

## Введение Ковариация и корреляция

Ковариация и корреляция являются математическими понятиями, которые также используются в статистике и теории вероятностей. Наиболее полезен для понимания переменных. Как правило, область науки о данных используется для сравнения выборок данных из разных групп населения, а ковариация используется для определения того, насколько две случайные переменные связаны друг с другом, тогда как корреляция используется для определения изменения одной переменной, если она влияет на другую переменную.

Оба термина связаны с линейной зависимостью между переменными. Другими словами, если одна переменная идет в возрастающем направлении так же, как другая переменная идет в том же направлении, это означает положительную корреляцию. Если обе переменные имеют противоположное направление, то это называется отрицательной корреляцией.

Когда нет отношений, нет никаких изменений. Корреляция объясняет, насколько изменение одной переменной приводит к изменению второй переменной.

## Образец

Выборка случайным образом выбирается из населения. Мы рассчитываем ковариацию и корреляцию на выборках, а не на всей совокупности.

## Ковариация

Ковариация зависит только от знака. Положительное значение показывает обе переменные в одном направлении. То же, что и Отрицательное значение показывает, что оба находятся в противоположном направлении. Ковариация - это измеренное использование для определения того, насколько переменная изменяется случайным образом. Ковариация является произведением единиц двух переменных. Значение ковариации лежит между  $-\infty$  и  $+\infty$ . Ковариация двух переменных ( $x$  и  $y$ ) может быть представлена как  $\text{cov}(x,y)$ .  $E[x]$  является ожидаемым значением или также называется средним значением выборки « $x$ ».



## Covariance Formula

For Population

$$\text{Cov}(x,y) = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{N}$$

For Sample

$$\text{Cov}(x,y) = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{(N - 1)}$$

Where,

- $\bar{x}$  = sample mean of x
- $\bar{y}$  = sample mean of y
- $x_i$  and  $y_i$  = the values of x and y for ith record in the sample.
- N = is the no of records in the sample

Значение формулы

1. Числитель показывает количество дисперсии x, умноженное на количество дисперсии y.
2. Единица ковариации показывает, единица x умножается на единицу y.
3. Следовательно, если мы изменим единицу измерения переменных, ковариация также будет иметь новое значение, но знак останется прежним.
4. Однако, если она положительна, обе переменные изменяются в одном направлении, если же она отрицательна, то они изменяются в противоположном направлении.

## Корреляции

Корреляция означает корреляцию между двумя переменными, которая является нормализованной версией ковариации. Диапазон коэффициентов корреляции всегда находится в диапазоне от -1 до 1. Коэффициент корреляции также известен как коэффициент корреляции Пирсона. Когда вы читаете о ковариации, это говорит только о направлении, но этого недостаточно, чтобы полностью понять взаимосвязь. Итак, мы делим ковариацию на стандартное отклонение x и y соответственно.

Коэффициент корреляции между случайными величинами X и Y, вы должны разделить выборочную ковариацию X и Y на произведение выборочного насыщения X и Y соответственно.

## Значение

1. -1 и +1 указывают на то, что обе переменные имеют идеальную линейную зависимость.
2. Отрицательный означает, что они обратно пропорциональны друг другу с коэффициентом значения коэффициента корреляции.
3. Положительные означают, что они прямо пропорциональны друг другу, средние изменяются в одном направлении с коэффициентом значения коэффициента корреляции.
4. если коэффициент корреляции равен 0, то это означает отсутствие линейной зависимости между переменными.

$$\text{Correlation} = \frac{\text{Cov}(x, y)}{\sigma_x * \sigma_y}$$

Where,

- Correlation = sample correlation between X and Y
- Cov(X,Y) = sample covariance between X and Y

$\sigma_x$

- = sample standard deviation of X

•

$\sigma_y$

- = sample standard deviation of Y

## Разница между ковариацией и корреляцией

Корреляция — это просто нормализованная форма ковариации. Очевидно, важно быть точным в формулировках при обсуждении этих двух, но концептуально они почти идентичны.

Значение коэффициента корреляции находится в диапазоне [-1 - 1]. -1 указывает на отрицательную связь. 1 означает положительное отношение. 0 означает отсутствие отношений.

```

import os
import sys
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

X = np.random.rand(50)    #random number
Y = 2 * X + np.random.normal(0, 0.1, 50).....

#covariance
cov_matrix = np.cov(X, Y)    #calculate covariance between x & y
print('The Covariance of X and Y: %.2f'%cov_matrix[0, 1]) .....

The covariance of X and Y: 0.21

#Correlation
.....
cor_matrix = np.corrcoef(X, Y) #calculate correlation between x & y
.....
print(Correlation of X and Y: %.2f'%cor_matrix[0, 1])

Correlation of X and Y: 0.99

```

Чтобы понять, как выглядят коррелированные данные, давайте построим два коррелированных набора данных.