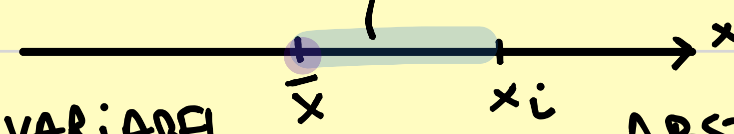
$$\boxed{m_2 = VAR(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{N}}$$

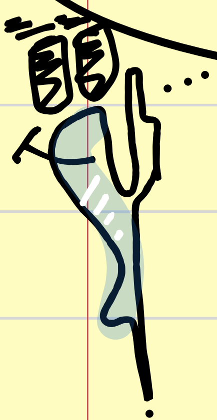
$(x_i - \bar{x}) \equiv$ Abstand



1 VARIABEL

ABSTAND ZUM MITTELWERT

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



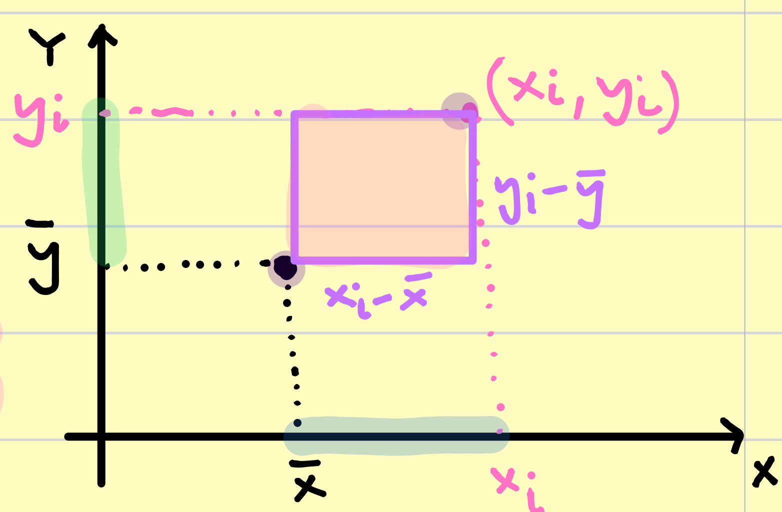
KOVARIANZMATRIX

(*) N−1 Inferenz Statistik

Definition KOVARIANZ von 2 variabeln X und Y als:

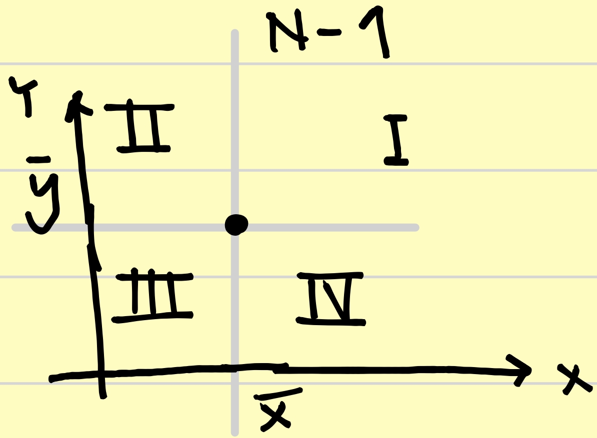
$$\boxed{Kov = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{N-1}} \quad (*)$$

$$= \frac{\sum_{i=1}^{n} \text{FLÄCHE der VIERECKEN ZUM MITTELWERT}}{N-1}$$



$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{N}$$



Die Kovarianz hat einen Symbol!! (anders als die Varianz)

Interpretation der Kovarianz:

$Kov(X,Y) \approx 0 \rightarrow$ Die meisten Punkte sind um den Mittelwert verteilt.

$Kov(X,Y) \gg 0$ / $\ll 0 \rightarrow$ Die meisten Punkte sind vom Mittelwert entfernt.

Beispiel. Gegeben wird ein Kennzahlensystem mit 2 KPI (Key Performance Indikatoren): Durchlaufzeit und €/Stück. Bitte ermitteln Sie die Kovarianz beider Kennzahlen.

	DLZ (Tage)	€/Stück
KW 1	6'3	230
KW 2	4'7	180
KW 3	3'2	170
KW 4	3'8	175

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

$$\overline{\frac{€}{Stück}} = \frac{\sum €/Stk_i}{4} = \frac{230+180+170+175}{4} = 188'75$$

$$KOV\left(DLZ, \frac{€}{Stück}\right) = \frac{\sum_{i=1}^{4}\left(DLZ_i - \overline{DLZ}\right)\left(\frac{€}{Stk_i} - \overline{\frac{€}{Stück}}\right)}{4-1} =$$

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

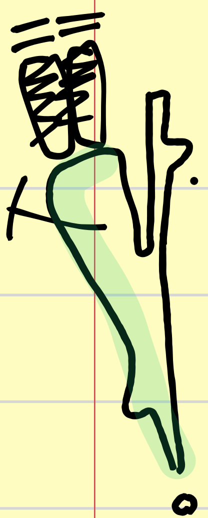
$= \ldots$

KOVARIANZ ✓

Die Kovarianzmatrix von 3 Variabeln X, Y, Z wird definiert als:

$$KOV(X,Y,Z) = \begin{array}{c|ccc} & X & Y & Z \\ \hline X & VAR(X) & KOV(X,Y) & KOV(X,Z) \\ Y & KOV(Y,X) & VAR(Y) & KOV(Y,Z) \\ Z & KOV(Z,X) & KOV(Z,Y) & VAR(Z) \end{array}$$

so mit N DIMENSIONEN erweiterbar ...



Die Kovarianzmatrix ist symmetrisch.

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

Übung. Gegeben wird ein Kennzahlensystem mit 3 KPIs: DLZ, €/Stück, Qualität
(Tage) (€/Stück) $\left(\frac{\#\text{Fehler}}{\text{Million produzierte Teile}} = \text{ppm}\right)$

	X DLZ (Tage)	Y €/Stück	Z Qualität (ppm)	
KW1	6'3	320	3200	$\bar{x} = 4'5$
KW2	4'7	180	4700	$\bar{y} = 188'75$
KW3	3'2	170	2100	$\bar{z} = 2875$
KW4	3'8	175	1500	

Bitte die Kovarianzmatrix des Systems ermitteln.

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

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

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

$$\mathrm{VAR}(Z) = \frac{\sum (z_i-\bar{z})^2}{N} = \frac{(3200-2875)^2+(4700-2875)^2+(2100-2875)^2+(1500-2875)^2}{4}$$

$$\mathrm{kov}(X,Y) = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{N-1} = \frac{(6'3-4'5)(320-188'75) + (4'7-4'5)(180-188'75) + (3'2-4'5)(170-188'75) + (3'8-4'5)(175-188'75)}{4-1} =$$

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

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

KOVARIANZMATRIX ✓

KORRELATION · PRODUKT-MOMENT VON PEARSON

$\rho \equiv$ Rho = KORRELATIONS-KOEFFIZIENT von PEARSON $\in [-1,1]$

Wird definiert als:

$$\boxed{\rho_{XY} = \rho = \frac{\mathrm{kov}(X,Y)}{\sqrt{\mathrm{VAR}(X)}\sqrt{\mathrm{VAR}(Y)}}}$$

$$\boxed{\rho = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}}$$

$\rho > 0$: POSITIVE KORRELATION

$\rho = 0$: KEINE KORRELATION

$\rho < 0$: NEGATIVE KORRELATION

Beispiel:

20% Änderung der Kosten $\frac{€}{\text{Stück}}$, wie beeinflussen das Gesamtergebnis vom Werk?

$\rho = -0{,}6$: wenn die Kosten senken, erhöht sich das Ergebnis um 60%.

$$\underbrace{0{,}2}_{\text{Änderung der €/Stück}} \cdot \underbrace{(-0{,}6)}_{\text{Korrelation Pearson}} \longrightarrow -0{,}12 \ (12\%)$$

12% Ergebniserhöhung

Übung. Bitte ermitteln Sie um wie viel verbessert sich die Qualität eines Produktes (Senkung der Ausschußzahlen), wenn sich die DLZ Werte um 15% verbessern (Senkung der DLZ).

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

$$\bar{x} = \frac{7{,}3 + 6{,}7 + 5{,}8 + 5{,}6}{4} = 6{,}35$$

$$\bar{y} = \frac{3200 + 2700 + 1900 + 1700}{4} = 2375$$

$$\rho(x,y) = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} =$$

$$= \frac{(7'3-6'35)(3200-2375)+(6'7-6'35)(2700-2375)+ \\ +(5'8-6'35)(1900-2375)+(5'6-6'35)(1700-2375)}{\sqrt{(7'3-6'35)^2+(6'7-6'35)^2+(5'8-6'35)^2+(5'6-6'35)^2} \cdot \sqrt{(3200-2375)^2+(2700-2375)^2+(1900-2375)^2+(1700-2375)^2}} =$$

$$= \dots$$

15% Senkung DLZ $\longrightarrow$?% Senkung ppm

$$-0'15 \cdot \rho_{XY} \longrightarrow - \underbrace{\dots}\ \%\ \text{ppm}$$

PEARSON'S KORRELATION ✓

KORRELATIONSMATRIX

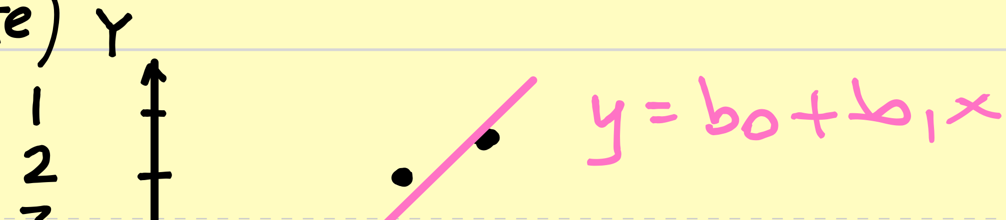
$$KORR(X,Y,Z) = \begin{bmatrix} 1 & \rho_{XY} & \rho_{XZ} \\ \rho_{YX} & 1 & \rho_{YZ} \\ \rho_{ZX} & \rho_{ZY} & 1 \end{bmatrix}$$ (per Definition)

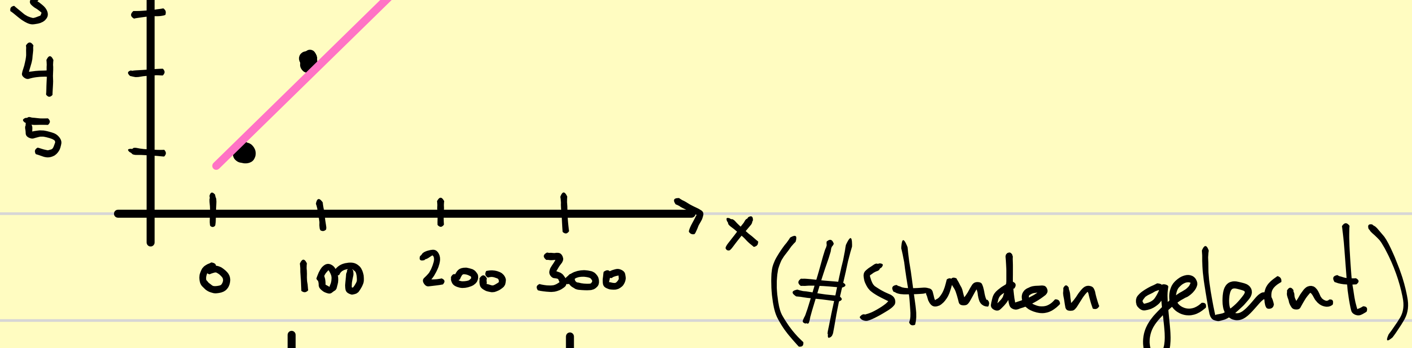
Analog KOVMATRIX $KORR(X,Y) = KORR(Y,X)$ KORRMATRIX ✓

PROGNOSE. Zukunftsvorhersage.

$\boxed{\text{Lineare Prognose. Lineare Regression}}$

(Note)





#Std	Note
X	Y
250	1,3
180	2,0
100	3,7
20	5,0

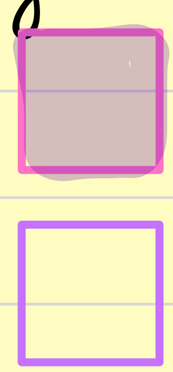
Gegeben sind die Daten und wir wollen eine Regression kalkulieren damit, wenn die #Stunden bekannt sind, die Note hervorgesagt werden kann.

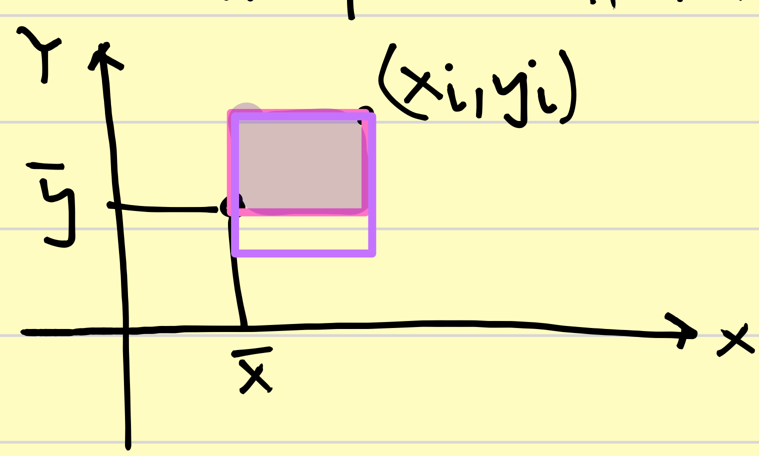
Schritt 1. Mittelwert der Variabeln ermitteln.

$$\bar{x} = \frac{250+180+100+20}{4} = 137'5 \qquad \bar{y} = \frac{1'3+2+3'7+5'0}{4} = 3$$

Schritt 2. Neigung der Linie .. b_1 ermitteln.

$$b_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$





- wenn beide Flächen gleich sind $\rightarrow b_1 = 1 \rightarrow 45°$
- wenn die obere Fläche größer als die untere $\rightarrow b_1 > 1 \rightarrow > 45$
- " " " " kleiner " " " $\rightarrow b_1 < 1 \rightarrow < 45$

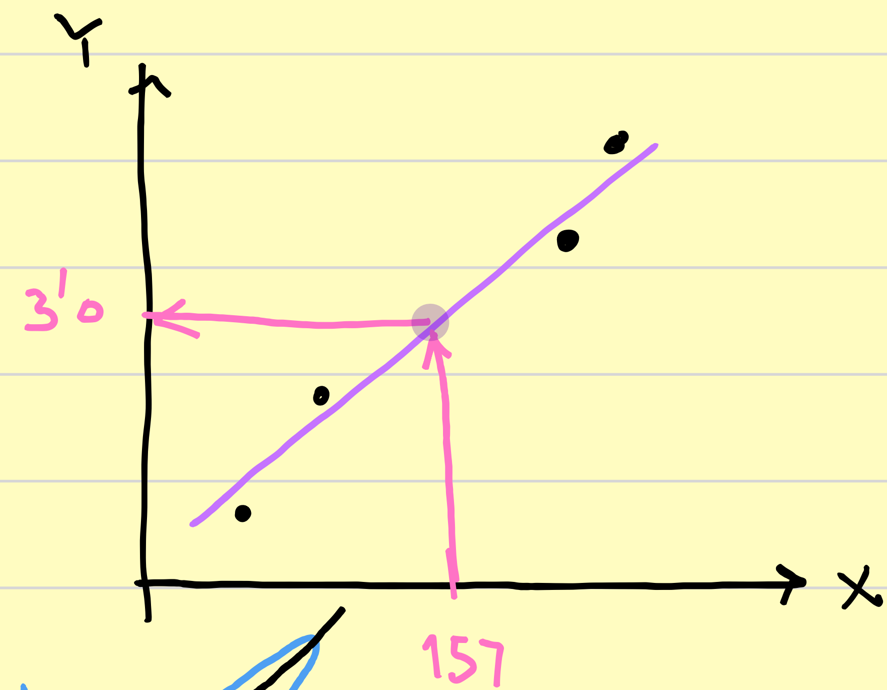
$$b_1 = \frac{(250-137'5)(1'3-3)+(180-137'5)(2-3)+(100-137'5)(3'7-3)+ \\ +(20-137'5)(5-3)}{(250-137'5)^2+(180-137'5)^2+(100-137'5)^2+(20-137'5)^2}$$

(*)

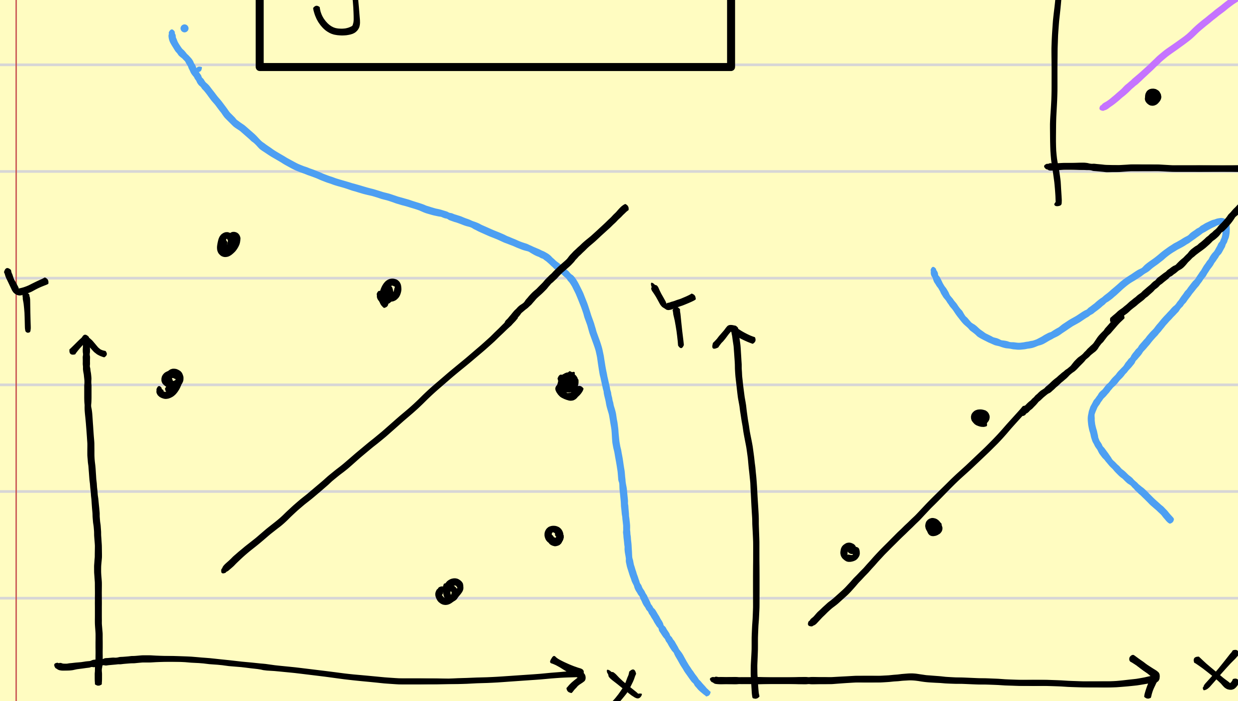
Schritt 3. Da die lineare Regression durch den Mittelwert geht, erfüllt der Mittelwert die Gleichung der Linie:

$$\bar{y} = b_1 \bar{x} + b_0 \longrightarrow b_0 = \bar{y} - b_1 \bar{x} = \ldots = b_0 \checkmark$$

bereits bekannt (*)



$$\boxed{y = b_1 x + b_0}$$



- HAUPTKOMPONENTEN -