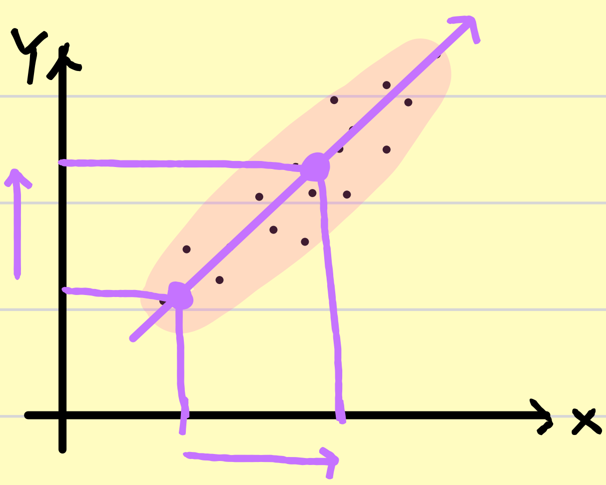
20250409_WIN2_Statistik

KORRELATION (Pearson's Korrelation)

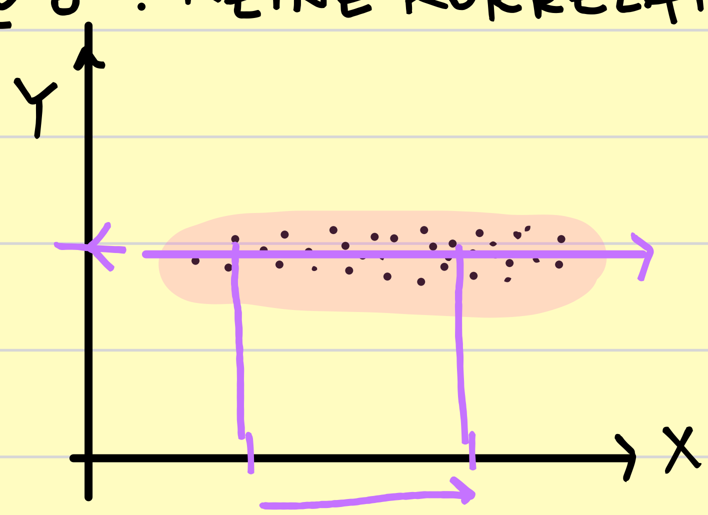
$$\boxed{\rho \equiv \text{Rho} \equiv \rho(X,Y) = \frac{\operatorname{Kov}(X,Y)}{\sigma_X\,\sigma_Y}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \in [-1,1]$$

Interpretation:

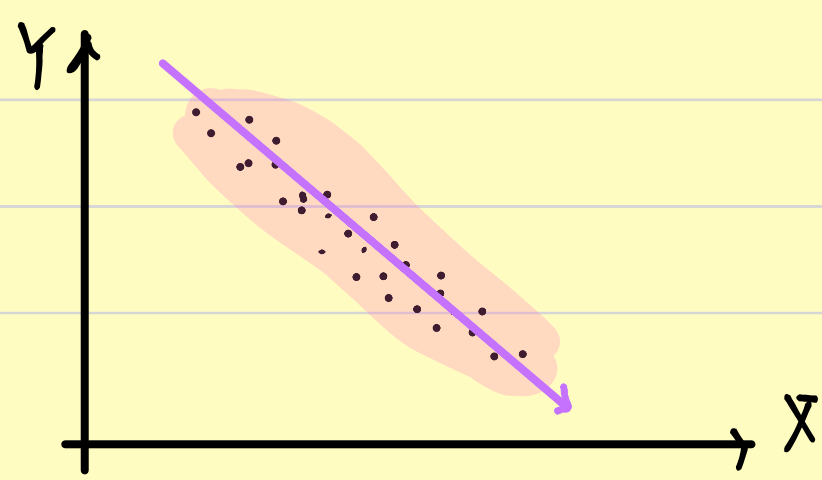
$\rho > 0$: POSITIVE KORRELATION. Die Variablen verhalten sich so, dass wenn eine wächst, die andere auch wächst.



$\rho \simeq 0$: KEINE KORRELATION. Die Variablen verhalten sich so, dass wenn eine sich ändert, die andere sich nicht ändert.



$\rho < 0$: NEGATIVE KORRELATION: Die Variablen 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 die DLZ sich um 15% verbessern würde.

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

$\downarrow$ NORMIERUNG

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

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

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

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

$\bar{x}^* = 1 \qquad \sigma_{x^*} = 0$

$\bar{y}^* = 1 \qquad \sigma_{y^*} = 0$

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

$$= \frac{(x_1^* - y_1^*)(x_2^* - y_2^*)(x_3^* - y_3^*)(x_4^* - y_4^*)}{(x_1^* + x_2^* + x_3^* + x_4^*)(y_1^* + y_2^* + y_3^* + y_4^*)} = \rho^*$$

Reduktion der DLZ

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

DLZ verbessern bedeutet ≡ DLZ sinkt.

Q verbessern bedeutet ≡ ppm sinkt.

KORRELATIONSMATRIX

- Definition für 3 Variabeln : $KORR(X,Y,Z) = \begin{bmatrix} 1 & \rho_{XY} & \rho_{XZ} \\ \rho_{YX} & 1 & \rho_{YZ} \\ \rho_{ZX} & \rho_{ZY} & 1 \end{bmatrix}$

- Korrelation ist biyektiv:

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

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

Übung: bitte ermitteln Sie die Korrelationsmatrix des Kennzahlensystems mit 3 KPIs (DLZ, $\frac{€}{St}$, Q):

	DURCHLAUFZEIT	$\frac{€}{\text{Stück}}$	QUALITÄT
KW1	6'3	320	3200
KW2	4'7	180	4700
KW3	3'2	170	2100
KW4	3'8	179	1500

1. NORMIEREN:

$\bar{x}, \sigma_x \; ; \; \bar{y}, \sigma_y \; ; \; \bar{z} \; ; \; \sigma_z$

$$x^* = \frac{x - \bar{x}}{\sigma_x} \qquad y^* = \frac{y - \bar{y}}{\sigma_y} \qquad z^* = \frac{z - \bar{z}}{\sigma_z}$$

2. KORRELATIONSMATRIX:

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