Prof. Dr. Dörte Haftendorn, Okt 2011

Zwei-Variablen-Statistik, Vergleich

Zwei-Variablen-Statistik, **lineare, exponetielle, Potenz-, quadratisch Regression weiter unten**

Haftendorn 5/08   www.mathematik-verstehen.de

In dieser Datei werden die wichtigsten Möglichleiten der Zwei-Variabeln-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.

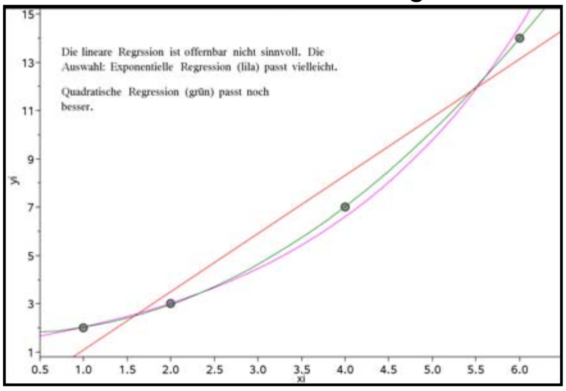
In den nachfolgenden Problemen geht es um Regressionen. Dort wird auch der eigene Umgang mit den Tabellenspalten gezeigt.

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 xi, links in der Mitte wählt man yi 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, Blatt4 und Blatt5 zeigen, wie man mit dem eingebauten Befehl twovar umgeht. Katalog, twovar suchen, Assistenten einschalten, xi,yi eintragen.....

1.1

Zwei-Variablen-Statistik, Vergleich



1.2

Zwei-Variablen-Statistik, Vergleich

	A i	B xi	C yi	D	E zw	F	G	H	I
•					=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_1X	1.
4	4	6	14	Σx²	57.			MedianX	1.5
5				sx := $s_{n-1}x$	2.21736			Q_3X	3.
6				σx := $\sigma_n x$	1.92029			MaxX	5.
7				n	4.			MinY	6.
8				$\bar{y}$	6.5			Q_1Y	2.
9				Σy	26.			MedianY	2.5
10				Σy²	258.			Q_3Y	5.
11				sy := $s_{n-1}y$	5.44671			MaxY	10.5
12				σy := $\sigma_n y$	4.71699			SSX := Σ(x−x̄)²...	14.
13				Σxy	120.			SSY := Σ(y−ȳ)²...	14.75
14				MinX	0.9798				89.
15				Q_1X	1.				
16				MedianX	1.5				
17				Q_3X	3.				
18				MaxX	5				

B7

1.3

Zwei-Variablen-Statistik, Vergleich

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.

TwoVar **xi**,**yi**,1: **stat.results** ▸

"Titel"	"Statistiken mit zwei Variablen"
"$\bar{x}$"	3.25
"Σx"	13.
"Σx²"	57.
"sx := $s_{n-1}x$"	2.21736
"σx := $\sigma_n x$"	1.92029
"n"	4.
"$\bar{y}$"	6.5
"Σy"	26.
"Σy²"	258.
"sy := $s_{n-1}y$"	5.44671
"σy := $\sigma_n y$"	4.71699
"Σxy"	120.
"r"	0.9798
"MinX"	1.
"Q_1X"	1.5
"MedianX"	3.
"Q_3X"	5.
"MaxX"	6.
"MinY"	2.

1.4

Zwei-Variablen-Statistik, Vergleich

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.Q_1Y** ▸ 2.5 **stat.MedianY** ▸ 5. **stat.Q_3Y** ▸ 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" Statndardabweichungen der Grundgesamtheit.

1.5

Lineare Regression

Lineare Regression, **Exponetiell, 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".*

Blatt2. definiert die Daten und rechnet nach der Formel der Formelsammlung

$\mathbf{m}:=\dfrac{\mathbf{n}\cdot \mathrm{sum}(\mathbf{xi}\cdot\mathbf{yi})-\mathrm{sum}(\mathbf{xi}\cdot\mathrm{sum}(\mathbf{yi}))}{\mathbf{n}\cdot\mathrm{sum}(\mathbf{xi}^2)-(\mathrm{sum}(\mathbf{xi}))^2}$ In den gelben Feldern sind Variable definiert worden xq:=....

Diese sind dann im ganzen Problem dieser Datei verfügbar approx(**m**) ▸ 2.40678 .

b=**yq**−**m**·**xq**   approx(**b**) ▸ −1.32203   Die Ausgleichsgerade ist $y=2.40678\cdot x-1.32203$

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 Verwendungder eingebauten linearen Regression

Blatt5 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, viesualisiert und angezeigt wird.

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

2.1

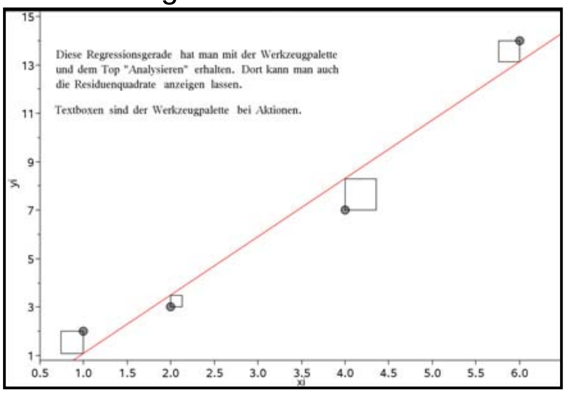
Lineare Regression

	A i	B xi	C yi	D xiyi	E xiq	F	G	H	I
•				=xi*yi	=xi^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**	**−1.32203**
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									

A1 1

2.2

Lineare Regression



2.3

Lineare Regression

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** ▸

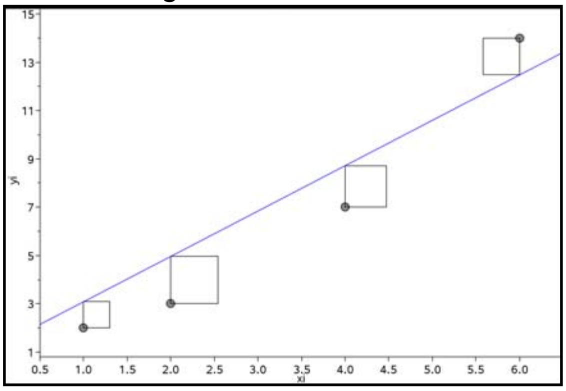
$$\begin{bmatrix} \text{"Titel"} & \text{"Lineare Regression (mx+b)"} \\ \text{"RegEqn"} & \text{"m*x+b"} \\ \text{"m"} & 2.40678 \\ \text{"b"} & -1.32203 \\ \text{"r²"} & 0.960008 \\ \text{"r"} & 0.9798 \\ \text{"Resid"} & \text{"\{...\}"} \end{bmatrix}$$

Die gesuchte Gleichung ist $f2(x)$ ▸ $2.40678\cdot x-1.32203$, der Korretationskoeffizient 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 $yi[1]-f2(xi[1])$ ▸ 0.915254

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

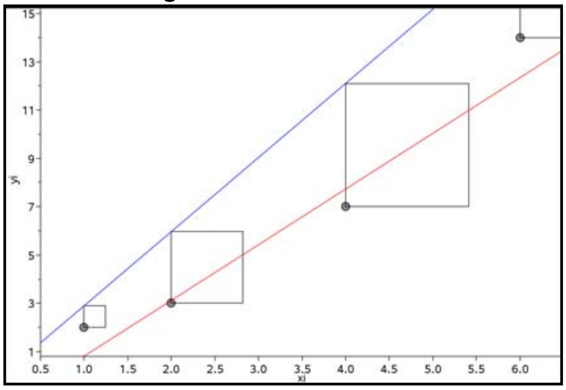
2.4

Lineare Regression



2.5

Lineare Regression



2.6

Exponentielle Regression

Regression Exponetiell

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(yi) (für "von Hand")

Blatt 2 zeigt die Arbeit von Hand, Blatt3 die Ausgleichsgerade der einfach logarithmierten Daten, Blatt 4 die eingebaute exponentielle Regression und Blatt 5 die Ausgleichskurve der Originaldaten.

Blatt2. definiert die Daten und rechnet den Logarithmus der yi aus. Dann macht man damit eine lineare Regression (siehe Problem 1). Nach der Formel der Formelsammlung ist

$$m:=\frac{n\cdot \text{sum}(xi\cdot lnyi)-\text{sum}(xi\cdot \text{sum}(lnyi))}{n\cdot \text{sum}(xi^2)-(\text{sum}(xi))^2}$$ In den gelben Feldern sind Variable definiert worden xq:=....

Diese sind dann im ganzen Problem dieser Datei verfügbar approx(**m**). **b**=**yq**-**m**·**xq** 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 $a:=e^{bb}$ ▸ 1.37677 und $k:=mm$ ▸ 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.

Blatt5 zeigt die Verwendung der eingebauten exponentiellen Regression

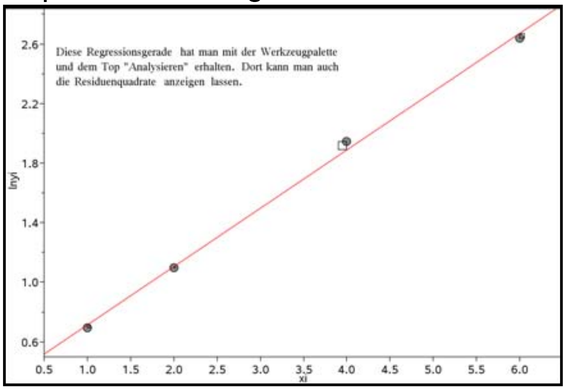
3.1

Exponentielle Regression

	A i	B xi	C yi	D xiyi	E xiq	F	G	H	I	J	K lnyi	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(lnyi)	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...				

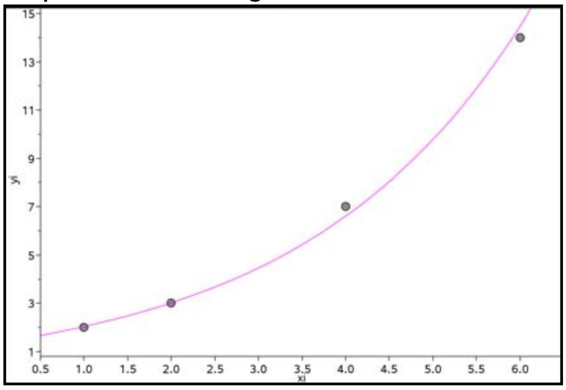
3.2

Exponentielle Regression



3.3

Exponentielle Regression



3.4

Exponentielle Regression

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

$$\begin{bmatrix} \text{"Titel"} & \text{"Exponentielle Regression"} \\ \text{"RegEqn"} & \text{"a*b^x"} \\ \text{"a"} & 1.37677 \\ \text{"b"} & 1.48014 \\ \text{"r²"} & 0.997878 \\ \text{"r"} & 0.998939 \\ \text{"Resid"} & \text{"\{...\}"} \\ \text{"ResidTrans"} & \text{"\{...\}"} \end{bmatrix}$$

Die gesuchte Gleichung ist $f1(x)$ ▸ $1.37677\cdot(1.48014)^x$, der Korreltationskoeffizient 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 $yi[1]-f1(xi[1])$ ▸ -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 umgrechneter Basis aus. Es gilt $e^k=$**stat.b** also $\ln(\textbf{stat.b})$ ▸ 0.392135, das ist tatsächlich das k aus der Berechnung "von Hand".

3.5

Potenz- Regression, doppellogarith.

Logarithmiert ist diese Gleichung $\ln(y)=\mathbf{k}\cdot\ln(x)+\ln(\mathbf{a})$. Man braucht also ln(xi und ln(yi) (für "von Hand")

Blatt 2 zeigt die Arbeit von Hand, Blatt3 die Ausgleichsgerade der doppelt logarithmierten Daten, Blatt 4 die eingebaute Potenz-Regression und Blatt 5 die Ausgleichskurve der Originaldaten.

Blatt2. definiert die Daten und rechnet den Logarithmus der xi und der yi aus. Dann macht man damit eine lineare Regression (siehe Problem 1). Nach der Formel der Formelsammlung ist

$\mathbf{m}:=\frac{\mathbf{n}\cdot\mathrm{sum}(\mathbf{lnxi}\cdot\mathbf{lnyi})-\mathrm{sum}(\mathbf{lnxi}\cdot\mathrm{sum}(\mathbf{lnyi}))}{\mathbf{n}\cdot\mathrm{sum}(\mathbf{lnxi}^2)-(\mathrm{sum}(\mathbf{lnxi}))^2}$ In den gelben Feldern sind Variable definiert worden lnxq:=....

$\mathbf{b}=\mathbf{lnyq}-\mathbf{m}\cdot\mathbf{lnxq}$ $\mathrm{approx}(\mathbf{b}) \cdot 0.548514$ Die Ausgleichsgerade ist $y=1.08046\cdot x+0.548514$ Wegen $\ln(y)=\mathbf{k}\cdot\ln(x)+\ln(\mathbf{a})$ ist die Potenz-Regressionskurve nun mit $\mathbf{bb}=\ln(\mathbf{a})$ also $\mathbf{a}:=e^{\mathbf{bb}} \cdot 1.73068$ und $\mathbf{k}:=\mathbf{mm} \cdot 1.08046$ bestimmt worden.

Blatt 3. Data&Statistics-Blatt. Unten trägt man "lnxi" ein links "lnyi" 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.

Blatt5 zeigt die Verwendungder eingebautenexponentiellen Regression

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.

Man nennt dies: einfach-logarithmische Darstellung.

Blatt 4 zeigt die Deutung in den ursprünglichen Daten.

Blatt5 zeigt die Verwendungder eingebautenexponentiellen Regression

4.1

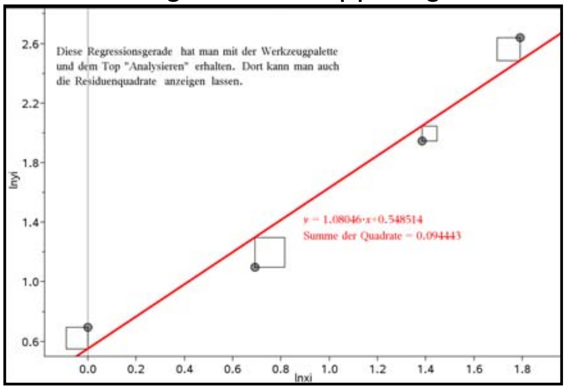
Potenz- Regression, doppellogarith.

	A i	B xi	C yi	D lnxil...	E xiq	F	G	H	I	J lnxi	K lnyi	L	M
•				=lnxi*lxy	=lnxi²					=approx	=approx(lr		
1	1	1	2	0.	0.	3.8712	6.37673	8.18766	5.61267	0.	0.693147		
2	2	2	3	0.6931...	0.4...	sum(ln...	sum(lnyi)	sum(lnxi*l...	sum(ln...	0.6931...	1.09861		
3	3	4	7	1.3862...	1.9...	lnxq	lnyq	m	mm	1.3862...	1.94591		
4	4	6	14	1.7917...	3.2...	0.9678	1.59418	1.08046	1.08046	1.7917...	2.63906		
5						n		b	bb				
6						4		0.548514	0.5485...				
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													

F7

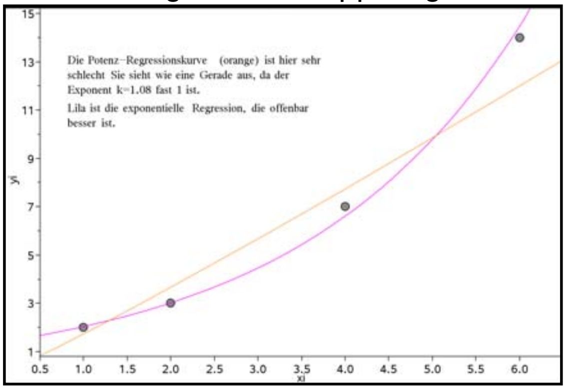
4.2

Potenz- Regression, doppellogarith.



4.3

Potenz- Regression, doppellogarith.



4.4

Potenz- Regression, doppellogarith.

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 f1(x) ein. 3. er zeigt mit stat.results die relevanten Parameter an.

PowerReg **xi**,**yi**,1: CopyVar **stat.RegEqn**,**f1**: **stat.results** •
$\begin{bmatrix} \text{"Titel"} & \text{"Potenzregression"} \\ \text{"RegEqn"} & \text{"a*x^b"} \\ \text{"a"} & 1.73068 \\ \text{"b"} & 1.08046 \\ \text{"r²"} & 0.958449 \\ \text{"r"} & 0.979004 \\ \text{"Resid"} & \text{"\{...\}"} \\ \text{"ResidTrans"} & \text{"\{...\}"} \end{bmatrix}$

Die gesuchte Gleichung ist $\mathbf{f1}(x) \cdot 1.73068\cdot x^{1.08046}$, der Korreltationskoeffizient 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 $\mathbf{yi}[1]-\mathbf{f1}(\mathbf{xi}[1]) \cdot 0.269321$

Achtung: es gibt in jedem Problem nur eine einzige Variable stat.results. Wenn man z.B. nun LinRegMx für lnxi und lnyi 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

4.5

Quadratische Regression

Potenzfunktion und Quadratische Regression

Haftendorn 5/08 www.mathematik-verstehen.de

Im 4. Problem wird zu gegebenen Daten die quadratische Regressionskurve $y=a\cdot x^2+\mathbf{b}\cdot x+c$ berechnet.

Das geschieht nur noch auf die eingebaute Art. Eine Berechnungsformel verwendet eine Matrix, das ist hier intern programmiert.

Die Daten werden diesesmal hier definiert: **xi**:=$\{1,2,4,6\}$ und **yi**:=$\{2,3,7,14\}$

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

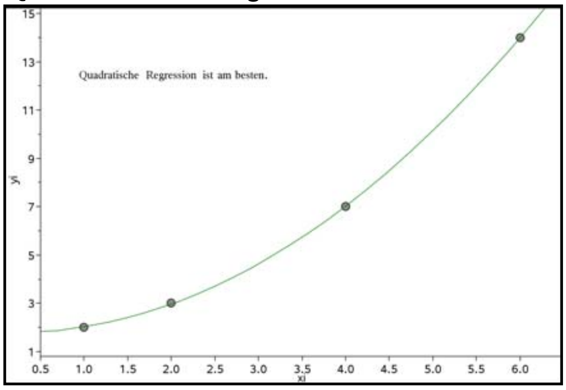
Blatt 2. Data&Statistics-Blatt. Unten trägt man "xi" ein links "yi" Man klickt dazu unten in der Mitte und bekommt mögliche Variablen angeboten. Dann ist die Regressionsgerade (Werkzeugpalette)eingetragen.

Blatt 3 zeigt die Deutung in den ursprünglichen Daten.

Blatt4 zeigt die Verwendungder eingebautenexponentiellen Regression

5.1

Quadratische Regression



5.2

Quadratische 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 f1(x) ein. 3. er zeigt mit stat.results die relevanten Parameter an.

QuadReg **xi**,**yi**,1: CopyVar **stat.RegEqn**,**f1**: **stat.results** •
$\begin{bmatrix} \text{"Titel"} & \text{"Quadratische Regression "} \\ \text{"RegEqn"} & \text{"a*x^2+b*x+c "} \\ \text{"a"} & 0.363065 \\ \text{"b"} & -0.146985 \\ \text{"c"} & 1.80402 \\ \text{"R²"} & 0.999972 \\ \text{"Resid"} & \text{"\{...\}"} \end{bmatrix}$

Die gesuchte Gleichung ist $\mathbf{f1}(x) \cdot 0.363065\cdot x^2-0.146985\cdot x+1.80402$, der Korreltaionskoeffizient ist $\sqrt{\mathbf{stat.r^2}} \cdot 0.999986$ 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 $\mathbf{yi}[1]-\mathbf{f1}(\mathbf{xi}[1]) \cdot -0.020101$

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.

5.3