4.1.2 Lineare Regression (Aufgaben)

x	–2	–1	3	4	6
y	0	0.5	2	2	5

1. Gegeben sind die nebenstehenden Werte:
   a) Berechne „von Hand“ die Gleichung der Regressionsgeraden.
   b) Erstelle ein Streudiagramm und zeichne die Regressionsgerade sowie die Residuen ein.

2. Bei einer zufällig ausgewählten Gruppe von Zuschauern an einem Basketballspiel kamen die nebenstehenden Messungen zustande. (sh. auch „Beschreibende Statistik“)
   a) Erstelle ein Streudiagramm mit der Körpergrösse in x-Richtung und dem Gewicht in y-Richtung.
   b) Berechne die Gleichung der Regressionsgeraden und zeichne die Gerade im Streudiagramm ein.
   c) Gib das Bestimmtheitsmass an und interpretiere es kurz.
   d) Mache Voraussagen für die folgenden Personen und kommentiere sie kurz:
   Person A: Grösse = 1.77 m
   Person B: Grösse = 1.35 m
   Person C: Gewicht = 91 kg

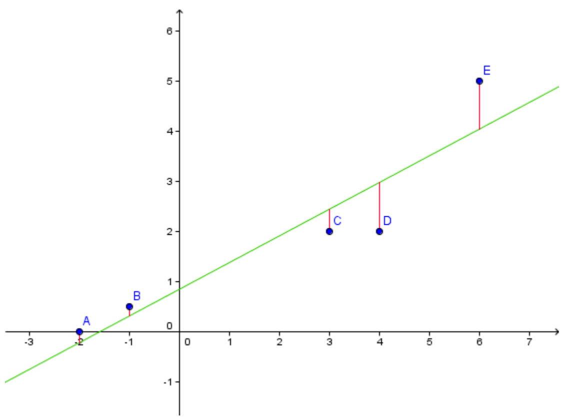
Körpergrösse	Gewicht
1.55	51
1.57	50
1.62	55
1.68	52
1.75	60
1.75	68
1.81	78
1.83	91
1.87	84
1.89	81
1.9	90
1.92	105
1.95	95
1.95	99
1.99	100
2.02	101

Lösungen

1. a)

$$n = 5$$
$$\bar{x} = \tfrac{1}{5}(-2-1+3+4+6) = 2$$
$$\bar{y} = \tfrac{1}{5}(0+0.5+2+2+5) = 1.9$$
$$m = \frac{-2\cdot 0 - 1\cdot 0.5 + 3\cdot 2 + 4\cdot 2 + 6\cdot 5 - 5\cdot 2\cdot 1.9}{(-2)^2 + (-1)^2 + 3^2 + 4^2 + 6^2 - 5\cdot 2^2} = \underline{0.533}$$
$$q = 1.9 - 0.533\cdot 2 = \underline{0.835}$$
$$\Rightarrow \quad \underline{\underline{g_{reg}: \ \hat{y} = 0.533\cdot x + 0.835}}$$

b)



2.

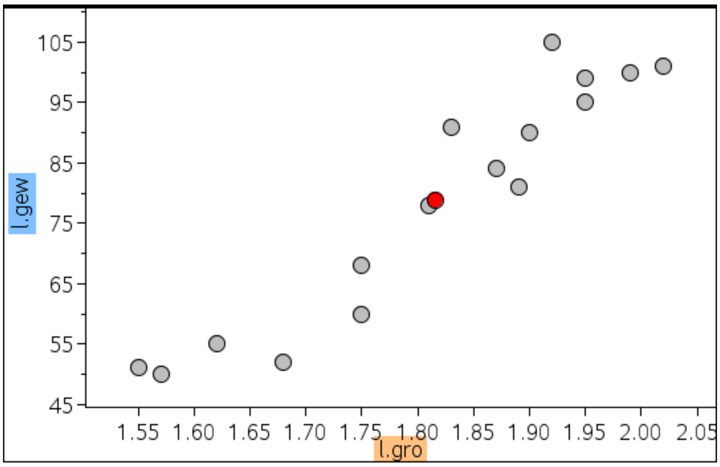
{1.68,1.75,1.75,1.81,1.83,1.87,1.89,1.9,1.92,1.95,1.95,1.99,2.02} → *l.gro*

{1.55,1.57,1.62,1.68,1.75,1.75,1.81,1.83,1.87,1.89,1.9,1.92,1.95,1.95,1.

{51,50,55,52,60,68,78,91,84,81,90,105,95,99,100,101} → *l.gew*

{51,50,55,52,60,68,78,91,84,81,90,105,95,99,100,101}

a)

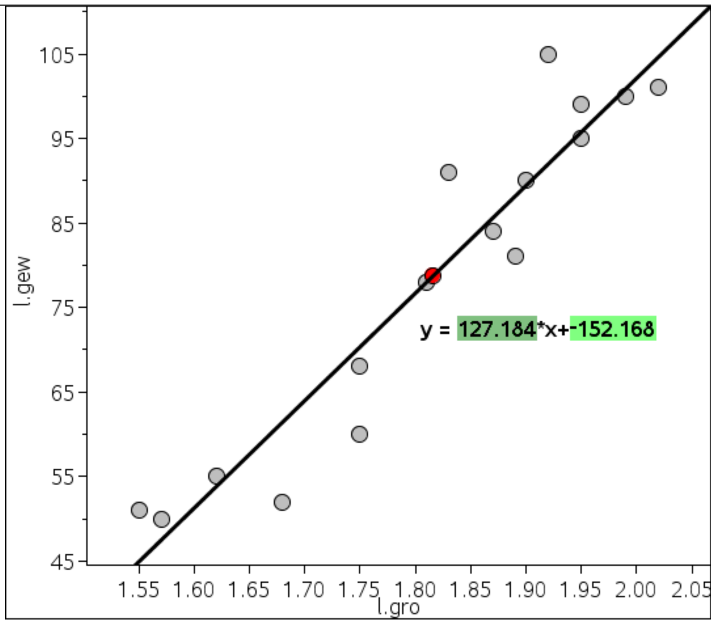


b)

$$\frac{\mathrm{sum}(l.gro\cdot l.gew)-\mathrm{dim}(l.gro)\cdot \mathrm{mean}(l.gro)\cdot \mathrm{mean}(l.gew)}{\mathrm{sum}(l.gro^2)-\mathrm{dim}(l.gro)\cdot(\mathrm{mean}(l.gro))^2}\to m.reg$$

127.184

$\mathrm{mean}(l.gew)-m.reg\cdot \mathrm{mean}(l.gro)\to q.reg$ -152.168



Bemerkung: Die Regressionsgerade geht genau durch den Schwerpunkt (rot) aller Mespunkte, dessen Koordinaten die entsprechenden Mittelwerte sind.

c)

$$\left(\frac{\frac{1}{\mathrm{dim}(l.gew)-1}\cdot \mathrm{sum}((l.gew-\mathrm{mean}(l.gew))\cdot(l.gro-\mathrm{mean}(l.gro)))}{s.gro\cdot s.gew}\right)^2\to reg.r2$$

.898

Unser Modell erklärt die gemessenen Datenpunkte also zu knapp 90%, was sehr gut ist. Die restlichen 10% müssen nach unserem Modell zufällige Schwankungen sein.

d)

$m.reg \cdot 1.77 + q.reg$	72.947
$m.reg \cdot 1.35 + q.reg$	19.530
$\dfrac{91 - q.reg}{m.reg}$	1.912

73 kg für Person A ist eine vernünftige Voraussage.
19.5 kg hingegen für Person B ist sinnlos (das würde einem BMI von 10.7 entsprechen.).
Eine Grösse von 1.91 m für Person C ist zulässig. Wir sehen zwar, dass in der Rohdatenliste tatsächlich eine Person mit 91 kg vorkommt und nur 1.83 m gross ist, aber das ist mit unserem Bestimmtheitsmass vereinbar.

Die oben zusammengestellten Berechnungen können alle mit einem direkt eingebauten Befehl ausgeführt werden:

LinRegMx *l.gro*,*l.gew*,1: CopyVar *stat.RegEqn*,*f1*: *stat.results*

$$\begin{bmatrix} \text{"Title"} & \text{"Linear Regression (mx+b)"} \\ \text{"RegEqn"} & \text{"m*x+b"} \\ \text{"m"} & 127.184 \\ \text{"b"} & -152.168 \\ \text{"r}^2\text{"} & .898 \\ \text{"r"} & .948 \\ \text{"Resid"} & \text{"\{...\}"} \end{bmatrix}$$

$f1(1.77)$	72.947
$f1(1.35)$	19.530
$\text{solve}(f1(groesse)=91, groesse)$	$groesse=1.912$