Statistik für Informatik
Übungsblatt 3 - WiSe2017/18

Lex Winandy

11. In einem Gebiet werden an verschiedenen Tagen die Tageshöchsttemperatur und die Anzahl der Autounfälle ermittelt.

Temperatur(t)	0.5	19.5	-5.1	-8.2	32.4	11.2
Unfälle(u)	55	27	113	82	36	31

(a) Berechne die Kovarianz, das Bestimmtheitsmaß und den Korrelationskoeffizienten. Interpretiere das Ergebnis.

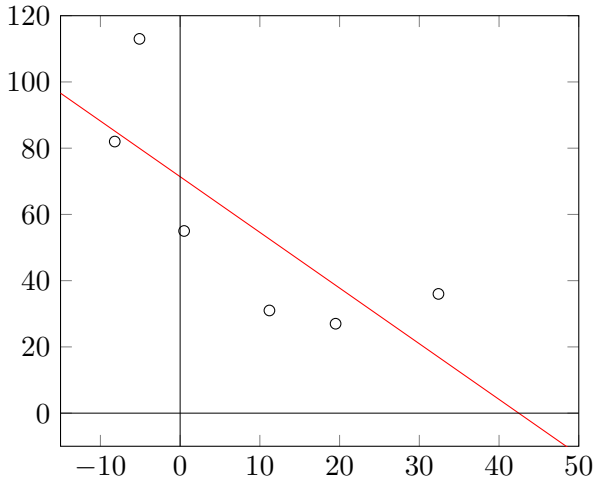
(b) Zeichne einen Scatterplot, berechne die Regressionsgerade und zeichne diese dort ein.

(a)

$$
\begin{aligned}
\bar{t} &= \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{0.5 + 19.5 - 5.1 - 8.2 + 32.4 + 11.2}{6} = \frac{50.3}{6} = 8.38\bar{3} \\
\bar{u} &= \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{55 + 27 + 113 + 82 + 36 + 31}{6} = \frac{344}{6} = 57.\bar{3} \\
s_{t,u} &= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} t_i u_i\right) - n\bar{t}\bar{u}\right) \\
&= \frac{0.5 \cdot 55 + 19.5 \cdot 27 + \ldots + 11.2 \cdot 31 - 6 \cdot 8.38\bar{3} \cdot 57.\bar{3}}{5} = \frac{-2063.1754}{5} = -412.635 \\
s_t^2 &= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} t_i^2\right) - n\bar{t}^2\right) \\
&= \frac{0.5^2 + 19.5^2 + 5.1^2 + 8.2^2 + 32.4^2 + 11.2^2 - 6 \cdot 8.38\bar{3}^2}{5} = \frac{1227.3}{5} = 245.46 \\
s_t &= \sqrt{s_t^2} = \sqrt{245.46} = 15.667 \\
s_u^2 &= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} u_i^2\right) - n\bar{u}^2\right) \\
&= \frac{55^2 + 27^2 + 113^2 + 82^2 + 36^2 + 31^2 - 6 \cdot 57.\bar{3}^2}{5} = \frac{5804.26}{5} = 1160.852 \\
s_u &= \sqrt{s_u^2} = \sqrt{1160.852} = 34.071 \\
r_{t,u} &= \frac{s_{t,u}}{s_t \cdot s_u} = \frac{-412.635}{15.667 \cdot 34.071} = -0.773 \\
r_{t,u}^2 &= (-0.773)^2 = 0.5976
\end{aligned}
\tag{1}
$$

(b) $a = \frac{s_{t,u}}{s_t^2} = \frac{-412.635}{245.46} = -1.681$

$b = \bar{u} - a \cdot \bar{t} = 57.3 - (-1.681) \cdot 8.383 = 71.3918$



12. Bei hundert Studenten wird gezählt, wieviele Mathematiknote x und Informatiknote y haben. Gib die vollständige Kontingenztabelle (inklusive Randhäufigkeiten)

x\y	1	2	3	4	5
1	7	5	2	1	
2	8	7	4	1	
3		8	9	5	1
4		1	9	7	4
5			2	10	9

mit absoluten sowie relativen Häufigkeiten an. Berechne $\bar{x}$, $\bar{y}$, s_x , s_y , $s_{x,y}$, $r_{x,y}$ und interpretiere die Ergebnisse.

x\y	1	2	3	4	5	$\sum$
1	7	5	2	1		15
2	8	7	4	1		20
3		8	9	5	1	23
4		1	9	7	4	21
5			2	10	9	21
$\sum$	15	21	26	24	14	100

x\y	1	2	3	4	5	$\sum$
1	0.07	0.05	0.02	0.01		0.15
2	0.08	0.07	0.04	0.01		0.2
3		0.08	0.09	0.05	0.01	0.23
4		0.01	0.09	0.07	0.04	0.21
5			0.02	0.1	0.09	0.21
$\sum$	0.15	0.21	0.26	0.24	0.14	1

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{15+40+69+84+105}{100} = \frac{313}{100} = 3.13$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{15+42+78+96+70}{100} = \frac{301}{100} = 3.01$$

$$s_{x,y} = \frac{1}{n-1}\left(\left(\sum_{i=1}^{l}\sum_{j=1}^{m} H_{i,j}x_i y_i\right) - n\bar{x}\bar{y}\right)$$

$$= \frac{7+10+...+225-100\cdot 3.13\cdot 3.01}{99} = \frac{127.87}{99} = 1.292$$

$$s_x^2 = \frac{15\cdot 1^2+20\cdot 2^2+23\cdot 3^2+21\cdot 4^2+21\cdot 5^2-100\cdot 3.13^2}{99} = \frac{183.31}{99} = 1.852$$

$$s_x = \sqrt{s_x^2} = \sqrt{1.852} = 1.361$$

$$s_y^2 = \frac{15\cdot 1^2+21\cdot 2^2+26\cdot 3^2+24\cdot 4^2+14\cdot 5^2-100\cdot 3.01^2}{99} = \frac{148.91}{99} = 1.5041$$

$$s_y = \sqrt{s_y^2} = \sqrt{1.5041} = 1.2264$$

$$r_{x,y} = \frac{s_{x,y}}{s_x\cdot s_y} = \frac{1.292}{1.361\cdot 1.2264} = 0.7741$$

13. In einem Ort werden über das Jahr folgende Temperaturen gemessen: Approx-

Tag d	30	105	175	261	338
Temperatur T	-4	10	25	15	1

imiere diese Daten mittels linearer Regression und dem Modell

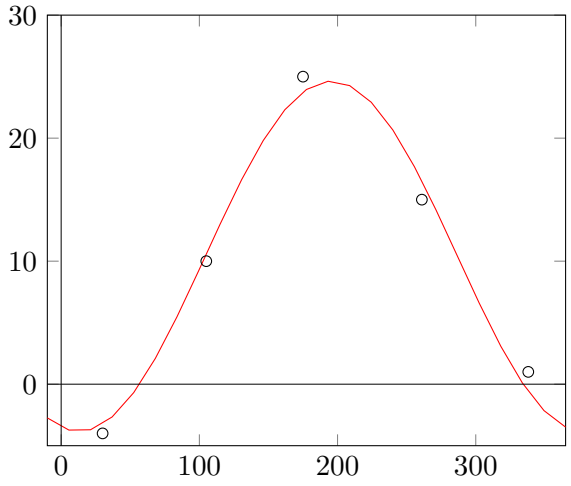
$$T = a + b\cos\left(\frac{2\pi d}{365}\right) + c\sin\left(\frac{2\pi d}{365}\right)$$

$f_{(1)} = 1$, $f_{(2)} = \cos\frac{2\pi}{365}d$, $f_{(3)} = \sin\frac{2\pi}{365}d$

$$\begin{pmatrix} n & \sum_{i=1}^n f_{(2)}x_i & \sum_{i=1}^n f_{(3)}x_i \\ \sum_{i=1}^n f_{(2)}x_i & \sum_{i=1}^n \left(f_{(2)}x_i\right)^2 & \sum_{i=1}^n \left(f_{(2)}x_i \cdot f_{(3)}x_i\right) \\ \sum_{i=1}^n f_{(3)}x_i & \sum_{i=1}^n \left(f_{(2)}x_i \cdot f_{(3)}x_i\right) & \sum_{i=1}^n \left(f_{(3)}x_i\right)^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n \left(f_{(2)}x_i \cdot y_i\right) \\ \sum_{i=1}^n \left(f_{(3)}x_i \cdot y_i\right) \end{pmatrix}$$

$$\begin{pmatrix} 5 & 0.3196 & 0.17 \\ 0.3196 & 2.641 & -0.1144 \\ 0.17 & -0.1144 & 2.3589 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 47 \\ -32.9871 \\ -4.1236 \end{pmatrix}$$

$\Rightarrow a = 10.3954$, $b = -13.8857$, $c = -3.1707$
$T = 10.3954 - 13.8857 \cdot \cos\left(\frac{2\pi d}{365}\right) - 3.1707 \cdot \sin\left(\frac{2\pi d}{365}\right)$

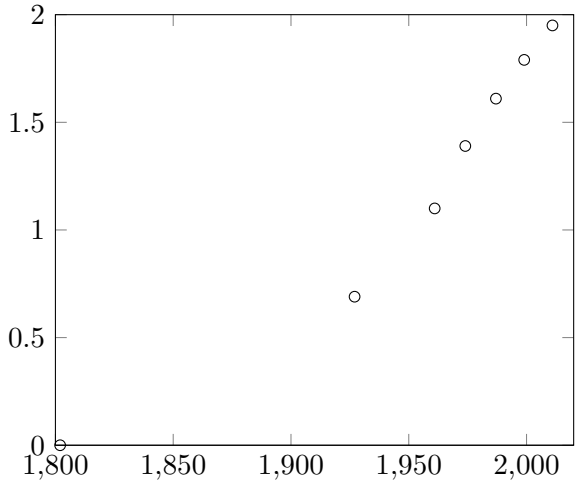


14. Die Erdbevölkerung hatte in folgenden Jahren folgende Ausmaße erreicht: Model-

Jahr x_i	1802	1927	1961	1974	1987	1999	2011
Mrd. Ew. y_i	1	2	3	4	5	6	7

liere das Wachstum mit $y \approx f(x) = b \cdot a^x$ (exponentielles Wachstum). Um a und b zu finden, benötigt man eigentlich Methoden der nichtlinearen Regression, da $f(x)$ in a nicht linear ist. Durch Logarithmieren kann es aber auf lineare Regression zurückgeführt werden. Zeichne einen Scatterplot von $\log y$ über x, berechne die Regressionsgerade und rechne daraus a und b aus. Zeichne dann den Scatterplot von y über x und die Regressionskurve.

Jahr x_i	1802	1927	1961	1974	1987	1999	2011
Mrd. Ew. y_i	1	2	3	4	5	6	7
$\ln(y_i)$	0	0.69	1.1	1.39	1.61	1.79	1.95



$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1802 + 1927 + 1961 + 1974 + 1987 + 1999 + 2011}{7} = \frac{13661}{7} = 1951.6$$

$$\ln(\bar{y}) = \frac{1}{n}\sum_{i=1}^{n} \ln(y_i) = \frac{0 + 0.69 + 1.1 + 1.39 + 1.61 + 1.79 + 1.95}{7} = \frac{8.53}{7} = 1.22$$

$$s_{x,\ln(y)} = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i \ln(y_i)\right) - n \cdot \bar{x} \cdot \overline{\ln(y)}\right)$$

$$= \frac{1802 \cdot 0 + 1927 \cdot 0.69 + ... + 1999 \cdot 1.79 + 2011 \cdot 1.95 - 7 \cdot 1951.6 \cdot 1.22}{6} = \frac{262.656}{6} = 43.78$$

$$s_x^2 = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2\right)$$

$$= \frac{1802^2 + 1927^2 + 1961^2 + 1974^2 + 1987^2 + 1999^2 + 2011^2 - 7 \cdot 1951.6^2}{6} = \frac{29823.08}{6} = 4970.51$$

$$\ln(y) = ax + b$$

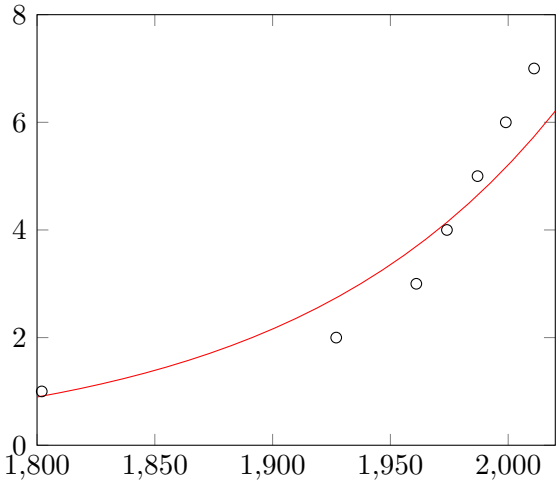
$$\Rightarrow a = \frac{s_{x,\ln(y)}}{s_x^2} = \frac{43.78}{4970.51} = 0.0088$$

$$\Rightarrow b = \overline{\ln(y)} - a\bar{x} = 1.22 - 0.0088 \cdot 1951.6 = -15.9488$$

$$y = e^{-15.95 + 0.0088x} = e^{-15.95} \cdot \left(e^{0.0088}\right)^x$$

$$\Rightarrow b = 5.2$$

$$\Rightarrow a = 1.009$$


8
6
4
2
0
1,800
1,850
1,900
1,950
2,000