

# Kovarianz



**Kovarianz** ist ein statistisches Maß, das die Richtung der linearen Beziehung zwischen zwei Variablen beschreibt. Es gibt an, ob und wie zwei Variablen gemeinsam variieren – also ob ein Anstieg in einer Variablen mit einem Anstieg oder einem Abstieg in der anderen Variablen verbunden ist.

## Formel für die Kovarianz:

Für zwei Variablen  $X$  und  $Y$  mit jeweils  $n$  Datenpunkten lautet die Formel der Kovarianz:

$$\text{Kov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

- $X_i$  und  $Y_i$  sind die einzelnen Datenpunkte der beiden Variablen.
- $\bar{X}$  und  $\bar{Y}$  sind die Mittelwerte der beiden Variablen.
- $n$  ist die Anzahl der Datenpunkte.

## Interpretation:

- **Positive Kovarianz:** Wenn die Kovarianz positiv ist, bedeutet das, dass, wenn eine Variable größer wird, die andere Variable tendenziell auch größer wird. Es gibt eine **positive lineare Beziehung**.
  - Beispiel: Wenn die Anzahl der Studienstunden steigt, steigt möglicherweise auch die Prüfungsnote – die Kovarianz wäre positiv.
- **Negative Kovarianz:** Eine negative Kovarianz bedeutet, dass, wenn eine Variable größer wird, die andere Variable tendenziell kleiner wird. Es gibt eine **negative lineare Beziehung**.
  - Beispiel: Wenn die Außentemperatur steigt, sinkt möglicherweise der Energieverbrauch für Heizung – die Kovarianz wäre negativ.
- **Kovarianz nahe Null:** Wenn die Kovarianz nahe bei null liegt, bedeutet das, dass es keine starke lineare Beziehung zwischen den beiden Variablen gibt. Sie sind also weitgehend unabhängig voneinander.

## Einschränkungen:

- **Skalierung:** Kovarianz ist abhängig von den Einheiten der Variablen. Wenn die Daten in anderen Einheiten gemessen werden, kann sich die Kovarianz stark

ändern. Daher wird häufig der **Korrelationskoeffizient** verwendet, siehe unten.

## Beispiel:

Nehmen wir an, wir haben die folgenden Datenpunkte:

X	Y
1	2
2	3
3	6
4	8
5	10

1. Berechne die Mittelwerte:  $\bar{X} = 3$ ,  $\bar{Y} = 5.8$ .
2. Berechne  $(X_i - \bar{X})$  und  $(Y_i - \bar{Y})$  für jeden Datenpunkt und multipliziere diese Abweichungen.
3. Die Kovarianz ist der Durchschnitt dieser Produkte.

Hier ist die Kovarianz positiv, da ein Anstieg von X auch mit einem Anstieg von Y einhergeht.

## Korrelation

Die **Korrelation** ist ein statistisches Maß, das die **Stärke und Richtung** einer linearen Beziehung zwischen zwei Variablen beschreibt. Sie baut auf der Kovarianz auf, ist jedoch **standardisiert**, sodass sie unabhängig von den Einheiten der Variablen ist und immer zwischen **-1** und **+1** liegt.

## Korrelationskoeffizient (Pearson-Korrelation):

Die gebräuchlichste Methode zur Berechnung der Korrelation ist der **Pearson-Korrelationskoeffizient**  $r$ , der durch folgende Formel definiert ist:

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

- $\text{Kov}(X, Y)$  ist die Kovarianz von  $X$  und  $Y$ .
- $\sigma_X$  und  $\sigma_Y$  sind die Standardabweichungen der Variablen  $X$  und  $Y$ .
- Es gilt:  $|\text{Kov}(X, Y)| \leq \sigma_X \sigma_Y$ , deswegen bewegt sich der Wert des Koeffizienten zwischen -1 und 1

## Eigenschaften der Korrelation:

- **$r=+1$ : Perfekte positive Korrelation.** Wenn eine Variable steigt, steigt auch die andere im gleichen Verhältnis.
- **$r=-1$ : Perfekte negative Korrelation.** Wenn eine Variable steigt, sinkt die andere im gleichen Verhältnis.
- **$r=0$ : Keine Korrelation.** Es gibt keine lineare Beziehung zwischen den beiden Variablen.

## Beispiel:

Angenommen, wir haben die Daten von oben für  $X$  und  $Y$ :

### 1. Mittelwerte:

- $\bar{X} = 3$
- $\bar{Y} = 5.8$

### 2. Standardabweichungen:

- $\sigma_X$  und  $\sigma_Y$  werden basierend auf den Abweichungen vom Mittelwert berechnet.

3. **Korrelation:** Wenn die Korrelation berechnet wird (mithilfe der Kovarianz und den Standardabweichungen), wird ein Wert nahe **1** herauskommen, da  $X$  und  $Y$  hier positiv und nahezu proportional zusammenhängen.

## Unterschied zur Kovarianz:

- **Kovarianz:** Beschreibt nur die Richtung der Beziehung, aber die Werte können in beliebigen Einheiten variieren.

- **Korrelation:** Standardisiert die Kovarianz, sodass sie immer zwischen  $-1$  und  $+1$  liegt. Sie gibt nicht nur die Richtung, sondern auch die **Stärke** der Beziehung an.

## Interpretation der Korrelation:

- **Starke positive Korrelation** ( $r$  nahe  $1$ ): Wenn  $X$  steigt, steigt auch  $Y$  stark.
- **Schwache positive Korrelation** ( $r$  nahe  $0$ , aber positiv): Es gibt einen schwachen Zusammenhang zwischen  $X$  und  $Y$ , aber tendenziell steigen sie gemeinsam.
- **Starke negative Korrelation** ( $r$  nahe  $-1$ ): Wenn  $X$  steigt, sinkt  $Y$  stark.
- **Schwache negative Korrelation** ( $r$  nahe  $0$ , aber negativ): Es gibt einen schwachen Zusammenhang zwischen  $X$  und  $Y$ , aber tendenziell verhalten sie sich entgegengesetzt.

## Unterschied zwischen Korrelation und kausalem Zusammenhang

Eine **Korrelation** beschreibt die Stärke und Richtung des Zusammenhangs zwischen zwei Variablen.

Korrelation zeigt nur, dass die Variablen in einer Beziehung zueinander stehen, sagt aber nichts darüber aus, **warum** dieser Zusammenhang besteht.

**Kausalität** zeigt eine Ursache-Wirkungs-Beziehung, wobei eine Variable direkt die andere beeinflusst.

Ein **kausaler Zusammenhang** bedeutet, dass die eine Variable eine **Ursache** für Veränderungen in der anderen Variable ist. Das bedeutet, dass Veränderungen in der einen Variablen direkt zu Veränderungen in der anderen führen. Ein Beispiel wäre, dass eine erhöhte Sonneneinstrahlung (Ursache) zu einer höheren Temperatur (Wirkung) führt.

**Korrelation ist nicht gleich Kausalität.** Nur weil zwei Variablen korrelieren, heißt das nicht, dass die eine die andere verursacht. Es könnte eine **dritte Variable** existieren, die beide beeinflusst, oder der Zusammenhang könnte **rein zufällig** sein.

## Beispiel für Korrelation ohne Kausalität

Ein klassisches Beispiel für Korrelation ohne Kausalität ist der Zusammenhang zwischen der Anzahl von Eisverkäufen und der Anzahl von Badeunfällen: Beide steigen im Sommer an, weil die Temperaturen höher sind, aber das eine verursacht nicht das andere.

