Übungsblatt - Methode der kleinsten Quadrate

Aufgabe 1) Manuelle Bestimmung einer Ausgleichsgeraden

Gegeben ist folgender Datensatz:

x	1	2,5	3,5	5
y	2,5	3,5	5	5

a) Berechnen Sie die Ausgleichsgerade $\hat{y} = ax + b$ mithilfe der Methode des kleinsten Quadrate.

$$m = \frac{\sum_{i=1}^{4} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{4} (x_i - \bar{x})^2} \qquad b = \bar{y} - m\bar{x}$$

Formeln aus der Vorlesung $\bar{x}$ und $\bar{y}$, die Mittelwerte der bekannten x und y Werte.

$$\bar{x} = \frac{1}{4}(1 + 2,5 + 3,5 + 5) = \frac{1}{4} 12 = \underline{\underline{3}}$$

$$\bar{y} = \frac{1}{4}(2,5 + 3,5 + 5 + 5) = \frac{1}{4}(16) = \underline{\underline{4}}$$

Zähler von m: $(1-3)(2,5-4) + (2,5-3)(3,5-4) + (3,5-3)(5-4) + (5-3)(5-4)$

$= (-2)(-1.5) + (-0,5)(-0,5) + 0.5 \cdot 1 + 2 \cdot 1$

$= 3 + 0,25 + 0,5 + 2 = \underline{\underline{5,75}}$

Nenner von m: $(1-3)^2 + (2,5-3)^2 + (3,5-3)^2 + (5-3)^2$

$= (-2)^2 + (-0.5)^2 + (0,5)^2 + 2^2$

$= 4 + 0,25 + 0,25 + 4$

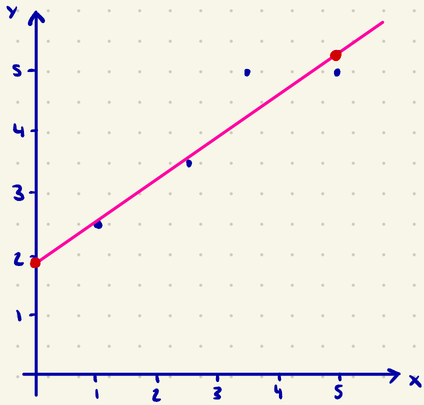
$= \underline{\underline{8,5}}$

$$m = \frac{5,75}{8,5} \approx 0,676 \qquad b = 4 - 0,676 \cdot 3 = 1,972$$

Die gesuchte Ausgleichsgerade ist somit:

$$\underline{\hat{y} = 0,676\,x + 1,972}$$

b) Tragen Sie die Datenpunkte in ein Koordinatensystem ein und zeichnen Sie die Ausgleichsgerade ein.



$\hat{y} = 0{,}676\,x + 1{,}972$

$f(0) = 1{,}972$

$f(5) = 0{,}676 \cdot 5 + 1{,}972$

$= 5{,}352$

c) Berechnen Sie die Residuen r_i, die quadrierten Abweichungen und den MSE der Ausgleichs_geraden

Gegeben war:

x	1	2,5	3,5	5
y	2,5	3,5	5	5

$\hat{y}_1 = f(1) \approx 2{,}647$ $\quad y_1 = 2{,}5$ $\quad r_1 = \hat{y}_1 - y_1 = 0{,}147$ $\quad r_1^2 = 0{,}022$

$\hat{y}_2 = f(2{,}5) \approx 3{,}662$ $\quad y_2 = 3{,}5$ $\quad r_2 = \hat{y}_2 - y_2 = 0{,}162$ $\quad r_2^2 = 0{,}026$

$\hat{y}_3 = f(3{,}5) \approx 4{,}338$ $\quad y_3 = 5$ $\quad r_3 = \hat{y}_3 - y_3 = -0{,}662$ $\quad r_3^2 = 0{,}438$

$\hat{y}_4 = f(5) \approx 5{,}352$ $\quad y_4 = 5$ $\quad r_4 = \hat{y}_4 - y_4 = 0{,}352$ $\quad r_4^2 = 0{,}124$

Der MSE $= \frac{1}{4}(0{,}022 + 0{,}026 + 0{,}438 + 0{,}124) = \underline{\underline{0{,}1525}}$

Aufgabe 2) Einfluss von Ausreißern

Ergänzen Sie den Datensatz aus Aufgabe 1 um einen Ausreißer bei $x=2$ mit $y=7$.

Die neue Datenbasis ist dann:

x	1	2,5	3,5	5	2
y	2,5	3,5	5	5	7

a) Wie verändert sich der MSE der in Aufgabe 1 bestimmten Ausgleichsgeraden?

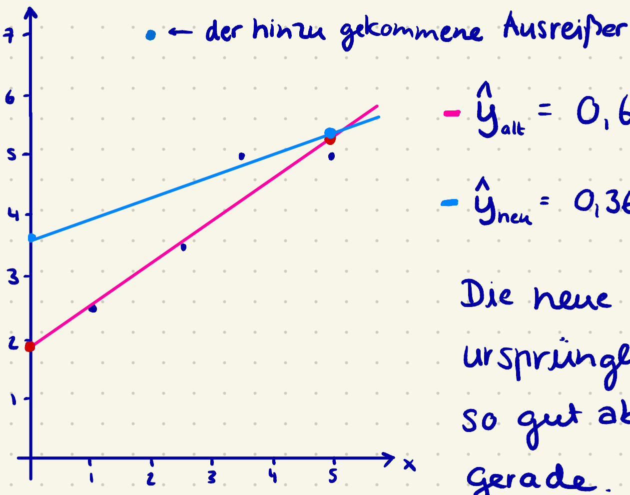
$\hat{y}_s = f(2) = 3{,}324 \qquad y_s = 7 \qquad r_s = \hat{y}_s - y_s = 3{,}676 \qquad r_s^2 = 13{,}513$

$$MSE = \frac{1}{5}(0{,}022 + 0{,}026 + 0{,}438 + 0{,}124 + 13{,}513) = \underline{\underline{2{,}8246}}$$

Der MSE der Ausgleichsgeraden hat sich deutlich verschlechtert.

b) Wenn wir für die angepassten Werte die Ausgleichsgerade neu bestimmen, erhalten wir für $m = 0{,}36$ und $b = 3{,}59$.

Vergleichen Sie die beiden Geraden graphisch. Und beschreiben Sie den Einfluss des Ausreißers auf das Modell.



$\hat{y}_{alt} = 0{,}676\,x + 1{,}972$

$\hat{y}_{neu} = 0{,}36\,x + 3{,}59$

Die neue Gerade bildet die ursprünglichen Punkte nicht mehr so gut ab wie die vorherige Gerade.

Aufgabe 3) Fehlermaße verlgleichen

Gegeben sind zwei Modelle A und B mit folgenden Residuen:

i	r_i (Modell A)	r_i (Modell B)
1	-0,2	-0,5
2	0,1	0,3
3	-0,3	0,1
4	0,4	-0,6

a) Berechnen Sie für beide Modelle: SSE, MSE und MAE

Modell A: $SSE = 0{,}04 + 0{,}01 + 0{,}09 + 0{,}16 = \underline{0{,}3}$

$MSE = \frac{1}{4}(SSE) = \frac{1}{4} \cdot 0{,}3 = \underline{0{,}075}$

$MAE = \frac{1}{4}(0{,}2 + 0{,}1 + 0{,}3 + 0{,}4) = \underline{0{,}25}$

Modell B: $SSE = 0{,}25 + 0{,}09 + 0{,}01 + 0{,}36 = \underline{0{,}71}$

$MSE = \frac{1}{4}(SSE) = \underline{0{,}1775}$

$MAE = 0{,}5 + 0{,}3 + 0{,}1 + 0{,}6 = \underline{1{,}5}$

b) Welches Modell würden Sie bevorzugen? Begründen Sie Ihre Entscheidung.

	Modell A	Modell B
SSE	0,3	0,71
MSE	0,075	0,1775
MAE	0,25	1,5

Modell A hat für jedes berechnete Fehlermaß den kleineren Wert, deshalb ist anzunehmen, dass Modell A das bessere Modell, und somit zu bevorzugen ist.