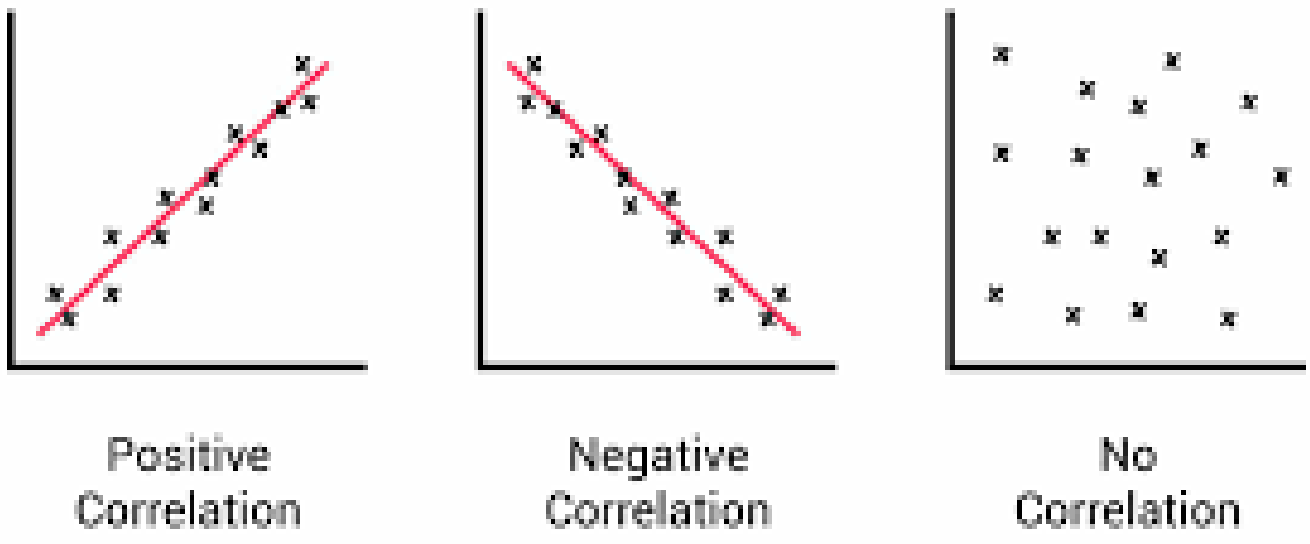
What is covariance?

Covariance measures the relationship between two variables, indicating how changes in one variable are associated with changes in another. It can be positive, negative, or zero, representing the nature of the relationship.



What is correlation?

Correlation quantifies the strength and direction of the linear relationship between two variables. The correlation coefficient ranges from -1 to 1, where 1 signifies a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 represents no linear relationship.



Analytics Vidhya

Advantages and Disadvantages of Covariance

Advantages	Disadvantages
Easy to Calculate	Restricted to Linear Relationships
Apprehends Relationship	Doesn’t Offer Relationship Magnitude
Beneficial in Portfolio Analysis	Scale Dependency

Analytics Vidhya

Advantages and Disadvantages of Correlation

Advantages	Disadvantages
Determining Non-Linear Relationships	Driven by Extreme Values
Standardized Criterion	Data Requirements
Robustness to Outliers	Limited to Bivariate Analysis
Scale Independencies	

Differences between Covariance and Correlation

Difference Grounds	Covariance	Correlation
Meaning	Covariance means two variables directly or inversely depend on one another.	Two variables are said to be in correlation if the change in one affects the other variable.
Values	Lie between -infinity to +infinity	Values lie between -1 to 1
Unit	It's a product of the unit of variables	It's a unit-free measure
Change in Scale	Even minor changes in scale affect Covariance	There won't be any change in correlation because of the scale
Measure of	Correlation	The scaled version of Covariance
Application	Market Research, Portfolio Analysis, and Risk Assistance	Medical Research, Data Analysis, and Forecasting

Analytics Vidhya

Practical Scenarios Where Covariance is Useful

- **Market Research:** Used in market research to understand how marketing endeavors influence business outcomes, such as sales revenue and advertising expenditure.

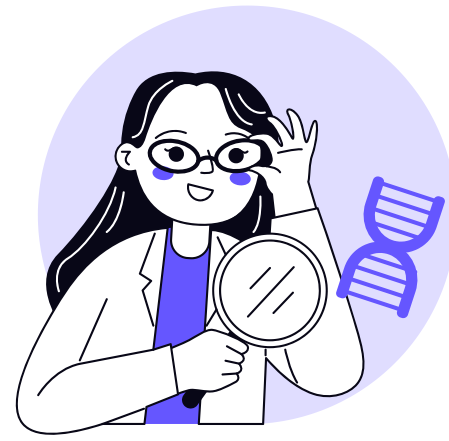
- **Risk Assessment:** It is used to assess potential losses and set appropriate premiums (such as claims frequency, health conditions, and age).

- **Portfolio Analysis:** Covariance has a profound application in finance for evaluating the relationship between different asset returns within a portfolio.

Analytics Vidhya

Practical Scenarios Where Correlation is Useful

- **Forecasting:** Used in determining the degree to which they can predict one variable based on the values of another variable.



- **Analysis of Data:** Data enthusiasts use correlation popularly to quantify and identify relationships between variables.

- **Medical Research:** Correlations help find associations between variables, such as lung cancer and smoking or cardiovascular disease risk and BMI (body mass index).

The formula for calculating the covarience

$$\sigma(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

Where,

x is the mean (average) of the X-variable

y is the mean (average) of the Y-variable

xi is the value of the X-variable

yi is the value of the Y-variable

n is the number of data points

The formula for calculating the correlation coefficient

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Where,

- **R denotes the correlation coefficient**
- **x denotes the mean (average) of values of the X variable**
- **xi is the value of variable X in the data sample**
- **y denotes the mean (average) of values of the Y variable**
- **yi is the value of variable y in the data sample**