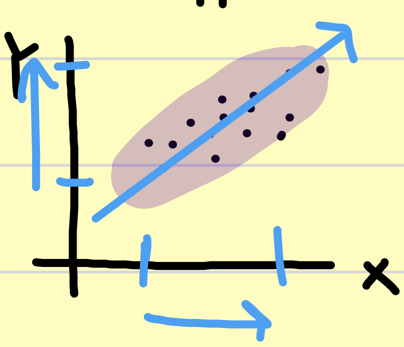
Korrelation (Pearson's Korrelation)

$$\boxed{\rho = \text{Rho} = \rho(X,Y) = \frac{\operatorname{kov}(X,Y)}{\sigma_X \sigma_Y}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

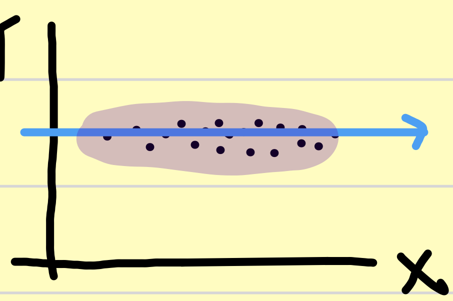
$\rho \in [-1,1]$

Interpretation:

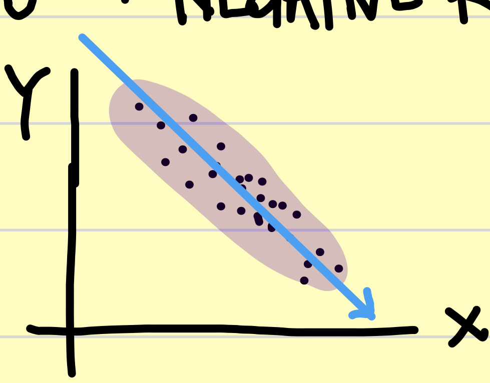
- $\rho > 0$: POSITIVE KORRELATION : wenn eine Variable wächst, die andere wächst auch.



- $\rho \approx 0$: KEINE KORRELATION : wenn eine Variable sich ändert, ändert sich die andere nicht.



- $\rho < 0$: NEGATIVE KORRELATION : wenn eine Variable wächst, die andere schrumpft.



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

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

1.SCHRITT : Normieren.

$$\bar{x} = \frac{7'3+6'7+5'8+5'6}{4} = 6'35 \; ; \; \bar{Y} = \frac{3200+2700+1900+1700}{4} = 2375$$

$$\sigma_x = \sqrt{\frac{(7'3-6'35)^2+(6'7-6'35)^2+(5'8-6'35)^2+(5'6-6'35)^2}{4-1}} = \dots$$

$$\sigma_y = \sqrt{\frac{(3200-2375)^2+(2700-2375)^2+(1900-2375)^2+(1700-2375)^2}{4-1}} = \dots$$

	DLZ^*	Q^*
KW1	$x_1^* = \frac{7'3-6'35}{\sigma_x}$	$\frac{3200-2375}{\sigma_y} = y_1^*$
KW2	$x_2^* = \frac{6'7-6'35}{\sigma_x}$	$\frac{2700-2375}{\sigma_y} = y_2^*$
KW3	$x_3^* = \frac{5'8-6'35}{\sigma_x}$	$\frac{1900-2375}{\sigma_y} = y_3^*$
KW4	$x_4^* = \frac{5'6-6'35}{\sigma_x}$	$\frac{1700-2375}{\sigma_y} = y_4^*$

$\bar{x}^* = 1 \quad \sigma_x^* = 0$

$\bar{y}^* = 1 \quad \sigma_y^* = 0$

$$\mathrm{KORR}(X^*,Y^*) = \frac{\sum (x_i^* - \bar{x}^*)(y_i^* - \bar{y}^*)}{\sqrt{\sum (x_i^* - \bar{x}^*)^2}\sqrt{\sum (y_i^* - \bar{y}^*)^2}} = \frac{\sum x_i^* - y_i^*}{\sum x_i^* \cdot \sum y_i^*} = \rho^*$$

$$-0{,}15 \cdot \rho^* \cdot 100 \rightarrow \underline{\qquad\qquad} \%$$

Änderung der Q in %

KORRELATIONSMATRIX

· Definition:

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

· Korrelation is bijektiv:

$$\rho_{XY} = \rho_{YX} \qquad \rho_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sigma_y \sigma_x} = \rho_{YX}$$

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

Übung: bitte ermitteln Sie die Korrelationsmatrix des Kennzahlensystems mit 3 KPIs (DLZ, Kosten, Q). Interpretieren Sie die Ergebnisse aus statistischer Sicht.

	x DLZ (Tage)	y Kosten (€/St)	z Qualität (ppm)
KW1	6'3	320	3200
KW2	4'7	180	4700
KW3	3'2	170	2100
KW4	3'8	179	1500

1. NORMIEREN:

$$\bar{x} = \frac{6'3+4'7+3'2+3'8}{4} ; \quad \bar{y} = \frac{320+180+170+179}{4} ; \quad \bar{z} = \frac{3200+4700+2100+1500}{4}$$

$$\sigma_x = \sqrt{\frac{(6'3-\bar{x})^2+(4'7-\bar{x})^2+(3'2-\bar{x})^2+(3'8-\bar{x})^2}{4-1}}$$

$$\sigma_y = \sqrt{\frac{(320-\bar{y})^2+(180-\bar{y})^2+(170-\bar{y})^2+(179-\bar{y})^2}{4-1}}$$

$$\sigma_z = \sqrt{\frac{(3200-\bar{z})^2+(4700-\bar{z})^2+(2100-\bar{z})^2+(1500-\bar{z})^2}{4-1}}$$

2.

	x^*	y^*	z^*
KW1	$\frac{6'3-\bar{x}}{\sigma_x}$	$\frac{320-\bar{y}}{\sigma_y}$	$\frac{3200-\bar{z}}{\sigma_z}$
KW2	$\frac{4'7-\bar{x}}{\sigma_x}$	$\frac{180-\bar{y}}{\sigma_y}$	$\frac{4700-\bar{z}}{\sigma_z}$
KW3	$\frac{3'2-\bar{x}}{\sigma_x}$	$\frac{170-\bar{y}}{\sigma_y}$	$\frac{2100-\bar{z}}{\sigma_z}$

KW4 $\frac{38-\bar{x}}{\sigma_x}$ $\frac{179-\bar{y}}{\sigma_y}$ $\frac{1500-\bar{z}}{\sigma_z}$

$$\rho_{x^*,y^*} = \frac{\sum x_i^* y_i^*}{\sum x_i^* \sum y_i^*} \qquad \rho_{x^*,z^*} = \frac{\sum x_i^* z_i^*}{\sum x_i^* \sum z_i^*} \qquad \rho_{y^*,z^*} = \frac{\sum y_i^* z_i^*}{\sum y_i^* \sum z_i^*}$$

$$\boxed{\mathrm{KORR}[x^*, y^*, z^*] = \begin{bmatrix} 1 & \rho_{x^*,y^*} & \rho_{x^*,z^*} \\ \rho_{x^*,y^*} & 1 & \rho_{y^*,z^*} \\ \rho_{x^*,z^*} & \rho_{y^*,z^*} & 1 \end{bmatrix}}$$