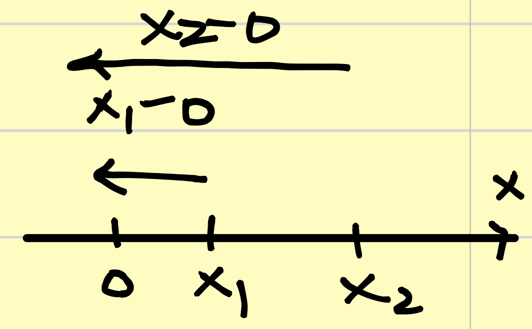
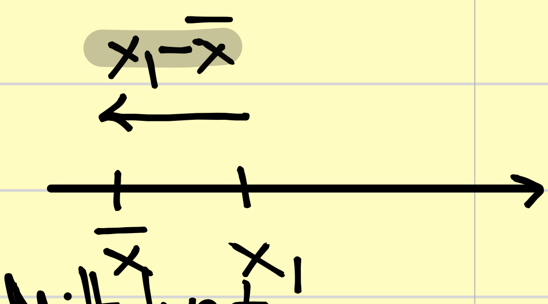
Hauptkomponentenanalyse

1 VARIABLE



- $x = \{x_1, x_2, \cdots, x_n\}$ $\quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$

MITTELWERT: die Summe der Abstände zum NULL.



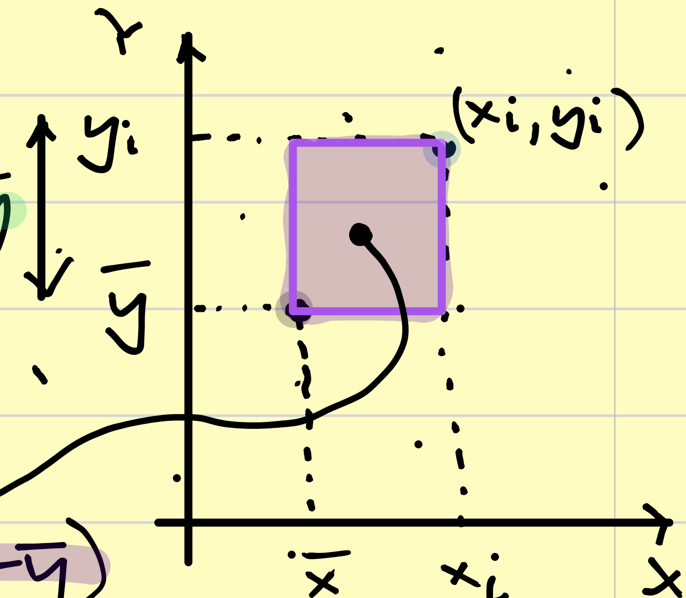
- $VAR(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}$

VARIANZ: die Summe der Abstände (in Quadrat) zum Mittelwert!

2 VARIABLEN

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

$\bar{x} = \frac{\sum x_i}{n} \quad \bar{y} = \frac{\sum y_i}{n}$



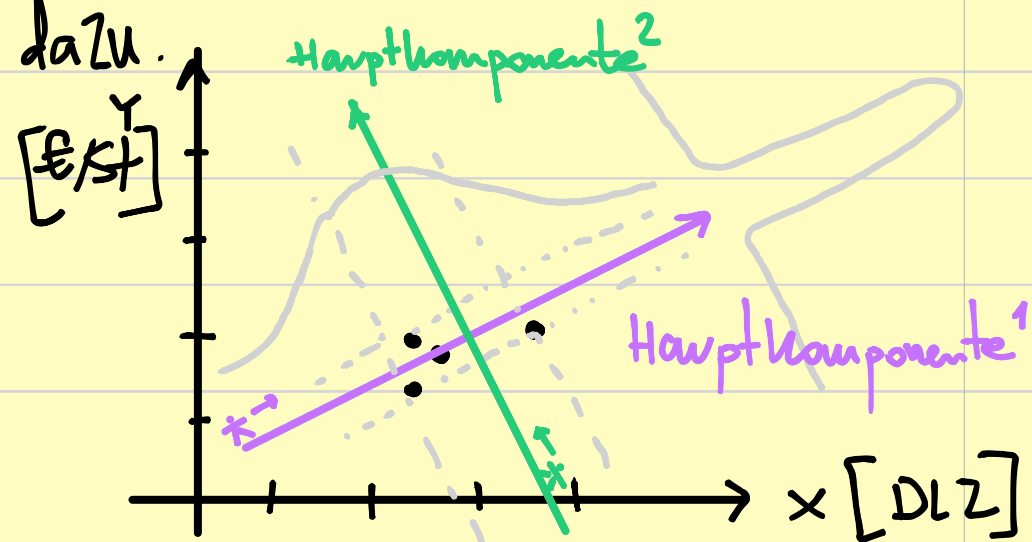
Fläche $= (x_i - \bar{x})(y_i - \bar{y})$

KOVARIANZ: die Summe der Flächen der Parallelogramme der Punkte zum Mittelwert.

- KOVARIANZMATRIX: $\text{KOV MATRIX}[X,Y] = A = \begin{bmatrix} VAR[X] & KOV[X,Y] \\ KOV[X,Y] & VAR[Y] \end{bmatrix}$

Beispiel. Ein Kennzahlensystem besteht aus 2 KPIs, DLZ & €/Stück. Bitte ermitteln Sie die Kov. MATRIX dazu.

	DLZ (Tage)	Kosten (€/St)
Jan	37	21
Feb	28	18
März	25	15
April	27	20



$$\text{Kov Matrix}\left[X,Y\right] = A = \begin{bmatrix} VAR[X] & KOV[X,Y] \\ KOV[X,Y] & VAR[Y] \end{bmatrix} = \begin{bmatrix} 21'187 & 7'87 \\ 7'87 & 5'25 \end{bmatrix}$$

- $\bar{x} = \frac{\sum x_i}{n} = \frac{37+28+25+27}{4} = 29'25$; $\bar{y} = \frac{\sum y_i}{n} = \frac{21+18+15+20}{4} = 18'5$

- $VAR[X] = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{(37-29'25)^2 + (28-29'25)^2 + (25-29'25)^2 + (27-29'25)^2}{4} = 21'187$

$VAR[Y] = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{(21-18'5)^2 + (18-18'5)^2 + (15-18'5)^2 + (20-18'5)^2}{4} = 7'87$

$KOV[X,Y] = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{(37-29'25)(21-18'5) + (28-29'25)(18-18'5) + (25-29'25)(15-18'5) + (27-29'25)(20-18'5)}{3}$

$= 5'25$

- Die erste HK beschreibt die Variabilität der Daten am besten. (schmale Gausskurve in die Richtung).
- Die zweite HK ist senkrecht zur HK¹ (und beschreibt der Rest der Variabilität der Daten).

$\boxed{\text{· Die Eigenvektoren der Kovarianzmatrix sind die Hauptkomponenten der Daten.}}$

$\boxed{A \cdot \vec{v} = \lambda \cdot \vec{v}}$ λ = Eigenwert $\vec{v}$ = Eigenvektoren

$\hookrightarrow$ 1. $\det(A - \lambda I) = 0 \rightarrow \lambda \longrightarrow \vec{v}$

$$A=\begin{bmatrix} 21'187 & 7'87 \\ 7'87 & 5'25 \end{bmatrix}$$

$$\det\left[A-\lambda I\right]=0 \rightarrow \det\left[\begin{bmatrix} 21'187 & 7'87 \\ 7'87 & 5'25 \end{bmatrix}-\lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right]=0$$

$$\rightarrow \det\left[\begin{bmatrix} 21'187-\lambda & 7'87 \\ 7'87 & 5'25-\lambda \end{bmatrix}\right]=0 \rightarrow$$

$$\boxed{\det\begin{bmatrix} a & c \\ b & d \end{bmatrix}=ad-cd} \rightarrow (21'187-\lambda)(5'25-\lambda)-7'87^2=0$$

$$\rightarrow \lambda^2-26'437\lambda+51'41=0$$

$$\boxed{\begin{gathered} ax^2+bx+c=0 \\ x=\frac{-b\pm\sqrt{b^2-4ac}}{2a} \end{gathered}}$$

$$\lambda=\frac{26'437\pm\sqrt{26'437^2-4.51'41}}{2}= \begin{cases} 24'32=\lambda_1. \\ 2'113=\lambda_2. \end{cases}$$

Eigenwerte

$$\boxed{\lambda_1=24'32} \rightarrow \vec{v_1}=\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix}$$

$$A\cdot\vec{v_1}=\lambda_1\vec{v_1} \rightarrow \begin{bmatrix} 21'187 & 7'87 \\ 7'87 & 5'25 \end{bmatrix}\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix}=24'32\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} \rightarrow$$

$$\rightarrow \begin{aligned} 21'187.v_{11}+7'87.v_{12}&=24'32v_{11} \\ 7'87.v_{11}+5'25.v_{12}&=24'32v_{12} \end{aligned} \rightarrow$$

$$7'87.v_{12}=(24'32-21'187)v_{11} \rightarrow v_{12}=1'5665v_{11}$$

$$v_{11}=1 \rightarrow v_{12}=1'5665 \qquad \vec{v_1}=\begin{bmatrix} 1 \\ 1'5665 \end{bmatrix}$$

Eine Änderung einer Einheit der Kosten €/Stück, bedeutet 1'5665 Einheiten Änderung der DLZ.

Um die Kosten 1€ zu reduzieren, müssen wir 1'5665 Tage DLZ reduzieren.