Kapitel VI

Korrelations- und Regressionsanalyse)

(Lösungen)

6. 1.

Jahre	y_i	x_i	x_i^2	$x_i \cdot y_i$	y_i^2
1997	8	-3	9	-24	64
1998	13	-2	4	-26	169
1999	15	-1	1	-15	225
2000	17	0	0	0	289
2001	18	1	1	18	324
2002	20	2	4	40	400
2003	21	3	9	63	441
Σ	112	0	28	56	1912

1.

$$7a_0 = 112 \quad \Rightarrow \quad a_0 = 16$$

$$28a_1 = 56 \quad \Rightarrow \quad a_1 = 2$$

$$y^* = 16 + 2x$$

2.

$$B_{yx} = \frac{(7 \cdot 56)^2}{(7 \cdot 28) \cdot (7 \cdot 1912 - 112^2)} = \frac{153664}{164640} \approx 0.93$$

Mit dem Bestimmtheitsmaß von 0.93 handelt es sich um eine sehr gute Funktion.

3.

$$y^*_{2004} = y^*(4) = 24 \text{ Mio. €}$$

6. 2.

Arbeitstabelle

y_i	x_i	y_i^2	x_i^2	$x_i \cdot y_i$
2	1.8	4	3.24	3.6
3	1.9	9	3.61	5.7
3	2.2	9	4.84	6.6
4	2.2	16	4.84	8.8
5	2.4	25	5.76	12.0
5	2.5	25	6.25	12.5
6	2.7	36	7.29	16.2
7	3.0.	49	9.00	21.0
8	3.2	64	10.24	25.6
9	3.3	81	10.89	29.7
10	3.6	100	12.96	36.0
62	28.8	418	78.92	177.7

a)

$$\begin{aligned} 11a_0 + 28.8\ a_1 &= 62 \\ 28a_0 + 78.92a_1 &= 177.7 \end{aligned} \qquad \Rightarrow \qquad a_0 = -5.722, \qquad a_1 = 4.37$$

$$\begin{aligned} 11a_0 + 62\ a_1 &= 28.8 \\ 62a_0 + 418a_1 &= 177.7 \end{aligned} \qquad \Rightarrow \qquad a_0 = 1.36, \qquad a_1 = 0.22$$

Damit lauten die beiden Regressionsfunktionen

$$y^* = -5.722 + 4.37x$$

und

$$x^* = 1.36 + 0.22y$$

b)

Die erste Gleichung gibt die Abhängigkeit des Alters von den Reparaturkosten an. Steigen die Reparaturkosten um 1 Einheit. so steigt das Alter um 4.37 Einheiten; Die zweite Gleichung gibt die Abhängigkeit der Reparaturkosten vom Alter der Lkw an. Steigt das Alter um eine Einheit. so steigen die Kosten um 0.22 Einheiten an. Die Kosten. die regelmäßig anfallen. betragen 1.36 Einheiten.

c)

(1)

$$R_{yx} = \frac{11 \cdot 177.7 - 28.8 \cdot 62}{\sqrt{(11 \cdot 78.92 - 28.8^2) \cdot (11 \cdot 418 - 62^2)}} = \frac{169.1}{170.776813} = 0.99018126$$

$$R_{xy} = \frac{11 \cdot 177.7 - 62 \cdot 28.8}{\sqrt{(11 \cdot 418 - 62^2) \cdot (11 \cdot 78.92 - 28.8^2)}} = \frac{169.1}{170.776813} = 0.99018126$$

(2)

Der Korrelationskoeffizient wird hier als geometrisches Mittel der Anstiege der beiden Regressionskoeffizienten berechnet:

$$R = \sqrt{4.37 \cdot 0.22} = 0.98\,.$$

d)

$$B = R^2 = 0.98^2 = 0.96$$

(Die Diskrepanz zwischen $B_{xy} = B_{yx} = 0.98045892$ und $B = 0.96$ ist auf Rundungsfehler zurückzuführen.)

Das Bestimmtheitsmaß besagt. dass 96% der Variation des Alters zu den Reparaturkosten durch die Regressionsfunktion $y^* = -5.722 + 4.37x$ erklärt werden kann.

6. 3.

a)

x_i	y_i	x_i^2	$x_i y_i$
34	35	1156	1190
37	26	1369	962
45	48	2025	2160
48	59	2304	2832
51	100	2601	5100
215	268	9455	12244

$$5a_0 + 215a_1 = 268$$
$$215a_0 + 9455a_1 = 12244$$

$$a_0 = -93.8286, \qquad a_1 = 3.4286$$
$$y^* = -93.8286 + 3.4286x\ .$$

b)

$$y^*(40) = 43.3133 \text{ Mio. €}$$

6. 4

Arbeitstabelle

x_i	y_i	x_i^2	$x_i \cdot y_i$
50	-10	2500	-500
60	-10	3600	-600
60	- 5	3600	-300
70	- 1	4900	- 70
100	-0.5	10000	- 50
340	-26.5	24600	-1520

a)

$$\begin{aligned} 5\ \ a_0 + \ 340\ a_1 &= -26.5 \\ 340a_0 + 24600a_1 &= -1520 \end{aligned} \quad \Rightarrow \quad a_0 = -18.2568, \quad a_1 = 0.1905$$

Die Regressionsfunktion lautet damit

$$y^* = -18.2568 + 0.1905x$$

b)

$$y^* = -18.2568 + 0.1905x := 0 \qquad \Rightarrow \qquad x \approx 95.84$$

Der Widerspruch ist damit zu begründen. dass die geschätzte Regressionsgerade nur eine Annäherung für die Daten ist.

6. 5

1)

$$\begin{aligned}
\sum_{i=1}^{n} y_i &= a_0 n + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 + a_3 \sum_{i=1}^{n} x_i^3 \\
\sum_{i=1}^{n} y_i \cdot x_i &= a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 + a_3 \sum_{i=1}^{n} x_i^4 \\
\sum_{i=1}^{n} y_i \cdot x_i^2 &= a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4 + a_3 \sum_{i=1}^{n} x_i^5 \\
\sum_{i=1}^{n} y_i \cdot x_i^3 &= a_0 \sum_{i=1}^{n} x_i^3 + a_1 \sum_{i=1}^{n} x_i^4 + a_2 \sum_{i=1}^{n} x_i^5 + a_3 \sum_{i=1}^{n} x_i^6
\end{aligned}$$

2)

$$\begin{aligned}
\sum_{i=1}^{n} \log y_i &= n \cdot a_0 + a_1 \sum_{i=1}^{n} \log x_i \\
\sum_{i=1}^{n} \log x_i \cdot \log y_i &= a_0 \sum_{i=1}^{n} \log x_i + a_1 \sum_{i=1}^{n} \left(\log x_i\right)^2
\end{aligned}$$

3)

$$\begin{aligned}
\sum_{i=1}^{n} y_i &= n \cdot a_0 + a_1 \sum_{i=1}^{n} \frac{1}{x_i^2} \\
\sum_{i=1}^{n} \frac{y_i}{x_i^2} &= a_0 \sum_{i=1}^{n} \frac{1}{x_i^2} + a_1 \sum_{i=1}^{n} \frac{1}{x_i^4}
\end{aligned}$$

4)

$$\begin{aligned}
\sum_{i=1}^{n} y_i &= a_0 + a_1 \sum_{i=1}^{n} x_i^2 \\
\sum_{i=1}^{n} x_i^2 \cdot y_i &= a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^4
\end{aligned}$$

5)

$$y^* = \frac{1}{a_0 + a_1 \cdot x + a_2 \cdot x^2}\,.$$

$$\sum_{i=1}^{n} \frac{1}{y_i} = a_0 n + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2$$
$$\sum_{i=1}^{n} \frac{x_i}{y_i} = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3$$
$$\sum_{i=1}^{n} \frac{x_i^2}{y_i} = a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4$$

6. 6.

1)

$$\frac{1}{y} = a + bx$$

2)

$$\log y = \log a + bx \log e$$

6. 7.

Arbeitstabelle

x_i	y_i	$1/x_i$	$1/x_i^{\,2}$	y_i/x_i
2	2.88	0.50000	0.25000	1.44000
3	2.22	0.33333	0.11111	0.74000
5	2.13	0.20000	0.04000	0.42600
8	2.06	0.12500	0.01563	0.25750
9	2.02	0.11111	0.01235	0.22444
10	2.03	0.10000	0.01000	0.20300
Σ	13.34	1.36944	0.43908179	3.29094

$$\begin{array}{l} 6\,a_0 + 1.36944\,a_1 = 13.34 \\ 1.36944 a_0 + 0.43908 a_1 = 3.29094 \end{array} \quad \Rightarrow \quad a_0 = 1.78, \quad a_1 = 1.93$$

Damit lautet die Regressionsfunktion:

$$y^* = 1.78 + 1.93 \frac{1}{x}\,.$$

6. 8.

$$y^* = \frac{a}{1 + e^{b-cx}}$$

Es gilt:

$$\frac{1}{y^*} = \frac{1+e^{b-cx}}{a}. \qquad \frac{a}{y^*} = 1+e^{b-cx}. \qquad \frac{a}{y^*} - 1 = e^{b-cx}$$

$$\ln\left(\frac{a}{y^*} - 1\right) = b - cx.$$

Sei

$$Y := \ln\left(\frac{a}{y^*} - 1\right). \qquad B := b. \qquad C := -c.$$

dann hat man

$$Y^* = B + Cx$$

Die Normalgleichungen lauten dann:

$$nB + C\sum_i x_i = \sum_i Y_i$$
$$B\sum_i x_i + C\sum_i x_i^2 = \sum_i x_i Y_i$$

Arbeitstabelle

x_i	y_i	Y_i	x_i^2	$x_i Y_i$
0.5	11.40	7.46930	0.25	3.73465
1.0	13.80	7.27813	1.00	7.27813
1.5	15.72	7.14777	2.25	10.72165
2.0	17.16	7.06005	4.00	14.12010
2.5	18.20	7.00116	6.25	17.50289
3.0	18.85	6.96603	9.00	20.89810
3.5	19.28	6.94345	12.25	24.30209
4.0	19.50	6.93210	16.00	27.72839
18.0		56.79799	51.00	126.28599

$$\begin{array}{l} 8B + 18C = 56.79799 \\ 18B + 51C = 126.28599 \end{array} \Rightarrow B = 7.4232103 \approx 7.42. \qquad C \approx -0.14376071 \approx -0.14.$$

d.h. $b \approx 7.42, \qquad c \approx 0.14.$

also

$$y^* = \frac{20000}{1+e^{7.42-0.14x}}$$

Das Bestimmtheitsmaß ist in der Regel nicht aussagekräftig.

6. 9.

1.

$$x = a_0 \cdot r^{a_1}$$

$$\lg x = \lg a_0 + a_1 \cdot \lg r$$

Sei

$$X := \lg x . \qquad A_0 := \lg a_0 . \qquad A_1 := a_1 . \qquad R := \lg r .$$

Das Bestimmungsgleichungssystem lautet:

$$n \cdot A_0 + A_1 \cdot \sum_i \lg r_i = \sum_i \lg x_i$$
$$A_0 \cdot \sum_i \lg r_i + A_1 \cdot \sum_i (\lg r_i)^2 = \sum_i \lg r_i \cdot \lg x_i$$

Arbeitstabelle

r_i	x_i	$\lg r_i$	$\lg x_i$	$(\lg r_i)^2$	$\lg x_i \cdot \lg r_i$
5.0	1.54	0.69897000	0.18752072	0.488559067	0.13107136
5.5	1.64	0.74036269	0.21484385	0.548136912	0.15906237
6.0	1.70	0.77815125	0.23044892	0.605519368	0.17932412
6.5	1.80	0.81291336	0.25527251	0.660828125	0.20751443
7.0	1.86	0.84509804	0.26951294	0.714190697	0.22776486
7.5	1.90	0.87506126	0.2787536	0.765732215	0.24392648
8.0	1.98	0.90308999	0.29666519	0.815571525	0.26791536
		5.65364659	1.73301773	4.598537909	1.41657898

$$7A_0 \quad + 5.65A_1 = 1.73$$
$$5.65A_0 + 4.60A_1 = 1.42$$

$$A_0 = -0.234 \quad \Rightarrow \quad a_0 = 0.583$$
$$A_1 = 0.596 \quad \Rightarrow \quad a_1 = 0.596$$

$$x = 0.583 \cdot r^{0.596}$$

2.

$$x(8.3) = 0.583 \cdot 1.8^{0.596} \approx 2.06$$

6. 10.

1.

$$y^* = a_0 \cdot a_1^{\,x}$$
$$\lg y^* = \lg a_0 \cdot a_1^{\,x}$$
$$\lg y^* = \lg a_0 + \lg a_1^{\,x}$$
$$\lg y^* = \lg a_0 + x \cdot \lg a_1$$
$$Y^* = A_0 + A_1 x$$

mit

$$Y^* := \lg y^*. \qquad A_0 := \lg a_0. \qquad A_1 := \lg a_1.$$

Das Bestimmungsgleichungssystem lautet:

$$n \cdot A_0 + A_1 \cdot \sum_i x_i = \sum_i Y_i$$
$$A_0 \cdot \sum_i x_i + A_1 \cdot \sum_i x_i^2 = \sum_i x_i \cdot Y_i .$$

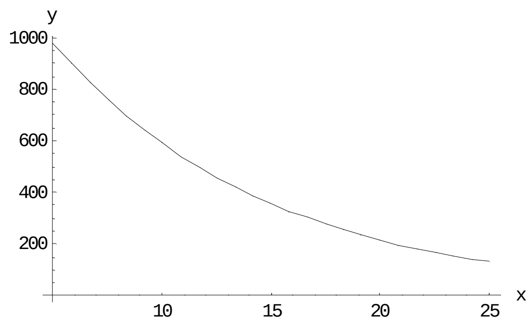
Arbeitstabelle

x_i	y_i	x_i^2	$\lg x_i$	$x_i \cdot \lg x_i$
20	220	400	2.34242268	46.8484536
18	260	324	2.41497335	43.4695203
15	350	225	2.54406804	38.1610207
12	480	144	2.68124124	32.1748948
10	600	100	2.77815125	27.7815125
75		1193	12.7608566	188.435402

$$5A_0 + 75A_1 = 12.7608566$$
$$75A_0 + 1193A_1 = 188.435402$$

$$A_0 = 3.208961095, \qquad a_0 = 1617.935093$$
$$A_1 = -0.043785985, \qquad a_1 = 0.90409489$$

$$y^* = 1617.935093 \cdot 0.90409489^x$$



2.

$$y^*(9) \approx 653.$$

6. 11.

1.

Jahr	Einwohnerzahl [Tausend]	Absolutes Wachstum	Relatives Wachstum [%]
2001	8000	-	
2002	8200	200	2.50
2003	9000	800	9.76
2004	9600	600	6.70
2005	10200	600	6.25

$$\bar{x}_g = \sqrt[4]{1.025 \cdot 1.0976 \cdot 1.0667 \cdot 1.0625} = 1.062636672 \approx 1.0626.$$

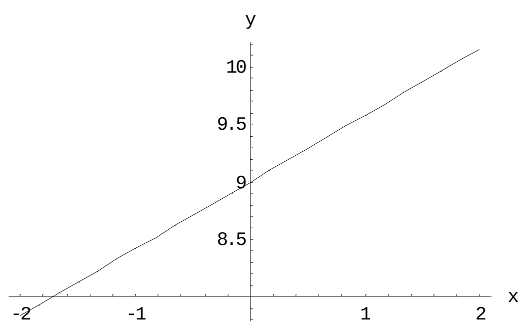
Die durchschnittliche Wachstumsrate beträgt also etwa 6.26%.

2.

x_i	y_i	x_i^2	$x_i y_i$
-2	8.0	4	-16.0
-1	8.2	1	-8.20
0	9.0	0	0.00
1	9.6	1	9.60
2	10.2	4	20.40
0	45.0	10	5.80

$$5a_0 = 45.0 \qquad a_0 = 9$$
$$10a_1 = 5.8 \qquad a_1 = 0.58$$

$$y^* = 9 + 0.58.$$



6. 12.

Arbeitstabelle

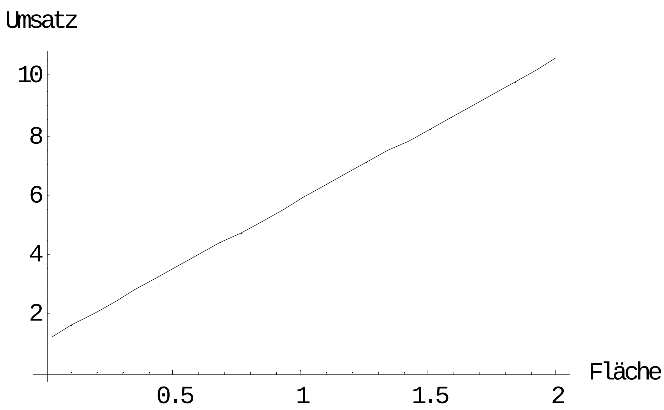
x_i	y_i	x_i^2	$x_i \cdot y_i$	y_i^2
0.31	2.93	0.0961	0.9083	8.5849
0.98	5.27	0.9604	5.1646	27.7729
1.21	6.85	1.4641	8.2885	46.9225
1.29	7.01	1.6641	9.0429	49.1401
1.12	7.02	1.2544	7.8624	49.2804
1.49	8.35	2.2201	12.4415	69.7225
0.78	4.33	0.6084	3.3774	18.7489
0.94	5.77	0.8836	5.4238	33.2929
8.12	47.53	9.1512	52.5094	303.4651

1.

$$\begin{cases} 8a_0 \quad +8.12a_1 = 47.53 \\ 8.12a_0 + 9.1512a_1 = 52.5094 \end{cases} \Rightarrow a_0 = 1.179377612 \approx 1.18, \quad a_1 = 4.69149989 \approx 4.69$$

$$y^* = 1.18 + 4.69x\,.$$

2.



3.

$$r_{yx} = \frac{8 \cdot 52.5094 - 8.12 \cdot 47.53}{\sqrt{\left(8 \cdot 9.15.12 - 8.12^2\right) \cdot \left(8 \cdot 303.4651 - 47.53^2\right)}} = 0.97449534.$$

Die Faktoren Verkaufsfläche und Jahresumsatz bewegen sich in gleicher Richtung.

$$B_{yx} = 0.94964116 \approx 0.95.$$

Der Umsatz ist zu etwa 95% von der Verkaufsfläche abhängig.

6. 13.

Arbeitstabelle

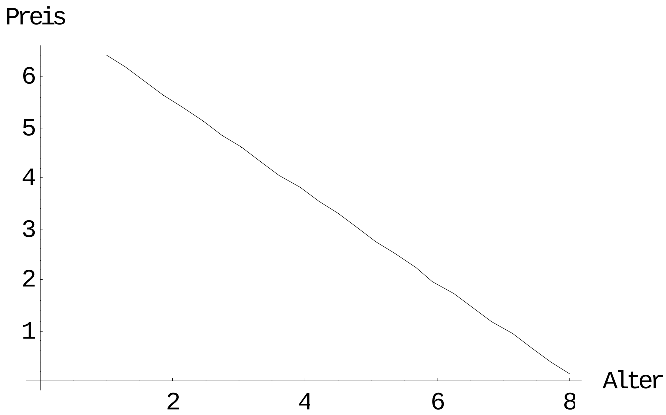
x_i	y_i	x_i^2	$x_i \cdot y_i$	y_i^2
3	4	9	12	16
5	3	25	15	9
3	4	9	12	16
6	1.5	36	9	2.25
1	8	1	8	64
2	4	4	8	16
1	7	1	7	49
7	2	49	14	4
28	33.5	134	85	176.25

1.

$$\begin{cases} 8a_0 + 28a_1 = 33.5 \\ 28a_0 + 134a_1 = 85.0 \end{cases} \Rightarrow a_0 = 7.322916667 \approx 7.32, \quad a_1 = -0.895833333 \approx -0.90 .$$

$$y^* = 7.32 - 0.90x .$$

2.



3.

$$r_{yx} = \frac{8 \cdot 85 - 28 \cdot 33.5}{\sqrt{\left(8 \cdot 134 - 28^2\right) \cdot \left(8 \cdot 126.75 - 33.5^2\right)}} = -0.8962224.$$

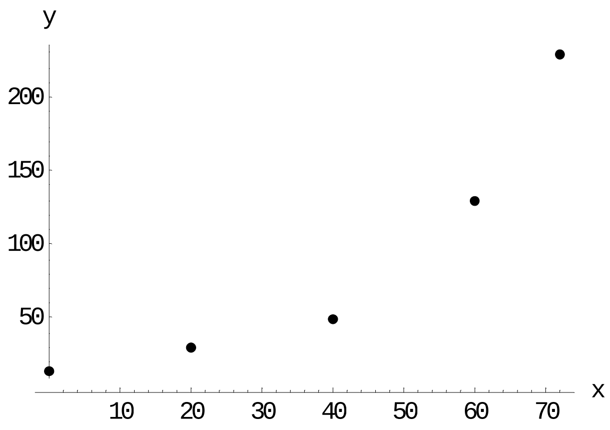
Die Faktoren Verkaufspreis und Alter bewegen sich in entgegengesetzter Richtung.

$$B_{yx} = 0.8032146 \approx 0.80.$$

Der Verkaufspreis ist zu etwa 80% vom Alter der Wagen abhängig.

6. 14.

1.



2.

$$y^* = a_0 \cdot e^{a_1 x}$$

$$\ln y^* = \ln a_0 \cdot e^{a_1 x}. \qquad \ln y^* = \ln a_0 + \ln \cdot e^{a_1 x}. \qquad \ln y^* = \ln a_0 + a_1 x.$$

Sei

$$A_0 := \ln a_0, \quad A_1 := a_1, \quad Y_i := \ln y_i.$$

$$\begin{cases} n \cdot A_0 + A_1 \cdot \sum_{i=1}^{5} x_i = \sum_{i=1}^{5} Y_i \\ A_0 \cdot \sum_{i=1}^{5} x_i + A_1 \cdot \sum_{i=1}^{5} x_i^2 = \sum_{i=1}^{5} x_i \cdot Y_i \end{cases}$$

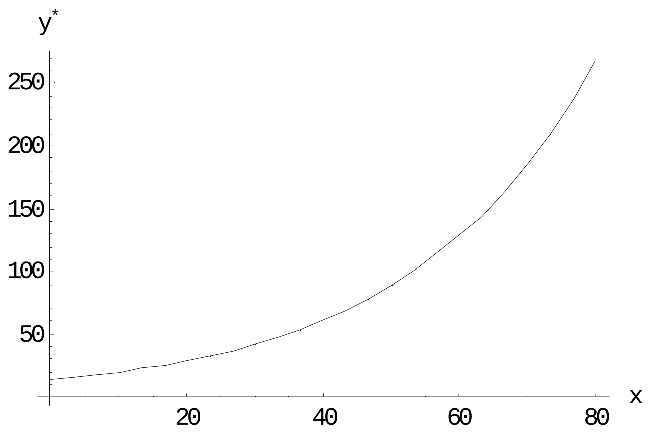
x_i	y_i	Y_i	x_i^2	$x_i \cdot Y_i$
0	15	2.70805020	0	0.00000000
20	30	3.40119738	400	68.0239476
40	50	3.91202301	1600	156.480920
60	130	4.86753445	3600	292.052067
72	230	5.43807931	5184	391.541710
192		20.3268843	10784	908.098645

$$\begin{cases} 5A_0 + 192A_1 = 20.3268843 \\ 192A_0 + 10784A_1 = 908.098645 \end{cases} \Rightarrow A_0 = 2.629583772, \quad A_1 = 0.037390445$$

$$a_0 = 13.86799645 \approx 13.868, \qquad a_1 = 0.037390445 \approx 0.037$$

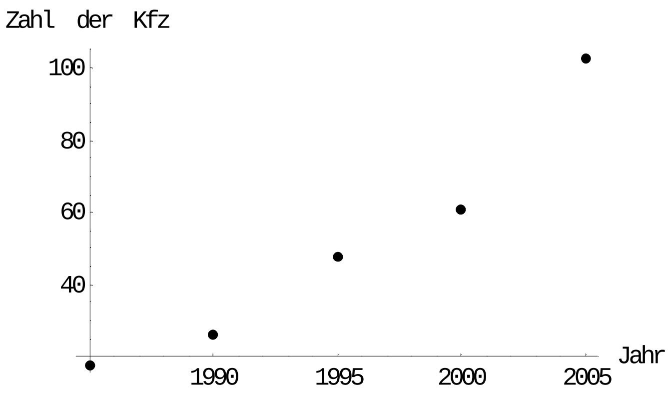
$$y^* = 13.868 \cdot e^{0.037x}$$

3.



6. 15.

1.



2.

$$y^* = a_0 \cdot e^{a_1 x}$$

$$\ln y^* = \ln a_0 \cdot e^{a_1 x}. \qquad \ln y^* = \ln a_0 + \ln \cdot e^{a_1 x}. \qquad \ln y^* = \ln a_0 + a_1 x .$$

Sei

$$A_0 := \ln a_0, \quad A_1 := a_1, \quad Y_i := \ln y_i .$$

$$\begin{cases} n \cdot A_0 \qquad\quad + A_1 \cdot \sum_{i=1}^{5} x_i = \sum_{i=1}^{5} Y_i \\ A_0 \cdot \sum_{i=1}^{5} x_i + A_1 \cdot \sum_{i=1}^{5} x_i^2 = \sum_{i=1}^{5} x_i \cdot Y_i \end{cases}$$

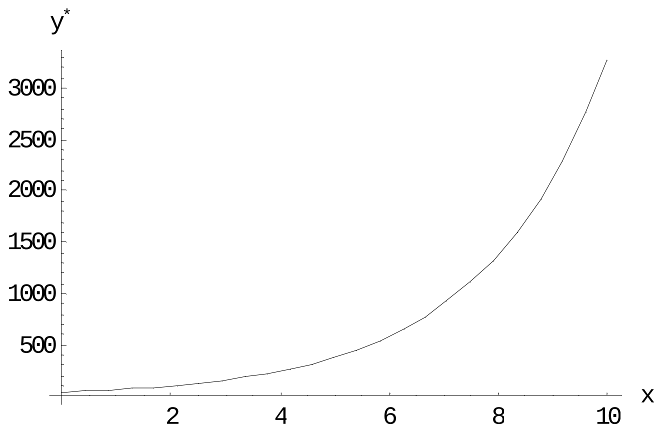
Jahr	y_i	x_i	x_i^2	Y_i	$x_i \cdot Y_i$
1985	18	-2	4	2,89037176	5,78074352
1990	26	-1	1	3,25809654	3,25809654
1995	48	0	0	3,87120101	0.00000000
2000	61	1	1	4,11087386	4,11087386
2005	103	2	4	4,63472899	9,26945798
		0	10	18,7652722	4,34149179

$5A_0 = 18.7652722, \quad A_0 = 3.75305444, \quad a_0 = 42.65115865$

$10A_1 = 4.34149179, \quad A_1 = a_1 = 0.434149179$.

$$y^* = 42.65115865e^{0.434149179x}$$

3.



(Letzte Aktualisierung: 26.02.08)