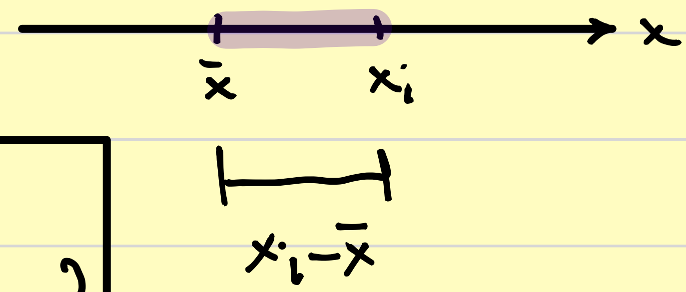
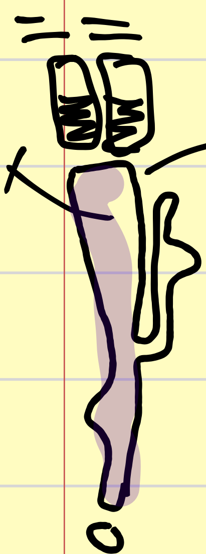
$$m_2 = VAR(x) = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}$$



1. VARIABEL

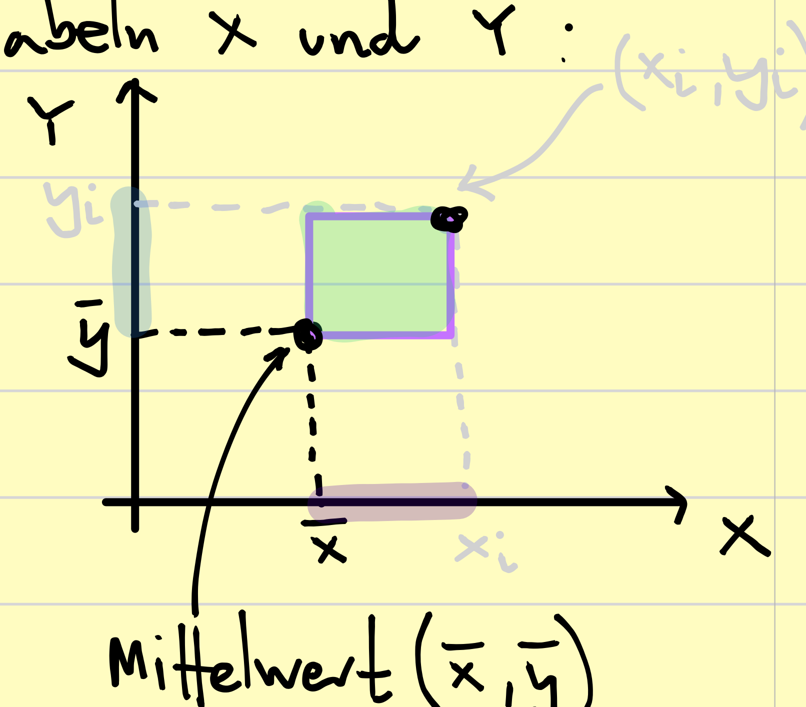


Was machen wir, wenn wir mehr als eine Variabel haben?

KOVARIANZMATRIX

Definition KOVARIANZ von ZWEI Variabeln X und Y :

$$\boxed{kov(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{N-1}}$$



Geometrische Interpretation:

- Die Varianz ist die summe der Abstände zum Mittelwert in Quadrat (immer positiv) geteilt durch N.
- Die Kovarianz ist die Summe der Viereckflächen zum Mittelwert ohne Quadrat (kann positiv/negativ sein) geteilt durch N-1.

Die Kovarianz hat einen Symbol.

- Interpretation der Kovarianz:
  - $kov(X,Y) \simeq 0$ → Die Meisten Punkte sind um den Mittelwert verteilt
  - $kov(X,Y) \gg 0$, $\ll 0$ → Die Meisten Punkte sind vom Mittelwert entfernt.

Beispiel. Gegeben wird ein Kennzahlensystem mit zwei KPIs (Key Performance Indicators). Kennzahlen: DURCHLAUFZEIT und €/Stück. Bitte ermitteln Sie die Kovarianz beider Kennzahlen.

	DLZ (Tage)	€/Stück
KW1	6'3	230
KW2	4'7	180
KW3	3'2	170
KW4	3'8	175

$$\overline{DLZ} = \frac{\sum DLZ_i}{4} = \frac{6'3+4'7+3'2+3'8}{4} = 4'5$$

$$\overline{\frac{€}{stk}} = \frac{\sum \frac{€}{stk}_i}{4} = \frac{230+180+170+175}{4} = 188'75$$

$$kov\left(DLZ, \frac{€}{St.}\right) = \frac{\sum_{i=1}^{4}\left(DLZ_i - \overline{DLZ}\right)\cdot\left(\frac{€}{stk}_i - \overline{\frac{€}{stk}}\right)}{4-1} =$$

$$= \frac{(6'3-4'5)\cdot(230-188'75)+(4'7-4'5)(180-188'75)+ \\ +(3'2-4'5)(170-188'75)+(3'8-4'5)(175-188'75)}{4-1} = \ldots =$$

- Die $\boxed{kov(X,Y) = kov(Y,X)}$

$$kov(X,Y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{N-1} = \frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{N-1} = kov(Y,X)$$

- KOVARIANZMATRIX

von 3 VARIABELN (erweiterbar auf N)

$$\boxed{\text{KOV.MATRIX}_{(X,Y,Z)} = A = \begin{array}{c|ccc} & X & Y & Z \\ \hline X & VAR(X) & KOV(X,Y) & KOV(X,Z) \\ Y & KOV(Y,X) & VAR(Y) & KOV(Y,Z) \\ Z & KOV(Z,X) & KOV(Z,Y) & VAR(Z) \end{array}}$$

· Die Kovarianzmatrix ist symmetrisch.

$$\boxed{A = \begin{bmatrix} VAR(X) & KOV(X,Y) & KOV(X,Z) \\ KOV(X,Y) & VAR(Y) & KOV(Y,Z) \\ KOV(X,Z) & KOV(Y,Z) & VAR(Z) \end{bmatrix}}$$

$KOV(X,Y) = KOV(Y,X)$

$KOV(X,Z) = KOV(Z,X)$

$KOV(Y,Z) = KOV(Z,Y)$

Beispiel. Gegeben wird ein Kennzahlensystem mit 3 KPIs: DURCHLAUFZEIT, $\frac{€}{\text{Stück}}$, Qualität. Bitte die Kovarianzmatrix vom System ermitteln.

	X DLZ (Tage)	Y €/Stück	Z Qualität (ppm)
KW1	6'3	320	3200
KW2	4'7	180	4700
KW3	3'2	170	2100
KW4	3'8	179	1500

$\bar{x} = 4'5 \qquad \bar{y} = 188'75 \qquad \bar{z} = 2875$

$$A = KOV(X,Y,Z) = \begin{bmatrix} VAR(X) & KOV(X,Y) & KOV(X,Z) \\ KOV(X,Y) & VAR(Y) & KOV(Y,Z) \\ KOV(X,Z) & KOV(Y,Z) & VAR(Z) \end{bmatrix}$$

$$VAR(X) = \frac{\sum (x_i - \bar{x})^2}{N} = \frac{(6'3-4'5)^2 + (4'7-4'5)^2 + (3'2-4'5)^2 + (3'8-4'5)^2}{4} = 1'365$$

$$VAR(Y) = \frac{\sum (y_i - \bar{y})^2}{N} = \frac{(320-188'75)^2 + (180-188'75)^2 + (170-188'75)^2 + (175-188'75)^2}{4} = 4460'9$$

$$VAR(Z) = \frac{\sum (z_i-\bar{z})^2}{N} = \frac{(3200-2875)^2+(4700-2875)^2+(2100-2875)^2+(1500-2875)^2}{4} = 1'48 \cdot 10^6$$

$$KOV(X,Y) = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{N-1} = \text{(siehe oben)}$$

$$KOV(X,Z) = \frac{\sum (x_i-\bar{x})(z_i-\bar{z})}{N-1} = \frac{(6'3-4'5)(3200-2875)+(4'7-4'5)(4700-2875)+(3'2-4'5)(2100-2875)+(3'8-4'5)(1500-2875)}{4-1}$$

$$KOV(Y,Z) = \frac{\sum (y_i-\bar{y})(z_i-\bar{z})}{N-1} = \frac{(320-188'75)(3200-2875)+(180-188'75)(4700-2875)+(170-188'75)(2100-2875)+(175-188'75)(1500-2875)}{4-1}$$

- In Management Systeme haben Kennzahlen in der Regel unterschiedliche Skala's. In dem Beispiel oben die Werte der DLZ sind alle zw. $[0,10]$, die Werte der $\frac{€}{Stk}$ sind alle zw $[150,400]$, und die Werte der Qualität sind alle zw. $[1500,3500]$. Diese Natur der Kennzahlen macht die Kovarianzrechnung UNBRAUCHBAR weil die Abstände in der X-Dimension eine andere Bedeutung als die Abstände in der Y-Dimension haben.
- Um diese Schwierigkeit zu lösen sind wir gezwungen vor der KOV.MATRIX Rechnung zu normieren. Wenn wir normieren, substrahieren wir den Mittelwert und teilen wir durch die std. Abweichung.

$$x = \{x_1, x_2, x_3, x_4, \dots\} \longrightarrow \quad \bar{x} = \frac{\sum x_i}{N} \quad \sigma_x = \sqrt{\frac{\sum (x_i-\bar{x})^2}{N}}$$

$\downarrow$ NORMIERUNG

$$x^* = \left\{ \frac{x_1-\bar{x}}{\sigma_x}, \frac{x_2-\bar{x}}{\sigma_x}, \frac{x_3-\bar{x}}{\sigma_x}, \dots \right\}$$

$$\downarrow$$
$$x^* = \{x_1^*, x_2^*, x_3^*, x_4^*, \dots\}$$

Durch die Normierung schaffen wir Homogeneität zw. den Variabeln und machen wir sie dadurch vergleichbar.

Erst nach der Normierung macht es Sinn die Kovarianz Matrix zu berechnen.

Beispiel. Gegeben wird ein Kennzahlensystem mit 3 KPIs [X, Y, Z]. Bitte ermitteln Sie die KOV. MATRIX.

X	Y	Z
...	...	...
...	...	...
...	...	...

1 → NORMIERUNG →

X^*	Y^*	Z^*
...	...	...
...	...	...
...	...	...

2 → KOV. MATRIX → A^*

□ Die resultierende KOV. MATRIX der normierten Variabeln A^* hat eine diagonale von 1, weil die normierten Variabeln X^*, Y^*, Z^* haben alle eine Varianz von 1.

□ Beispiel. Bitte ermitteln Sie die Kov. Matrix vom folgenden 3 KPIs Kennzahlensystem in einem Management Kontext.

	X DLZ (Tage)	Y €/Stück	Z Qualität (ppm)
KW1	6'3	320	3200
KW2	4'7	180	4700
KW3	3'2	170	2100
KW4	3'8	179	1500
	$\bar{x} = 4'5$	$\bar{y} = 188'75$	$\bar{z} = 2875$

VAR X = 1'365
VAR Y = 4460'9
VAR Z = 1481875

Schritt 1. Normieren.

	x^*	y^*	z^*
KW1	$x_1^* = \frac{6'3-4'5}{\sqrt{1'365}} = 1'541$	$y_1^* = \frac{320-188'75}{66'79} = 1'965$	$z_1^* = \frac{3200-2875}{1217} = 0'267$
KW2	$x_2^* = \frac{4'7-4'5}{\sqrt{1'365}} = 0'171$	$y_2^* = \frac{180-188'75}{66'79} = -0'131$	$z_2^* = \frac{4700-2875}{1217} = 1'5$
KW3	$x_3^* = \frac{3'2-4'5}{\sqrt{1'365}} = -1'113$	$y_3^* = \frac{170-188'75}{66'79} = -0'281$	$z_3^* = \frac{2100-2875}{1217} = -0'637$
KW4	$x_4^* = \frac{3'8-4'5}{\sqrt{1'365}} = -0'6$	$y_4^* = \frac{175-188'75}{66'79} = -0'206$	$z_4^* = \frac{1500-2875}{1217} = -1'13$

$\bar{x}^* = 0$ $\qquad$ $\bar{y}^* = 0$ $\qquad$ $\bar{z}^* = 0$

$\sigma_{x^*} = 1$ $\qquad$ $\sigma_{y^*} = 1$ $\qquad$ $\sigma_z^* = 1$

$$A^* = \begin{bmatrix} VAR(x^*) & KOV(x^*, y^*) & KOV(x^*, z^*) \\ & VAR(y^*) & KOV(y^*, z^*) \\ & & VAR(z^*) \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

$$VAR(x^*) = \sigma_{x^*}^2 = 1^2 = 1 = VAR(y^*) = VAR(z^*)$$

$$KOV(x^*, y^*) = \frac{\sum (x_i^* - \bar{x}^*)(y_i^* - \bar{y}^*)}{N-1} = \frac{\sum x_i^* y_i^*}{N-1} = \frac{1'541 \cdot 1'965 + 0'171 \cdot (-0'131) + (-1'113) \cdot (-0'281) + (-0'6) \cdot (-0'206)}{3}$$

$$\mathrm{Kov}(X^*, Z^*) = \frac{\sum (x_i^* - \cancel{\bar{x}^*})(z_i^* - \cancel{\bar{z}^*})}{N-1} = \frac{\sum x_i^* z_i^*}{N-1} = \frac{1'541 \cdot 0'267 + 0'171 \cdot 1'5 + (-1'113)\cdot(-0'637) + (-0'6)(-1'13)}{3}$$

$$\mathrm{Kov}(Y^*, Z^*) = \frac{\sum (y_i^* - \bar{y}^*)(z_i^* - \bar{z}^*)}{N-1} = \frac{1'965 \cdot 0'267 + (-0'131) \cdot 1'5 + (-0'281)\cdot(-0'637) + (-0'206)\cdot(-1'13)}{3}$$

KORRELATION. Produkt-Moment von Pearson

$\rho \equiv$ Rho $\equiv$ KORRELATIONSKOEFFIZIENT von PEARSON $\in [-1, 1]$

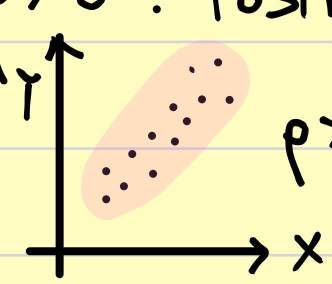
Wird definiert als:

$$\boxed{\rho_{XY} = \rho(X, Y) = \frac{\mathrm{Kov}(X, Y)}{\sigma_X \cdot \sigma_Y}}$$

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

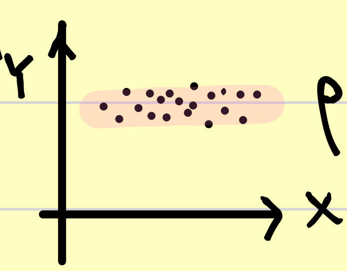
Interpretation

- $\rho > 0$. POSITIVE KORRELATION. Die Variabeln verhalten sich so, dass wenn eine wächst, die andere auch und umgekehrt.



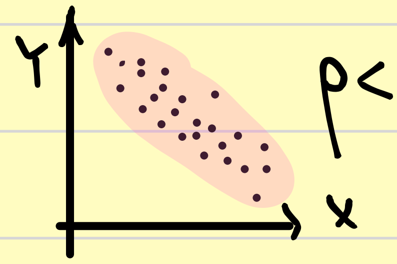
$\rho > 0$

- $\rho \cong 0$. KEINE KORRELATION. Die Variabeln verhalten sich so, dass wenn eine sich ändert, die andere sich nicht (ändert).



$\rho \cong 0$

II $\rho < 0$. NEGATIVE KORRELATION.



Die Variabeln verhalten sich so, dass wenn eine wächst, die andere schrumpft, und umgekehrt.

Beispiel. Bitte ermitteln Sie um wie viel sich die Qualität eines Produktes verbessert, wenn sich die DLZ um 15% verbessern würde.

	DLZ (Tage) X	Qualität (ppm) Y
KW1	7'3	3200
KW2	6'7	2700
KW3	5'8	1900
KW4	5'6	1700

$\bar{x} = 6{,}35 \qquad \bar{y} = 2375$

$$\rho_{XY} = \frac{\mathrm{KOV}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - x)^2}\sqrt{\sum (y_i - \bar{y})^2}} = \frac{(7{,}3-6{,}35)(3200-2375) + (6{,}7-6{,}35)(2700-2375) + (5{,}8-6{,}35)(1900-2375) + (5{,}6-6{,}35)(1700-2375)}{\sqrt{(7{,}3-6{,}35)^2 + (6{,}7-6{,}35)^2 + (5{,}8-6{,}35)^2 + (5{,}6-6{,}35)^2}\sqrt{(3200-2375)^2 + (2700-2375)^2 + (1900-2375)^2 + (1700-2375)^2}}$$

15% Senkung DLZ $\longrightarrow$? Senkung ppm

$-0{,}15 \cdot \rho_{XY} \longrightarrow - \underline{\qquad} \cdot 100\%$ ppm

KORRELATIONSMATRIX

I Definition für 3 Variabeln:

$$\boxed{KORR(X,Y,Z) = \begin{bmatrix} 1 & \rho_{XY} & \rho_{XZ} \\ \rho_{YX} & 1 & \rho_{YZ} \\ \rho_{ZX} & \rho_{ZY} & 1 \end{bmatrix}}$$

II $\boxed{\rho_{XY} = \rho_{YX}}$

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

$$\boxed{KORR(X,Y,Z) = \begin{bmatrix} 1 & \rho_{XY} & \rho_{XZ} \\ \rho_{XY} & 1 & \rho_{YZ} \\ \rho_{XZ} & \rho_{YZ} & 1 \end{bmatrix}}$$

Übung: bitte ermitteln Sie die Korrelationsmatrix des Kennzahlensystems mit 3 KPIs (DLZ, €/St., Qualität) oben.