

## Ich kann die Regression zweidimensionaler Daten erklären.

- c 1 Gegeben sind 5 Zahlenpaare  $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)$ , wobei alle  $x_i$  verschieden sind. Vervollständige den Satz, sodass eine mathematisch richtige Aussage entsteht. Wähle dazu die richtigen Satzteile aus.

Die lineare Regressionsfunktion  $f(x) = ax + b$  hat die Eigenschaft, dass die .... I. .... von  $f$  bezüglich  $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)$  minimal ist und dadurch den linearen Zusammenhang zwischen .... II. .... möglichst gut beschreibt.

I.
a. Summe der quadrierten Funktionswerte
b. Summe der quadratischen Abweichungen $\sum_{i=1}^5 f(x_i)^2 - y_i^2$
c. Summe der quadratischen Abweichungen $\sum_{i=1}^5 (f(x_i) - y_i)^2$

II.
a. den Funktionswerten von $f$
b. $x_1, x_2, x_3, x_4, x_5$ und $y_1, y_2, y_3, y_4, y_5$
c. $x_1, x_2, x_3, x_4, x_5$ und $f(x_1), f(x_2), f(x_3), f(x_4), f(x_5)$

- A, B, C 2 Eine neues Tablet wird anfangs um 1299€ angeboten. Nach 6 Monaten beträgt der Verkaufspreis 1219€ und nach 8 Monaten 1199€. Nimm an, dass der Verfall des Preises annähernd durch eine lineare Funktion beschrieben werden kann.

- Erkläre, welchen Sachverhalt man mit der linearen Regression in diesem Beispiel beschreiben kann.
- Erkläre anhand des Beispiels, was man unter der „Methode der kleinsten Quadrate“ versteht.
- Bestimme mit der Regressionsrechnung diejenige lineare Funktion  $f$ , für die die Summe der quadratischen Abweichungen bezüglich der angegebenen Zahlenpaare minimal ist. Berechne diese Summe der quadratischen Abweichungen.

- A, B 3 Um am Zentrum für Sportwissenschaft studieren zu können, muss man im Vorfeld einen sportlichen Eignungstest („Ergänzungsprüfung“) absolvieren. Im Zuge der Vorbereitungswochen für die Ergänzungsprüfung werden unter anderem die Leistungen für 20m-Sprints und 5er-Hop dokumentiert. In einer Trainingsgruppe erhält man folgende Werte

Laufzeit 20m-Sprint in sec	2,52	2,65	2,74	2,80	2,71
Sprungweite 5er-Hop in m	20,8	20,1	18,9	19,0	19,6

Bei der Datenauswertung soll der Zusammenhang zwischen den Ergebnissen der beiden Bewerbe möglichst gut durch eine lineare Funktion beschrieben werden. Die Datenauswertung wird von zwei Personen unabhängig durchgeführt. Jede der beiden Personen erhält eine andere Regressionsfunktion, nämlich  $f(x) = -7,3x + 38,7$  und  $g(x) = -7,1x + 38,8$ .

- Berechne für beide Funktionen die Summe der quadratischen Abweichungen bezüglich der in der Tabelle angegebenen Zahlenpaare.
- Begründe, welche der beiden Funktionen  $f$  und  $g$  die erreichte 5er-Hop-Weite besser voraussagt.

Lösungen zu:  
Ich kann die Regression zweidimensionaler Daten erklären.

- 1 Die lineare Regressionsfunktion  $f(x) = ax+b$  hat die Eigenschaft, dass die **I. c. die Summe der quadratischen Abweichungen**  $\sum_{i=1}^5 (f(x_i) - y_i)^2$  von  $f$  bezüglich  $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5)$  minimal ist und dadurch den linearen Zusammenhang zwischen **II. b.  $x_1, x_2, x_3, x_4, x_5$  und  $y_1, y_2, y_3, y_4, y_5$**  möglichst gut beschreibt.
- 2 a. Mit der linearen Regression kann man den linearen Zusammenhang zwischen dem Verkaufszeitpunkt und dem Tablet-Preis abschätzen. Man kann daraus abschätzen, welchen Preis das Tablet z.B. ein Jahr nach der Markteinführung haben wird.
- b. Die „Methode der kleinsten Quadrate“ ist die Vorgangsweise, mit der man jene Funktion ermittelt, die den Zusammenhang zwischen dem Verkaufszeitpunkt  $x$  und dem Tablet-Preis  $y$  möglichst gut beschreibt. Es wird dabei jene lineare Funktion  $P(x) = ax+b$  ( $x$ ...Anzahl der Monate seit Markteinführung des Tablets) bestimmt, für die die Summe der quadratischen Abweichungen von  $P(x)$  bezüglich des echten Tablet-Preises  $y$  minimal wird.
- c. mit Technologieeinsatz:  $f(x) = -12,69x + 1298,23$   
Berechnung der Summe der quadratischen Abweichungen (alle Zwischenergebnisse werden auf 2 Nachkommastellen gerundet angegeben):

$x_i$ Zeit seit der Markteinführung in Monaten	$y_i$ Verkaufspreis in €	$f(x_i)$	$(f(x_i) - y_i)^2$
0	1299	1298,23	0,59
6	1219	1222,08	9,47
8	1199	1196,69	5,33
<b>Summe der quadratischen Abweichungen:</b>			<b>15,38</b>

- 3 a. Wir erstellen eine Tabelle (alle Zwischenergebnisse werden auf 2 Nachkommastellen gerundet angegeben):

$x_i$ Laufzeit 20m-Sprint in sec	$y_i$ Sprungweite 5er-Hop in m	$f(x_i)$	$(f(x_i) - y_i)^2$	$g(x_i)$	$(g(x_i) - y_i)^2$
2,52	20,80	20,30	0,25	20,91	0,01
2,65	20,10	19,36	0,56	19,99	0,01
2,74	18,90	18,70	0,04	19,35	0,20
2,80	19,00	18,26	0,55	18,92	0,01
2,71	19,60	18,92	0,47	19,56	0,00
<b>Summe der quadratischen Abweichungen:</b>			<b>1,86</b>		<b>0,23</b>

- b. Funktion  $g$  sagt die 5er-Hop-Leistung besser voraus, da hier die Summe der quadratischen Abweichungen kleiner ist als bei Funktion  $f$ .