Covariance and correlation

Covariance and **correlation** are two statistical measures that quantify the relationship between two variables. Covariance measures the extent to which two variables change together, while correlation measures the strength and direction of the linear relationship between two variables.

Covariance

Covariance is a measure of how two variables vary together. It is calculated by taking the product of the deviations from the mean for each variable and then averaging the products. The sign of the covariance tells you whether the variables are positively or negatively correlated, and the magnitude of the covariance tells you how strong the correlation is.

A positive covariance indicates that the variables tend to move in the same direction. For example, the price of a stock and the volume of trading in that stock often have a positive covariance. As the price of the stock goes up, the volume of trading also tends to go up.

A negative covariance indicates that the variables tend to move in opposite directions. For example, the price of a stock and the interest rate on a loan that is secured by that stock often have a negative covariance. As the price of the stock goes up, the interest rate on the loan tends to go down.

```python
In [6]:  #function for calculating covrience
         import statistics as st
         x = [2,4,6]
         y = [3,5,7]
         def covarience(x,y):
             cov = [((X-st.mean(x))*(Y-st.mean(y)))/(len(x)-1) for X,Y in zip(x,y) ]
             return f"Covarience of x and y is: {sum(cov)}"
         covarience(x,y)
```

```
Out[6]:  'Covarience of x and y is: 4.0'
```

```python
In [8]:  #
         import numpy as np
         X = np.stack((x, y), axis=0)
         np.cov(X)
```

```
Out[8]:  array([[4., 4.],
                [4., 4.]])
```

Corelation

Correlation is a measure of the strength and direction of the linear relationship between two variables. It is calculated by dividing the covariance by the product of the standard deviations for each variable. The correlation coefficient can range from -1 to +1.

A correlation coefficient of +1 indicates a perfect positive linear relationship. This means that the variables move in the same direction and the changes in one variable are perfectly proportional to the changes in the other variable.

A correlation coefficient of -1 indicates a perfect negative linear relationship. This means that the variables move in opposite directions and the changes in one variable are perfectly inversely proportional to the changes in the other variable.

A correlation coefficient of 0 indicates that there is no linear relationship between the variables. This does not mean that there is no relationship between the variables, only that the relationship is not linear.

Pearson Corelation coeficient

Pearson correlation is a statistical measure that assesses the strength of a linear relationship between two variables. It is a measure of how much two variables change together, and it can be used to predict one variable from the other.

The Pearson correlation coefficient is a number between -1 and 1. A correlation coefficient of 1 indicates a perfect positive correlation, meaning that the two variables increase or decrease together. A correlation coefficient of -1 indicates a perfect negative correlation, meaning that the two variables increase or decrease in opposite directions. A correlation coefficient of 0 indicates no correlation, meaning that there is no relationship between the two variables.

The Pearson correlation coefficient is calculated using the following formula:

$r = (\sum(x-\bar{x})(y-\bar{y})) / sqrt(\sum(x-\bar{x})^2) * sqrt(\sum(y-\bar{y})^2)$

In [12]:
```python
#implementation of pearson corelation coefficient
import statistics as st
x = [2,4,6]
y = [3,5,7]
def PearsonCorelationCoeff(x,y):
    PCC = [((X-st.mean(x))*(Y-st.mean(y)))/(st.stdev(x)*st.stdev(y)) for X,Y i
    return sum(PCC)
PearsonCorelationCoeff(x,y)
```
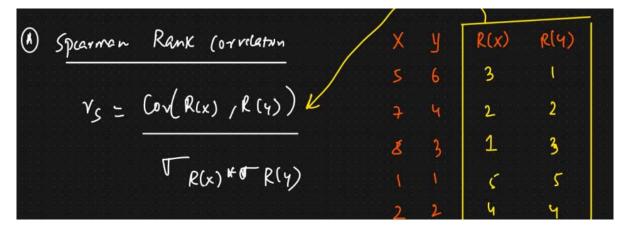
Out[12]: 2.0

In [19]:
```python
#
import numpy as np
import scipy
res= scipy.stats.pearsonr(x,y)
res.confidence_interval()
```

Out[19]: ConfidenceInterval(low=-1.0, high=1.0)

Spearman corelation coefficient

pearman rank correlation is a statistical measure that assesses the strength and direction of the monotonic relationship between two variables. A monotonic relationship is one in which the variables tend to increase or decrease together, but not necessarily in a linear fashion

formula for calculating the Spearman rank correlation coefficient:

In [21]:
```
#implementation

res = scipy.stats.spearmanr(x,y)
res
```

Out[21]: SignificanceResult(statistic=1.0, pvalue=0.0)