

Zwei-Variablen-Statistik, lineare, exponentielle, Potenz-, quadratisch Regression weiter unten

Haftendorn 5/08 www.mathematik-verstehen.de

In dieser Datei werden die wichtigsten Möglichkeiten der Zwei-Variablen-Statistik gezeigt.

Die TI-Datei enthält mehrere "Probleme" – hier sind sie mit Namen versehen, sonst mit Nummern.

Diese sind untereinander Variablen-geschützt. So kann hier auf stets fast gleiche Art mit denselben Daten verschiedenes vorgeführt werden.

Will man mit den eigenen Daten experimentieren, kopiert man sich entweder die ganze Datei oder in dieser Datei ein Problem (Problemnamen in der Seitenübersicht markieren, re-Maus Kopieren, re-Maus Einfügen).

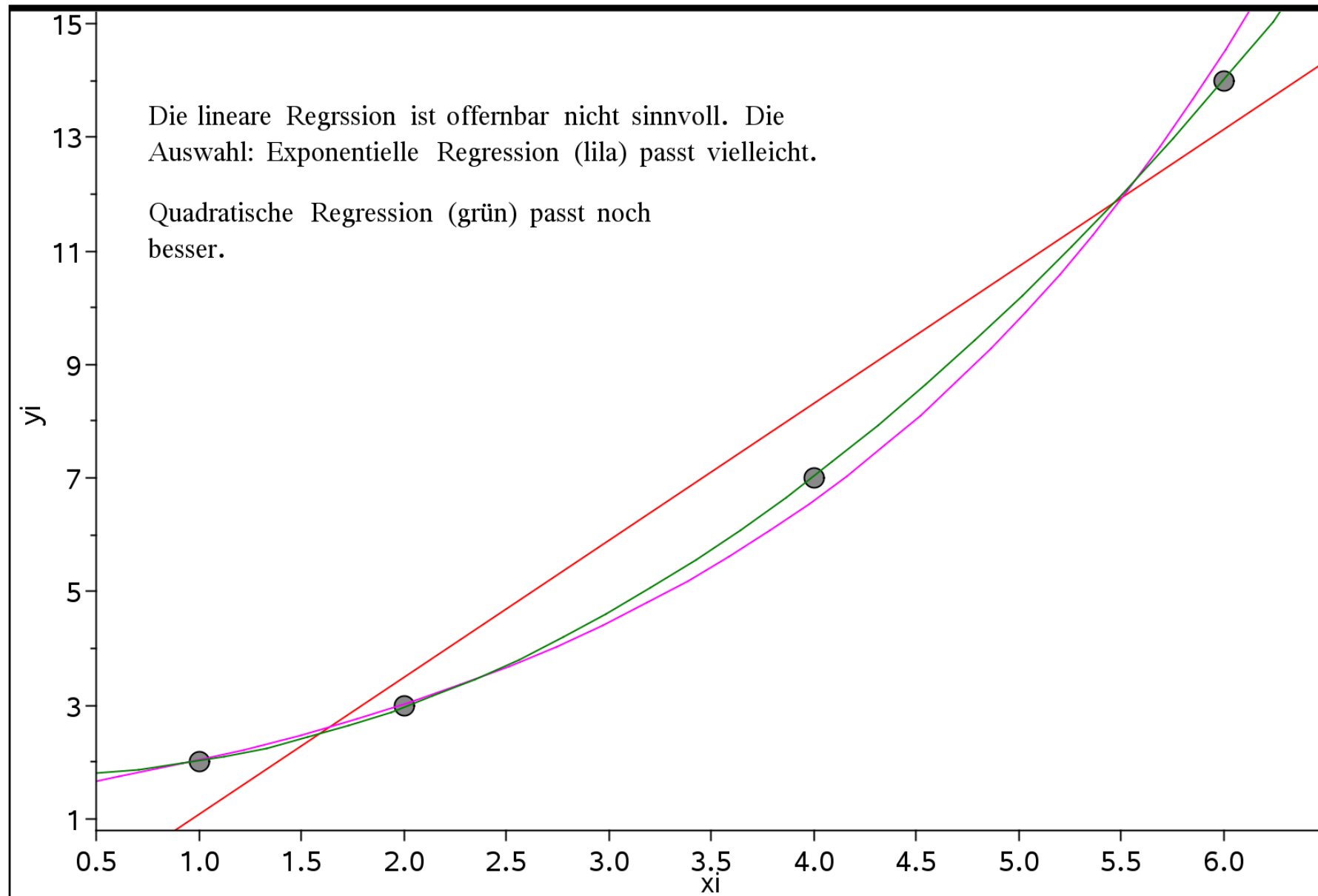
Im ersten Problem hier wird die Arbeit mit der **eingebauten twoVar-Statistik** gezeigt.

In den nachfolgenden Problemen geht es um Regressionen. Dort wird auch der eigene Umgang mit den Tabellenspalte gezeigt.

Die Definition der Daten kann in Mathezellen (strg m) so erfolgen $x:=\{1,2,4,6\}$ oder als Einträge in einem List&Spreadsheet-Fenster. So ist es hier in **Blatt 3** gemacht.

Blatt 2 zeigt das Data&Statistics-Fenster. Unten in der Mitte klickt man, es erscheinen Vorschläge, man wählt x_i , links in der Mitte wählt man y_i und es werden die Datenpunkte gezeichnet. Die Regressionskurven erhält man mit der Werkzeugpalette, Aktionen, Regression,... Das wird in den anderen Problemen ausführlich gemacht. **Blatt 3 rechts, Blatt 4 und Blatt 5** zeigen, wie man mit dem eingebauten Befehl `twovar`

umgeht. Katalog, `twovar` suchen, Assistenten einschalten, x_i, y_i eintragen.....|



	A _j	B _{xi}	C _{yi}	D	E _{zw}	F	G	H	I	J	K
◆					=twovar(xi,yi,1)						
1	1	1	2	Titel	Statistiken mit zw...						
2	2	2	3	\bar{x}	3.25			MinX	0.9798		
3	3	4	7	Σx	13.			Q ₁ X	1.		
4	4	6	14	Σx^2	57.			MedianX	1.5		
5				$s_x := s_{n-1}x$	2.21736			Q ₃ X	3.		
6				$\sigma_x := \sigma_n x$	1.92029			MaxX	5.		
7				n	4.			MinY	6.		
8				\bar{y}	6.5			Q ₁ Y	2.		
9				Σy	26.			MedianY	2.5		
10				Σy^2	258.			Q ₃ Y	5.		
11				$s_y := s_{n-1}y$	5.44671			MaxY	10.5		
12				$\sigma_y := \sigma_n y$	4.71699			SSX := $\Sigma(x-\bar{x})^2$	14.		
13				Σxy	120.			SSY := $\Sigma(y-\bar{y})^2$	14.75		
14				MinX	0.9798				89.		
15				Q ₁ X	1.						
16				MedianX	1.5						
17				Q ₃ X	3.						
A1		1									

1.3

Die eingebaute Zwei-Variablen-Statistik

Sie ist besonders nützlich, wenn man irgendwo die Zwischenwerte braucht. Wieder ist es günstig, die Eingabemaske des Assistenten zu verwenden.

	"Titel"	"Statistiken mit zwei Variablen"
	" \bar{x} "	3.25
	" ΣX "	13.
	" ΣX^2 "	57.
	" $s_X := s_{n-1}X$ "	2.21736
	" $\sigma_X := \sigma_{nX}$ "	1.92029
	"n"	4.
	" \bar{y} "	6.5
	" Σy "	26.
	" Σy^2 "	258.
	" $s_y := s_{n-1}Y$ "	5.44671
	" $\sigma_y := \sigma_{nY}$ "	4.71699
TwoVar xi,yi,1: stat.results ▶	" Σxy "	120.
	"r"	0.9798
	"MinX"	1.
	"Q ₁ X"	1.5
	"MedianX"	3.
	"Q ₃ X"	5.
	"MaxX"	6.
	"MinY"	2.

An die Einzelwerte kommt man so heran:

In eine Mathezelle **stat.** schreiben, dann erscheint –beim Schreiben des Punktes– ein Pulldownmenu, aus dem man auswählen kann.

stat.MinY ▶ 2. **stat.Q₁Y** ▶ 2.5 **stat.MedianY** ▶ 5. **stat.Q₃Y** ▶ 10.5 **stat.MaxY** ▶ 14.
stat.SSX ▶ 14.75 **stat.SSY** ▶ 89.

Übrigens hat man von den beiden Standardabweichungen in der beschreibenden Statistik zu nehmen **stat.σx** ▶ 1.92029 und **stat.σy** ▶ 4.71699

Dagegen hat man in der beurteilenden Statistik (induktiven Statistik) mit **stat.sx** ▶ 2.21736 und **stat.sy** ▶ 5.44671 bessere (nämlich erwartungstreue) Schätzer für die "wahren" Standardabweichungen der Grundgesamtheit.

Lineare Regression

Lineare Regression, Exponentiell, Potenz, Quadratisch weiter unten

Haftendorn 5/08 www.mathematik-verstehen.de

Im 1. Problem wird zu gegebenen Daten die Regressionsgerade $y = \mathbf{m} \cdot x + \mathbf{b}$ berechnet.

Das geschieht auf verschiedene Arten. *Achtung: nach dem Kopieren so eines Spreadsheets muss man in jeder Spalte eine Zahl neu "abschicken".*

Blatt 2. definiert die Daten und rechnet nach der Formel der Formelsammlung

$$\mathbf{m} = \frac{n \cdot \sum(\mathbf{x}_i \cdot \mathbf{y}_i) - \sum(\mathbf{x}_i) \cdot \sum(\mathbf{y}_i)}{n \cdot \sum(\mathbf{x}_i^2) - (\sum(\mathbf{x}_i))^2}$$

In den gelben Feldern sind Variable definiert worden $x_q := \dots$

Diese sind dann im ganzen Problem dieser Datei verfügbar $\text{approx}(\mathbf{m})$.

$\mathbf{b} = \mathbf{y}_q - \mathbf{m} \cdot \mathbf{x}_q$ $\text{approx}(\mathbf{b})$ Die Ausgleichsgerade ist $y = \mathbf{m} \cdot x + \mathbf{b}$

Blatt 3. Data&Statistics-Blatt. Unten trägt man "xi" ein links "yi" Man klickt dazu unten in der Mitte und bekommt mögliche Variablen angeboten. Dann ist die Regressionsgerade (Werkzeugpalette) eingetragen.

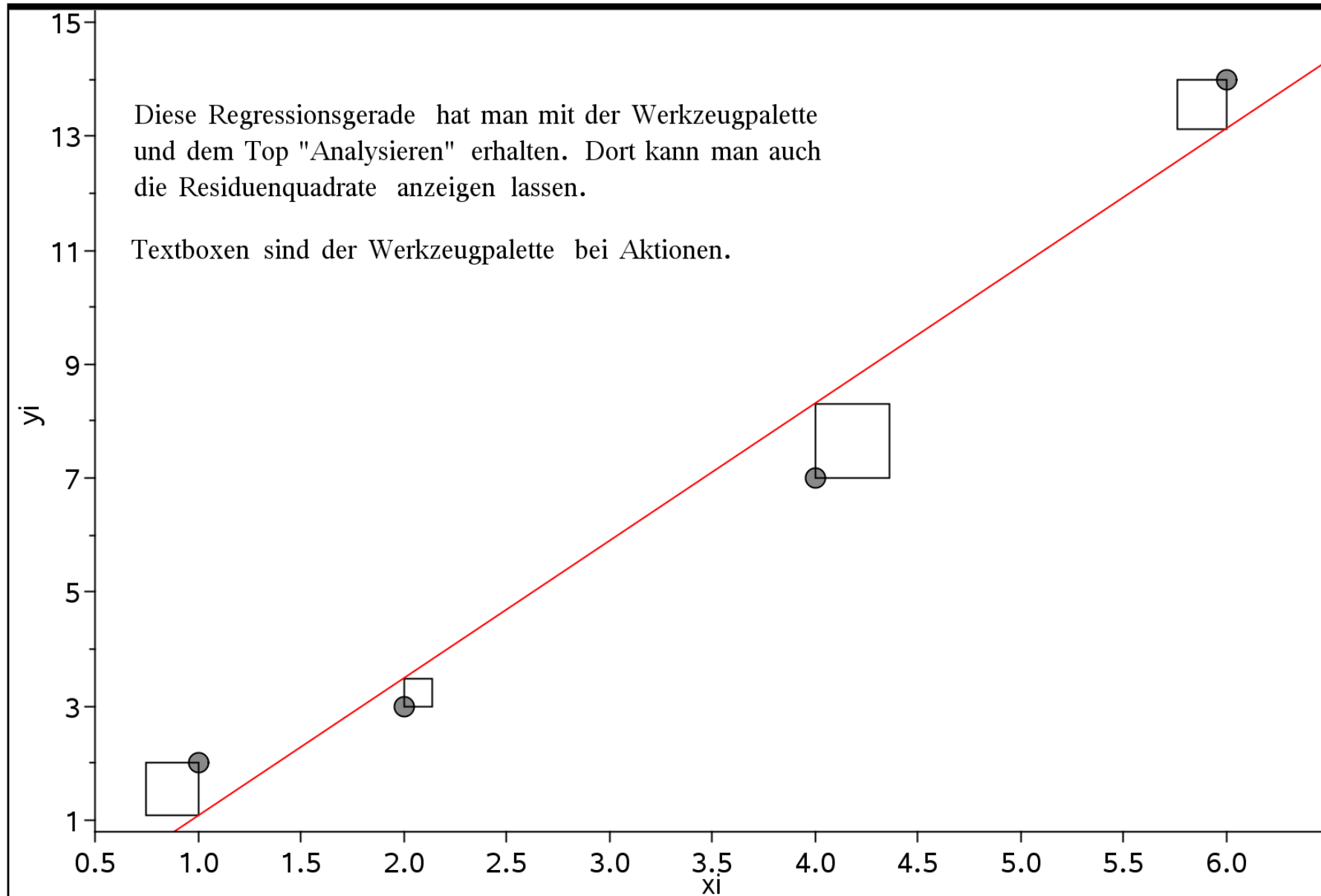
Blatt 4 zeigt die Verwendung der eingebauten linearen Regression

Blatt 5 zeigt eine didaktische Möglichkeit mit einer freien Geraden herumzuspielen. Man ziehe an der blauen Geraden (Mitte: verschieben, außen: drehen) und beobachte die Summe der Quadrate, visualisiert und angezeigt wird.

In **Blatt 6** kann man sehen, dass man mit der blauen Geraden nicht besser wird als mit der Regressionsgeraden. Weiteres in nachfolgenden Problemen.

	A _i	B _{xi}	C _{yi}	D _{x_iy_i}	E _{x_i²}	F	G	H	I	J	K	L	M	N
◆				=x _i *y _i	=x _i ²									
1	1	1	2	2	1	13	26	120	57					
2	2	2	3	6	4	sum(xi)	sum(yi)	sum(xiyi)	sum(xiq)					
3	3	4	7	28	16	xq	yq	m	mm					
4	4	6	14	84	36	13/4	13/2	142/59	2.40678					
5						n		b	bb					
6						4		-78/59	-1.322...					
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
AI	1													

2.2



2.3

Verwendung der eingebauten linearen Regression mit LinRegMx

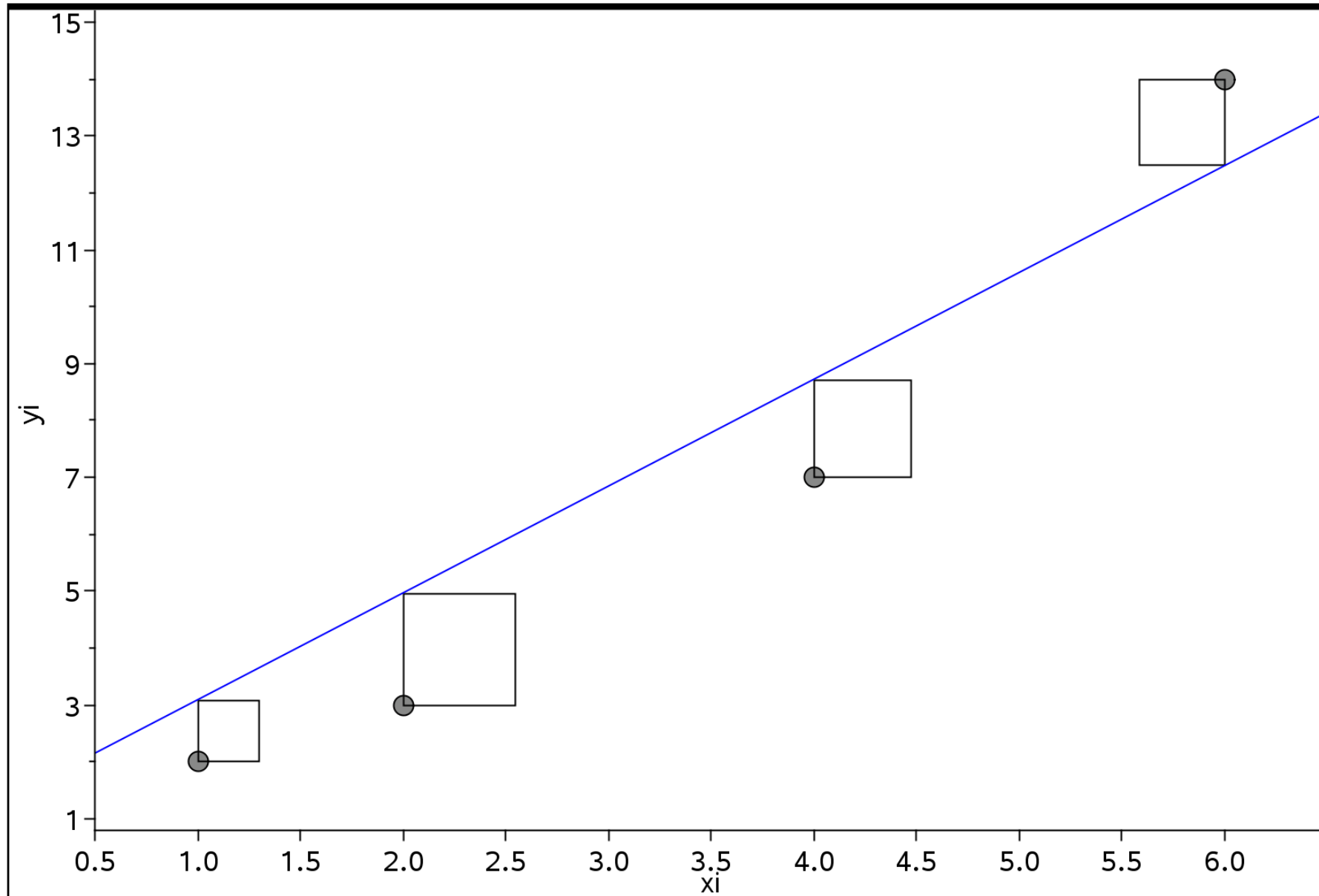
Wenn man diesen Befehl aus dem Katalog nimmt und den Assistenten zulässt, wird ganz bequem mit einer **Eingabemaste** gleich dreierlei erledigt: 1. der Befehl macht (im Hintergrund) die Berechnungen, 2. er trägt mit dem Namen `f2` die Regressionsfunktion als $f2(x)$ ein (`f1` war schon vergeben), 3. er zeigt mit `stat.results` die relevanten Parameter an.

LinRegMx `xi,yi,1: CopyVar stat.RegEqn,f2: stat.results` ▶

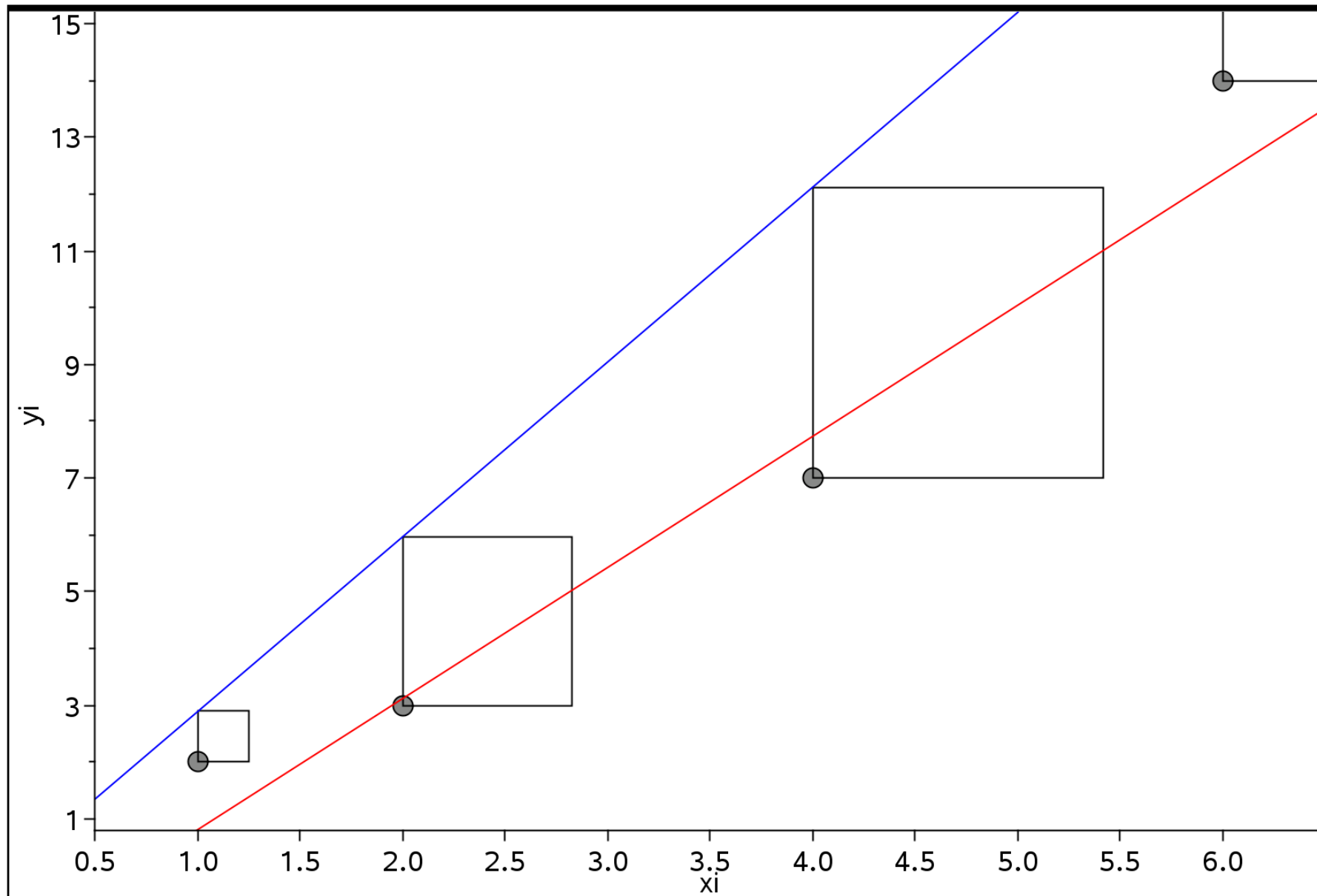
"Titel "	"Lineare Regression (mx+b) "
"RegEqn "	"m*x+b "
"m "	2.40678
"b "	-1.32203
"r ² "	0.960008
"r "	0.9798
"Resid "	" {...} "

Die gesuchte Gleichung ist $f2(x) \blacktriangleright 2.40678 \cdot x - 1.32203$, der Korrelationskoeffizient ist `stat.r` ▶ 0.9798 und die Residuen sind `stat.Resid` ▶ $\{0.915254, -0.491525, -1.30508, 0.881356\}$. Das sind die senkrechten "Fehler", also $yi[1] - f2(xi[1]) \blacktriangleright 0.915254$

Die "Fehlerquadrate" kann sind in Blatt 3 angezeigt. Ihre Summe erscheint, wenn man die Gerade anklickt.



2.5



Exponentielle Regression

Regression Exponentiell Haftendorn 5/08 www.mathematik-verstehen.de

Im 2. Problem wird zu gegebenen Daten die Regressionskurve $y = a \cdot e^{k \cdot x}$ berechnet.

Logarithmiert ist diese Gleichung $\ln(y) = k \cdot x + \ln(a)$. Man braucht also $\ln(y_i)$ (für "von Hand")

Blatt 2 zeigt die Arbeit von Hand, Blatt 3 die Ausgleichsgerade der einfach logarithmierten Daten, Blatt 4 die eingebaute exponentielle Regression und Blatt 5 die Ausgleichskurve der Originaldaten.

Blatt 2. definiert die Daten und rechnet den Logarithmus der y_i aus. Dann macht man damit eine lineare Regression (siehe Problem 1). Nach der Formel der Formelsammlung ist

$$m := \frac{n \cdot \sum(x_i \cdot \ln y_i) - \sum(x_i) \cdot \sum(\ln y_i)}{n \cdot \sum(x_i^2) - (\sum(x_i))^2}$$

In den gelben Feldern sind Variable definiert worden $x_q := \dots$

Diese sind dann im ganzen Problem dieser Datei verfügbar $\text{approx}(m)$. $b = y_q - m \cdot x_q$ $\text{approx}(b)$ Die Ausgleichsgerade ist $y = mm \cdot x + bb$ Wegen $\ln(y) = k \cdot x + \ln(a)$ ist die Exponentielle Regressionskurve nun mit $bb = \ln(a)$ also $a := e^{bb} \triangleright 1.37677$ und $k := mm \triangleright 0.392135$ bestimmt worden.

Blatt 3. Data&Statistics-Blatt. Unten trägt man "xi" ein links "lnyi" Man klickt dazu unten in der Mitte und bekommt mögliche Variablen angeboten. Dann ist die Regressionsgerade (Werkzeugpalette) eingetragen.

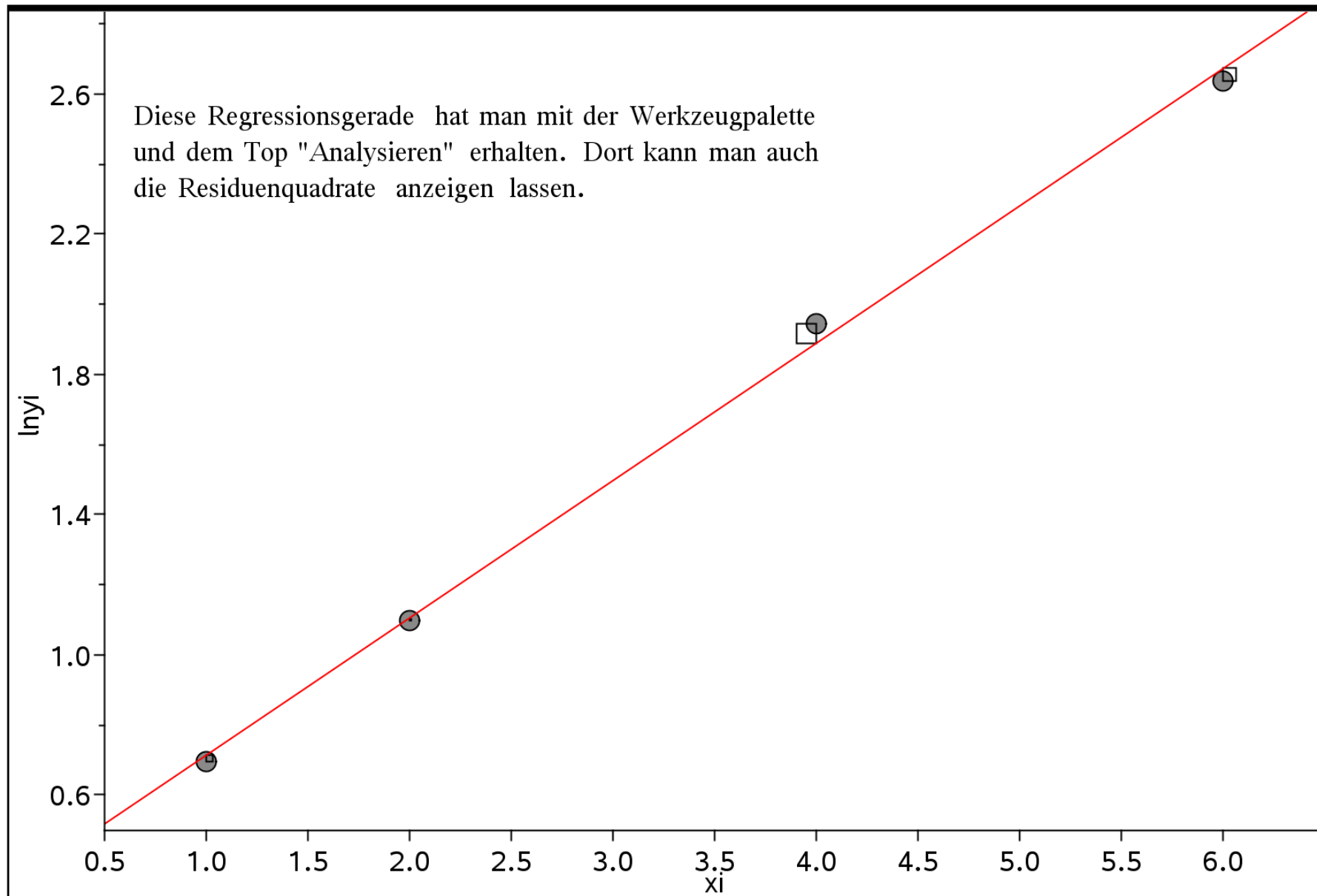
Man nennt dies: einfach-logarithmische Darstellung.

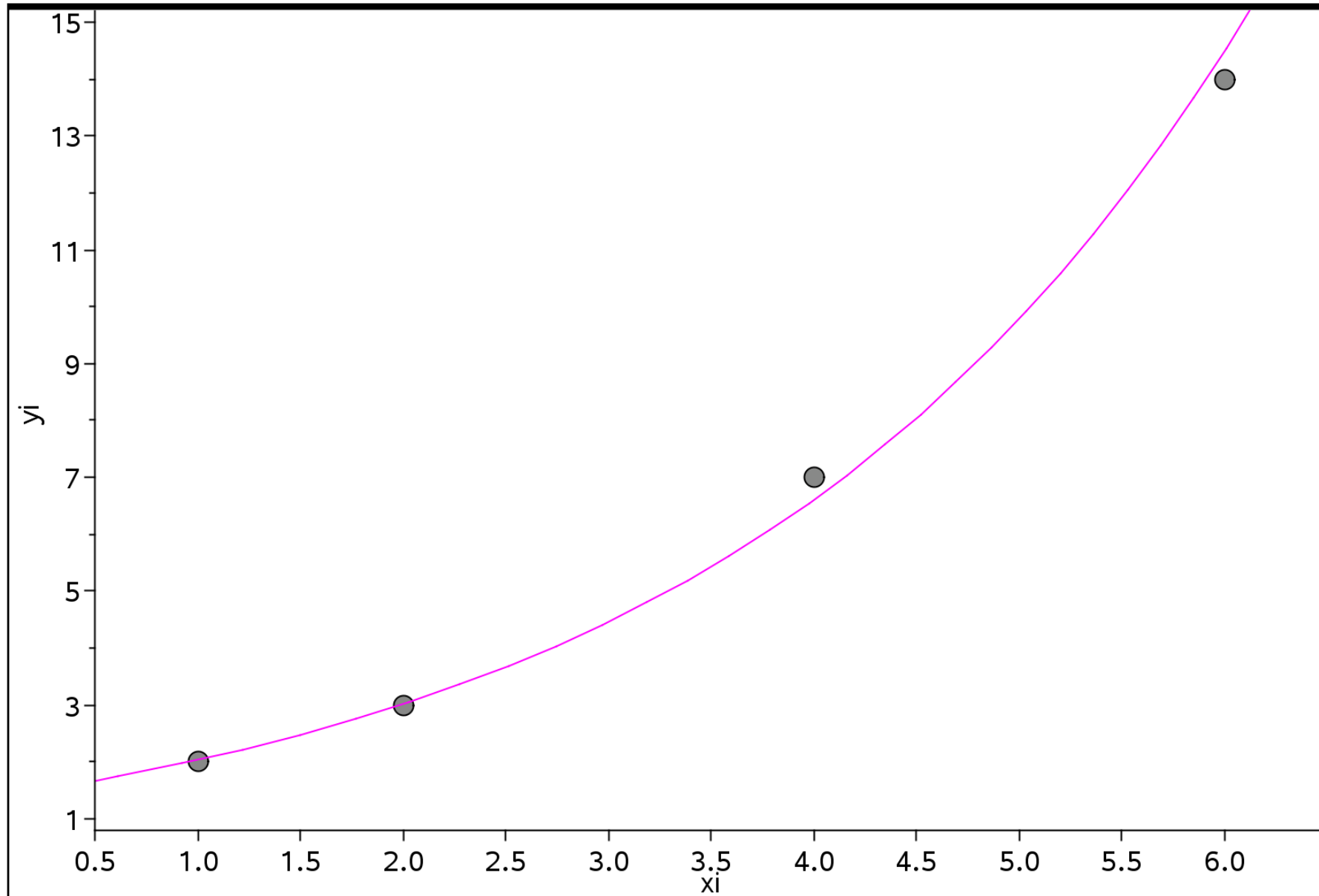
Blatt 4 zeigt die Deutung in den ursprünglichen Daten.

Blatt 5 zeigt die Verwendung der eingebauten exponentiellen Regression

	A _i	B _{xi}	C _{yi}	D _{x_iy_i}	E _{x_i²}	F	G	H	I	J	K _{lny_i}	L	M
◆					=xi^2						=approx(lr		
1	1	1	2	2	1	13	6.37673	26.5084	57		0.693147		
2	2	2	3	6	4	sum(xi)	sum(lny _i)	sum(xi*ln...	sum(xiq)		1.09861		
3	3	4	7	20	16	xq	yq	m	mm		1.94591		
4	4	6	14	84	36	13/4	1.59418	0.392135	0.3921...		2.63906		
5						n		b	bb				
6						4		0.319742	0.3197...				
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
A1													1

3.2





3.4

Verwendung der eingebauten exponentiellen Regression mit ExpReg

Wenn man diesen Befehl aus dem Katalog nimmt und den Assistenten zulässt, wird ganz bequem mit einer **Eingabemaste** gleich dreierlei erledigt: 1. der Befehl macht (im Hintergrund) die Berechnungen, 2. er trägt mit dem Namen f1 die Regressionsfunktion als $f1(x)$ ein, 3. er zeigt mit stat.results die relevanten Parameter an.

ExpReg $\mathbf{xi,yi,1}$: CopyVar **stat.RegEqn,f1: stat.results** ▶

"Titel"	"Exponentielle Regression"
"RegEqn"	"a*b^x"
"a"	1.37677
"b"	1.48014
"r ² "	0.997878
"r"	0.998939
"Resid"	" {... }"
"ResidTrans"	" {... }"

Die gesuchte Gleichung ist $f1(x) \blacktriangleright 1.37677 \cdot (1.48014)^x$, der Korrelationskoeffizient ist **stat.r** ▶ 0.998939

und die Residuen sind **stat.Resid** ▶ $\{-0.037814, -0.016245, 0.391986, -0.476888\}$. Das sind die senkrechten "Fehler", also $\mathbf{yi[1]} - f1(\mathbf{xi[1]}) \blacktriangleright -0.037814$

Achtung: es gibt in jedem Problem nur eine einzige Variable stat.results. Wenn man z.B. nun LinRegMx für xi und lnyi ausführte, würde die obige Ausgabe zerstört.

Der TI gibt die exponentielle Funktion mit umgerechneter Basis aus. Es gilt $e^k = \mathbf{stat.b}$ also $\ln(\mathbf{stat.b}) \blacktriangleright 0.392135$, das ist tatsächlich das k aus der Berechnung "von Hand".

Potenzfunktion als Regressionskurve**Haftendorn 5/08 www.mathematik-verstehen.de**

Im 3. Problem wird zu gegebenen Daten die Regressionskurve $y = a \cdot x^k$ berechnet.

Logarithmiert ist diese Gleichung $\ln(y) = k \cdot \ln(x) + \ln(a)$. Man braucht also $\ln(x_i)$ und $\ln(y_i)$ (für "von Hand")

Blatt 2 zeigt die Arbeit von Hand, Blatt 3 die Ausgleichsgerade der doppelt logarithmierten Daten, Blatt 4 die eingebaute Potenz-Regression und Blatt 5 die Ausgleichskurve der Originaldaten.

Blatt 2 definiert die Daten und rechnet den Logarithmus der x_i und der y_i aus. Dann macht man damit eine lineare Regression (siehe Problem 1). Nach der Formel der Formelsammlung ist

$$m := \frac{n \cdot \sum(\ln x_i \cdot \ln y_i) - \sum(\ln x_i) \cdot \sum(\ln y_i)}{n \cdot \sum(\ln x_i^2) - (\sum(\ln x_i))^2}$$

In den gelben Feldern sind Variable definiert worden

$\ln x_q := \dots$

$b = \ln y_q - m \cdot \ln x_q \approx 0.548514$ Die Ausgleichsgerade ist $y = 1.08046 \cdot x + 0.548514$ Wegen

$\ln(y) = k \cdot \ln(x) + \ln(a)$ ist die Potenz-Regressionkurve nun mit $b = \ln(a)$ also $a = e^b \approx 1.73068$ und $k = m \approx 1.08046$ bestimmt worden.

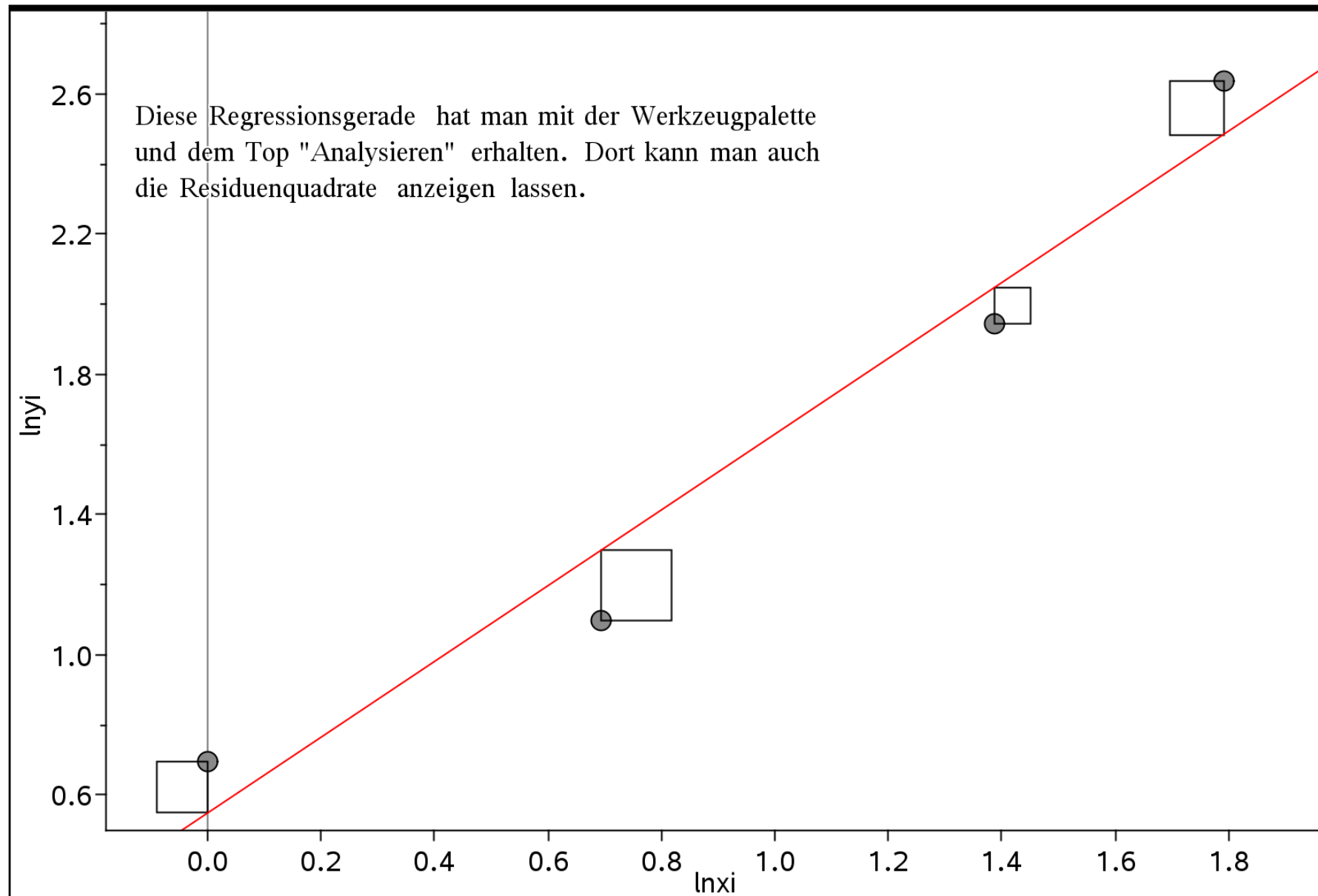
Blatt 3. Data&Statistics-Blatt. Unten trägt man " $\ln x_i$ " ein links " $\ln y_i$ " Man klickt dazu unten in der Mitte und bekommt mögliche Variablen angeboten. Dann ist die Regressionsgerade (Werkzeugpalette) eingetragen. Man nennt dies: doppelt-logarithmische Darstellung.

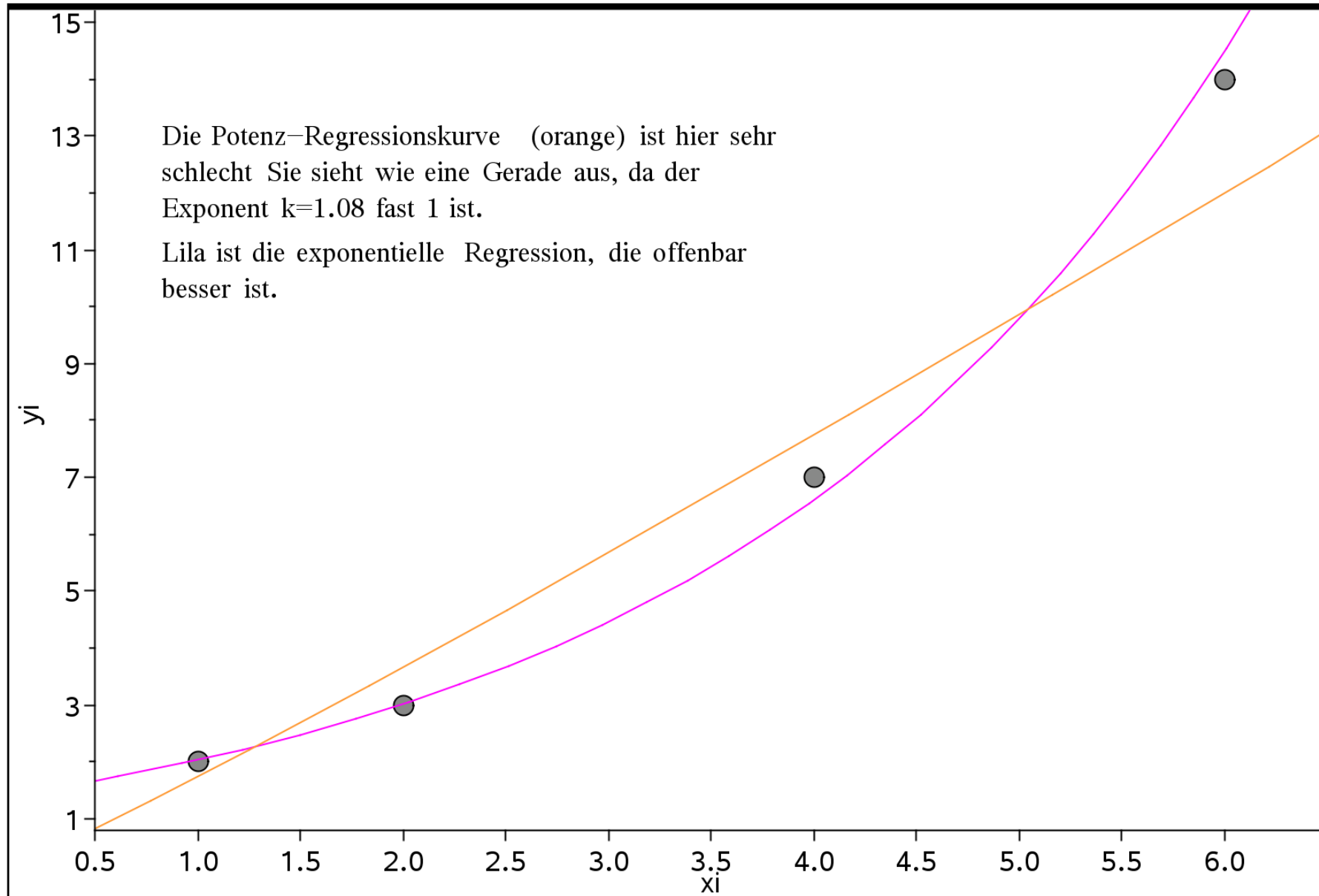
Blatt 4 zeigt die Deutung in den ursprünglichen Daten.

Blatt 5 zeigt die Verwendung der eingebauten exponentiellen Regression

	A	B	C	D	E	F	G	H	I	J	K	L	M
	i	x_i	y_i	$\ln x_i \ln y_i$	x_i^2					$\ln x_i$	$\ln y_i$		
◆				$= \ln x_i * \ln y_i$	$= \ln x_i^2$					$= \text{approx}$	$= \text{approx}(\ln$		
1	1	1	2	0.	0.	3.8712	6.37673	8.18766	5.61267	0.	0.693147		
2	2	2	3	0.693...	0.4...	sum(ln...	sum(lny_i)	sum(lnx_i*...	sum(ln...	0.693...	1.09861		
3	3	4	7	1.386...	1.9...	lnxq	lnyq	m	mm	1.386...	1.94591		
4	4	6	14	1.791...	3.2...	0.9678	1.59418	1.08046	1.08046	1.791...	2.63906		
5						n		b	bb				
6						4		0.548514	0.5485...				
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
A1		1											

4.2





Verwendung der eingebauten Potenz-Regression mit PowerReg

Wenn man diesen Befehl aus dem Katalog nimmt und den Assistenten zulässt, wird ganz bequem mit einer **Eingabemaste** gleich dreierlei erledigt: 1. der Befehl macht (im Hintergrund) die Berechnungen, 2. er trägt mit dem Namen f1 die Regressionsfunktion als $f_1(x)$ ein, 3. er zeigt mit stat.results die relevanten Parameter an.

PowerReg **xi,yi,1: CopyVar stat.RegEqn,f1: stat.results** ▶

"Titel"	"Potenzregression"
"RegEqn"	"a*x^b"
"a"	1.73068
"b"	1.08046
"r ² "	0.958449
"r"	0.979004
"Resid"	" {... }"
"ResidTrans"	" {... }"

Die gesuchte Gleichung ist $f_1(x) \triangleright 1.73068 \cdot x^{1.08046}$, der Korrelationskoeffizient ist **stat.r** ▶ 0.979004, der ist sehr viel schlechter als die r aus den anderen Problemen.

Die Residuen sind **stat.Resid** ▶ $\{0.269321, -0.659881, -0.739578, 2.00566\}$. Das sind die senkrechten "Fehler", also $y_i[1] - f_1(x_i[1]) \triangleright 0.269321$

Achtung: es gibt in jedem Problem nur eine einzige Variable stat.results. Wenn man z.B. nun LinRegMx für lnx_i und lny_i ausführte, würde die obige Ausgabe zerstört.

Die Werte **stat.a** ▶ 1.73068 und **stat.b** ▶ 1.08046 sind tatsächlich die aus der Berechnung "von Hand" auf Blatt 1

Quadratische Regression

Potenzfunktion und Quadratische Regression

Haftendorn 5/08 www.mathematik-verstehen.de

Im 4. Problem wird zu gegebenen Daten die quadratische Regressionskurve $y = a \cdot x^2 + b \cdot x + c$ berechnet. Das geschieht nur noch auf die eingebaute Art. Eine Berechnungsformel verwendet eine Matrix, das ist hier intern programmiert.

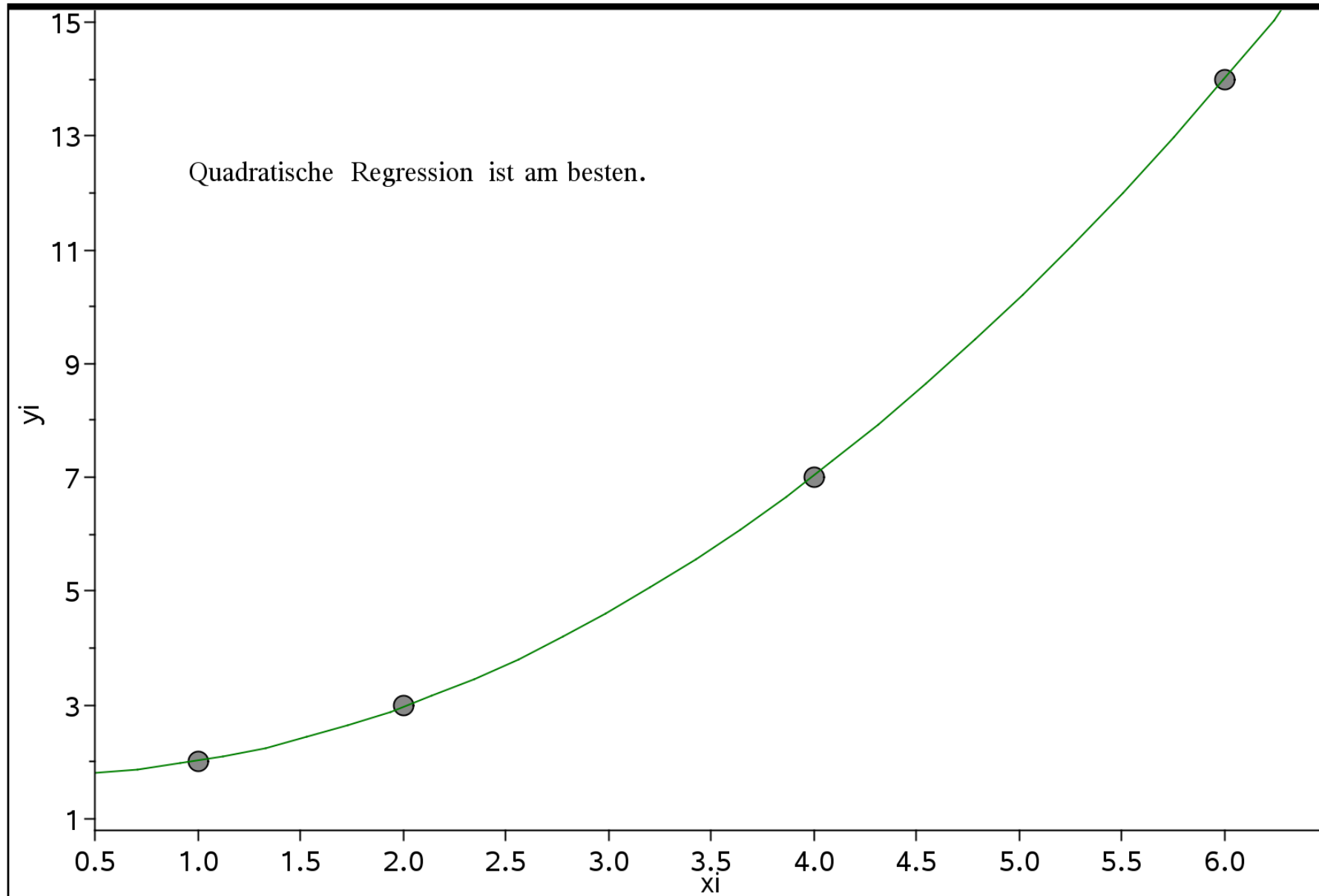
Die Daten werden diesmal hier definiert: $x_i = \{1, 2, 4, 6\}$ und $y_i = \{2, 3, 7, 14\}$

Achtung: Die Mathe-Zellen haben "Attribute" (re-Maus) Hier ist die Ausgabe unterdrückt, damit die Zahlen nicht zweimal da stehen.

Blatt 2. Data&Statistics-Blatt. Unten trägt man "xi" ein links "yi" Man klickt dazu unten in der Mitte und bekommt mögliche Variablen angeboten. Dann ist die Regressionsgerade (Werkzeugpalette) eingetragen.

Blatt 3 zeigt die Deutung in den ursprünglichen Daten.

Blatt 4 zeigt die Verwendung der eingebauten exponentiellen Regression



Verwendung der eingebauten quadratischen Regression mit QuadReg

Wenn man diesen Befehl aus dem Katalog nimmt und den Assistenten zulässt, wird ganz bequem mit einer **Eingabemaste** gleich dreierlei erledigt: 1. der Befehl macht (im Hintergrund) die Berechnungen, 2. er trägt mit dem Namen f1 die Regressionsfunktion als $f_1(x)$ ein, 3. er zeigt mit stat.results die relevanten Parameter an.

QuadReg **xi,yi,1: CopyVar stat.RegEqn,f1: stat.results** ▶

"Titel "	"Quadratische Regression "
"RegEqn "	"a*x^2+b*x+c "
"a"	0.363065
"b"	-0.146985
"c"	1.80402
"R ² "	0.999972
"Resid "	" {... } "

Die gesuchte Gleichung ist $f_1(x) \triangleright 0.363065 \cdot x^2 - 0.146985 \cdot x + 1.80402$, der Korreltaionskoeffizient ist $\sqrt{\text{stat.r}^2} \triangleright 0.999986$. Das ist der höchste dieser ganzen Serie. Die quadratische Regressionskurve ist also die beste Ausgleichskurve.

und die Residuen sind **stat.Resid** ▶ $\{-0.020101, 0.037688, -0.025126, 0.007538\}$. Das sind die senkrechten "Fehler", also $y_i[1] - f_1(x_i[1]) \triangleright -0.020101$

Achtung: es gibt in jedem Problem nur eine einzige Variable stat.results. Wenn man nun eine andere Regression zum Vergleich ausführte, würde die obige Ausgabe zerstört.