Kovarianz und Korrelationskoeffizient

$x := \begin{pmatrix} 21 \\ 28 \\ 35 \\ 46 \\ 55 \end{pmatrix} \quad y := \begin{pmatrix} 1850 \\ 1800 \\ 2230 \\ 2500 \\ 2560 \end{pmatrix}$ Ausprägungen der Merkmale x und y[1)]

$n := \text{länge}(x) = 5$ Anzahl der Wertepaare von x und y

$\text{ORIGIN} \equiv 1$ Startwert für alle verwendeten Indizes

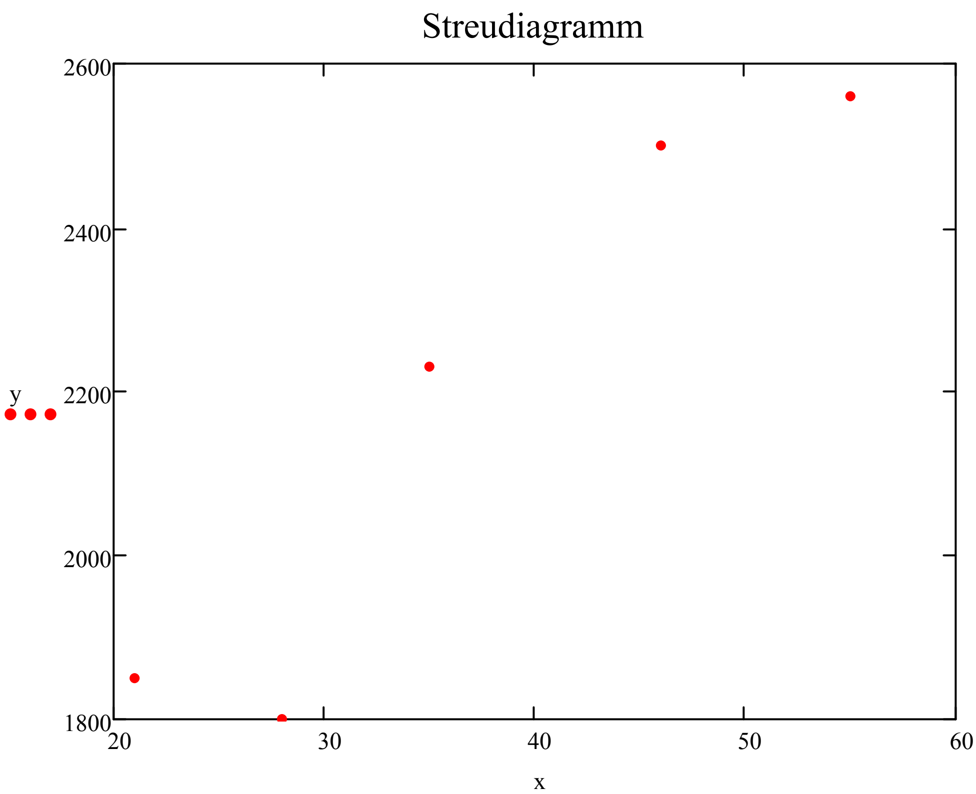
$i := \text{ORIGIN} .. n$ Index der Merkmalsausprägungen

$x_d := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = 37$ Durchschnittliche Ausprägung des Merkmals x

$\text{mittelwert}(x) = 37$

$y_d := \frac{1}{n} \cdot \sum_{i=1}^{n} y_i = 2188$ Durchschnittliche Ausprägung des Merkmals y

$\text{mittelwert}(y) = 2188$



1) Quelle für das Beispiel: A. Quatember, Statistik ohne Angst vor Formeln, 2. Aufl. München 2008, S. 67 ff.

Kovarianz und Korrelationskoeffizient

$\Delta x := \min(x) .. \max(x)$ Parameter für die folgenden Grafiken

$\Delta y := \min(y) .. \max(y)$

$B1 := \begin{cases} x_1 .. x_d & \text{if } x_1 < x_d \\ x_d .. x_1 & \text{otherwise} \end{cases}$ $H1 := \begin{cases} y_1 .. y_d & \text{if } y_1 < y_d \\ y_d .. y_1 & \text{otherwise} \end{cases}$

Abweichung 1:

$(x_1 - x_d)\cdot(y_1 - y_d) = 5408$

$B2 := \begin{cases} x_2 .. x_d & \text{if } x_2 < x_d \\ x_d .. x_2 & \text{otherwise} \end{cases}$ $H2 := \begin{cases} y_2 .. y_d & \text{if } y_2 < y_d \\ y_d .. y_2 & \text{otherwise} \end{cases}$

Abweichung 2:

$(x_2 - x_d)\cdot(y_2 - y_d) = 3492$

$B3 := \begin{cases} x_3 .. x_d & \text{if } x_3 < x_d \\ x_d .. x_3 & \text{otherwise} \end{cases}$ $H3 := \begin{cases} y_3 .. y_d & \text{if } y_3 < y_d \\ y_d .. y_3 & \text{otherwise} \end{cases}$

Abweichung 3:

$(x_3 - x_d)\cdot(y_3 - y_d) = -84$

$B4 := \begin{cases} x_4 .. x_d & \text{if } x_4 < x_d \\ x_d .. x_4 & \end{cases}$ $H4 := \begin{cases} y_4 .. y_d & \text{if } y_4 < y_d \\ y_d .. y_4 & \text{otherwise} \end{cases}$

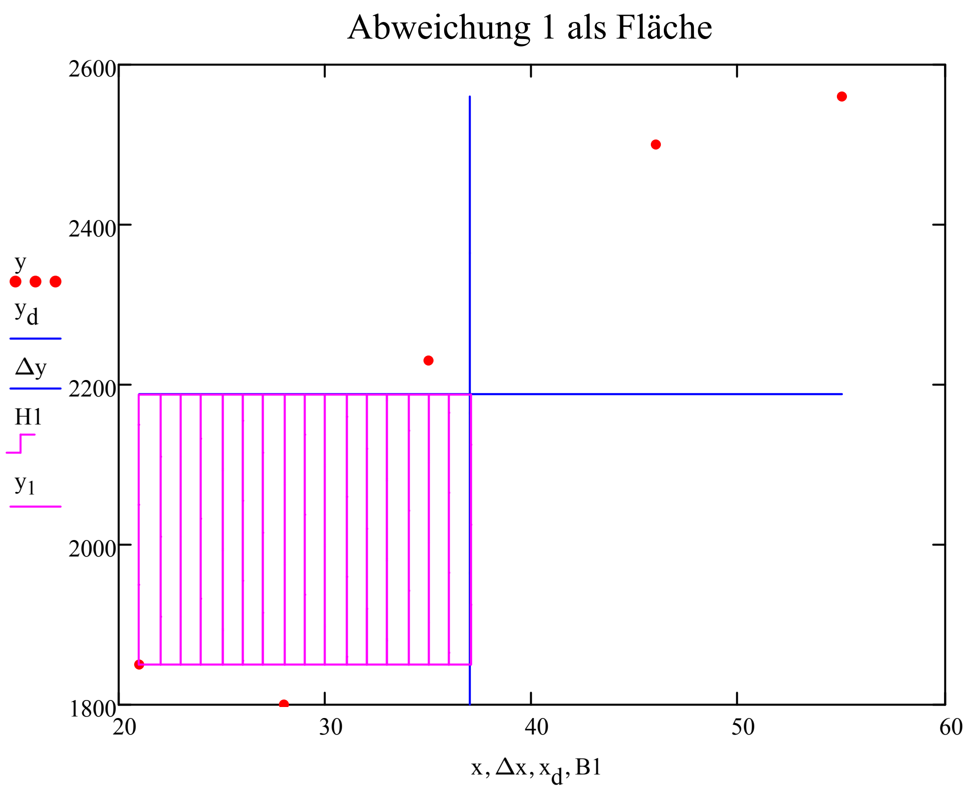
Abweichung 4:

$(x_4 - x_d)\cdot(y_4 - y_d) = 2808$

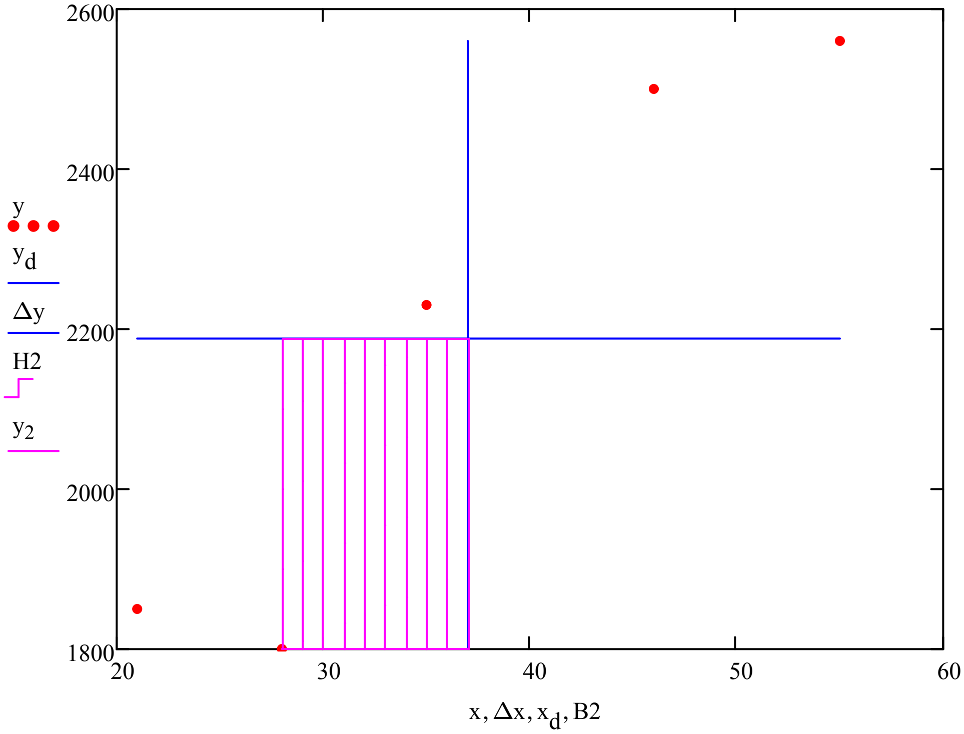
$B5 := \begin{cases} x_5 .. x_d & \text{if } x_5 < x_d \\ x_d .. x_5 & \text{otherwise} \end{cases}$ $H5 := \begin{cases} y_5 .. y_d & \text{if } y_5 < y_d \\ y_d .. y_5 & \text{otherwise} \end{cases}$

Abweichung 5:

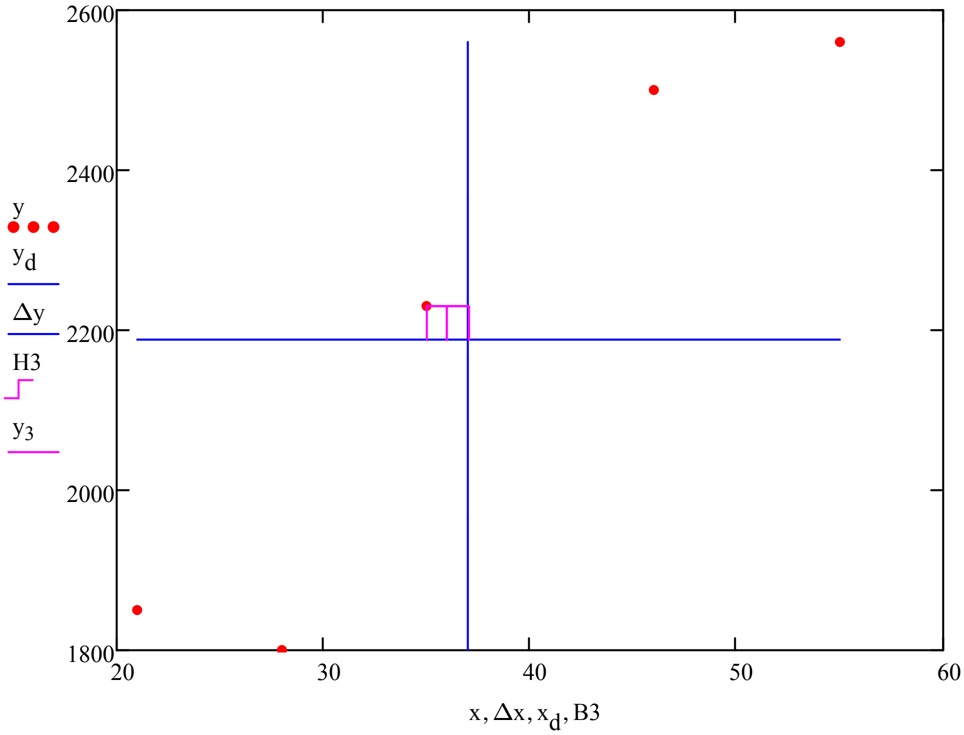
$(x_5 - x_d)\cdot(y_5 - y_d) = 6696$

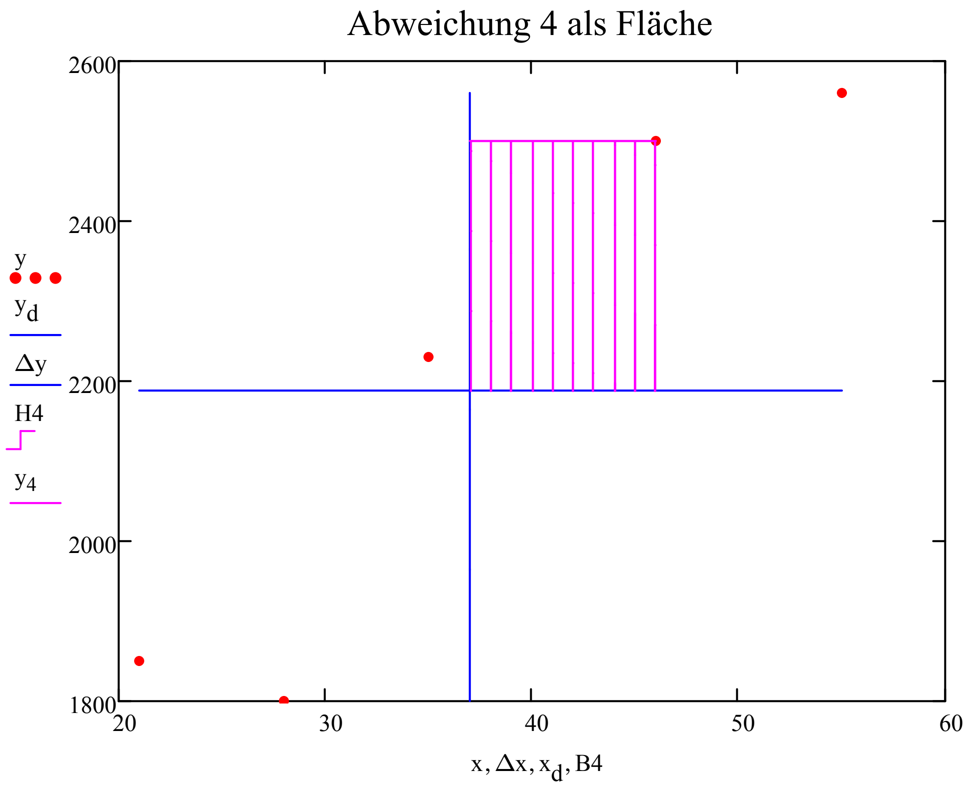


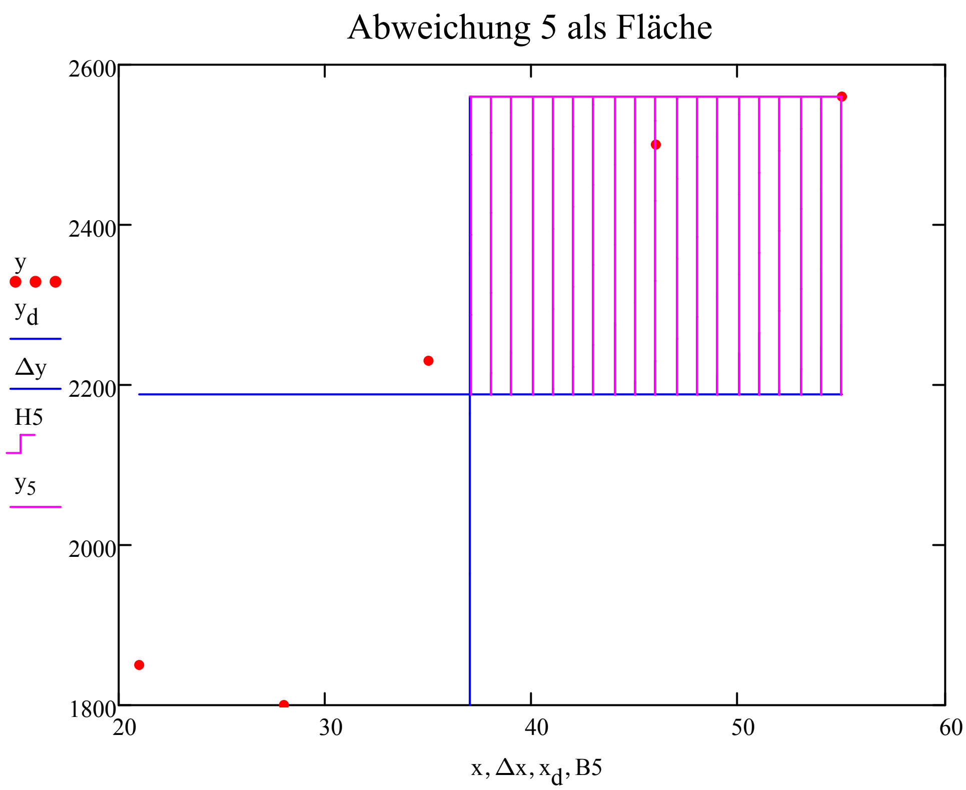
Abweichung 2 als Fläche



Abweichung 3 als Fläche




Abweichung 4 als Fläche
2600
2400
2200
2000
1800
20
30
40
50
60
y
y_d
Δy
H4
y_4
x, Δx, x_d, B4


Abweichung 5 als Fläche
2600
2400
2200
2000
1800
20
30
40
50
60
y
y_d
Δy
H5
y_5
x, Δx, x_d, B5

Kovarianz und Korrelationskoeffizient

Wenn x und y die Grundgesamtheit darstellen:

$$\frac{1}{n}\cdot\sum_{i=1}^{n}\left[\left(x_i - x_d\right)\cdot\left(y_i - y_d\right)\right] = 3664$$

Empirische Kovarianz

$\mathrm{kvar}(x, y) = 3664$

$$\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i - x_d\right)^2} = 12.215$$

Standardabweichung von x

$\mathrm{stdev}(x) = 12.215$

$$\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(y_i - y_d\right)^2} = 316.948$$

Standardabweichung von y

$\mathrm{stdev}(y) = 316.948$

Wenn x und y eine Stichprobe darstellen:

$$\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left[\left(x_i - x_d\right)\cdot\left(y_i - y_d\right)\right] = 4580$$

Kovarianz einer Stichprobe

$$\sqrt{\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(x_i - x_d\right)^2} = 13.657$$

Standardabweichung von x

$\mathrm{Stdev}(x) = 13.657$

$$\sqrt{\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(y_i - y_d\right)^2} = 354.359$$

Standardabweichung von y

$\mathrm{Stdev}(y) = 354.359$

Unabhängig davon, ob x und y die Grundgesamtheit oder eine Stichprobe darstellen:

$$\frac{\sum_{i=1}^{n}\left[\left(x_i - x_d\right)\cdot\left(y_i - y_d\right)\right]}{\sqrt{\sum_{i=1}^{n}\left(x_i - x_d\right)^2}\cdot\sqrt{\sum_{i=1}^{n}\left(y_i - y_d\right)^2}} = 0.946$$

Korrelationskoeffizient

$\mathrm{korr}(x, y) = 0.946$