Lineare Regression und Korrelation (s. auch Applet auf www.mathematik.ch)

Fragestellung: Die **lineare Regression** beschäftigt sich mit der folgenden Fragestellung: Gegeben sind n Punkte (x_i / y_i) , i = 1,.. ,n im (x,y)-Koordinatensystem (n > 1). Gesucht ist die **lineare** Funktion mit Gleichung $y = f(x) = ax + b$, die die Punkte 'optimal annähert'.
Die **Korrelationsrechnung** bestimmt nun ein Mass dafür, ob die Annahme eines linearen Zusammenhangs zwischen den x- und den y-Werten sinnvoll ist.

Lernziele:
1. Bedeutung der 'optimalen' Annäherung.
2. Definition und Berechnung von Mittelwert und Varianz der x-Werte bzw. der y-Werte sowie der Kovarianz.
3. Bestimmung der Parameter a und b in der Geradengleichung $y = f(x) = ax + b$.
4. Begriff des Korrelationskoeffizienten r_{xy} und seine Bedeutung als Mass für die Qualität eines linearen Zusammenhangs der x- und y-Werte.
5. Berechnung des Korrelationskoeffizienten r_{xy} .

Vorgehen: Studium von *Arbeitsblatt 1: Lineare Regression* und das *Arbeitsblatt 2: Korrelationsrechnung*.
Lernkontrolle: Lösen der jeweils angegebenen Aufgaben.

Arbeitsblatt 1 : Lineare Regression

Bevor wir auf den Begriff 'optimal annähern' eingehen, wollen wir noch einige Formeln, die von früher her bekannt sind, zusammenstellen: siehe auch DMK/DPK Formeln und Tafeln, p. 85.

Mittelwert der x-Werte: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ Mittelwert der y-Werte: $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Varianz der x-Werte: $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ Varianz der y-Werte: $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$

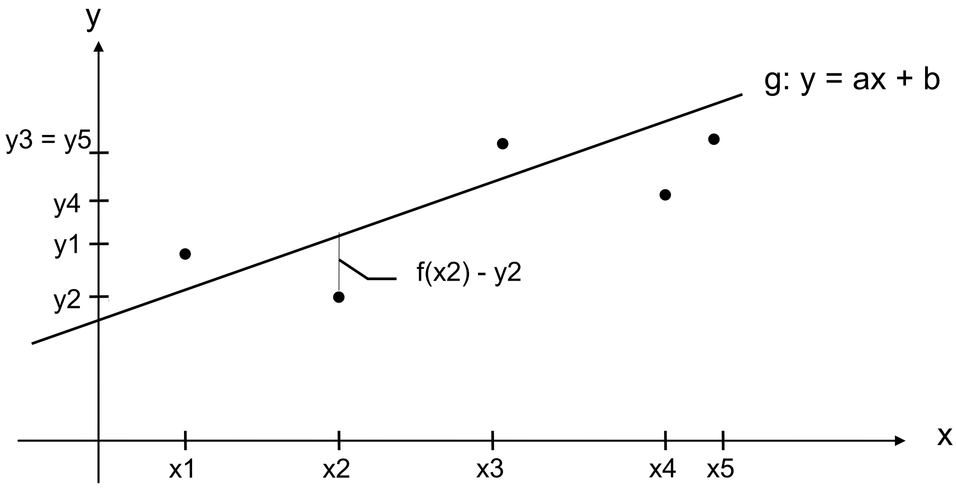
Wie lässt sich nun der Begriff 'optimal annähern' mathematisch präzisieren? Als praktisch brauchbar und auch theoretisch gut fundiert hat sich die **Summe der Quadrate** der Abweichungen $f(x_i) - y_i$ zwischen Funktionswert und y-Koordinate der Punkte (x_i / y_i) herausgestellt. Wir definieren die **Fehlerquadratsumme** F :

$$F := \sum_{i=1}^{n}(f(x_i) - y_i)^2 \; .$$

Da $f(x_i) = ax_i + b$, so wird F in Abhängigkeit von a und b zu $F(a,b) = \sum_{i=1}^{n}(ax_i + b - y_i)^2$.

Wir suchen nun diejenige Gerade g, für die diese Fehlerquadratsumme **minimal** ist. Es sind also die zwei Unbekannten a und b zu berechnen. Mit Mitteln der höheren

Mathematik kann man zeigen, dass der Punkt ($\bar{x}$ / $\bar{y}$) immer auf dieser gesuchten Geraden g liegt.



Da also $\bar{y} = f(\bar{x}) = a\bar{x} + b$, so lässt sich b durch a ausdrücken: $b = \bar{y} - a\bar{x}$. F ist also nur noch abhängig von einer Variablen a:

$F(a) = \sum_{i=1}^{n}(ax_i + \bar{y} - a\bar{x} - y_i)^2$. Da F(a) minimal werden soll, so bildet man die erste Ableitung F'(a) {summandenweise Ableiten nach a, Kettenregel beachten!} und setzt sie gleich Null:

$F'(a) = \sum_{i=1}^{n} 2(ax_i + \bar{y} - a\bar{x} - y_i)\cdot(x_i - \bar{x}) = 0$. Division mit 2 und Ausrechnen der Klammer

liefert: $\sum_{i=1}^{n}(ax_i^2 - 2a\bar{x}x_i + \bar{y}x_i - \bar{y}\bar{x} + a\bar{x}^2 - x_i y_i + \bar{x}y_i) = 0$, also

$a\left(\sum_{i=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}^2)\right) = \sum_{i=1}^{n}(-\bar{y}x_i + \bar{y}\bar{x} + x_i y_i - \bar{x}y_i)$. Berücksichtigt man $n\bar{x} = \sum_{i=1}^{n} x_i$,

$n\bar{y} = \sum_{i=1}^{n} y_i$ und die Rechenregeln für Summen {konstante Faktoren nach vorne!}, so gilt:

$a\left(\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right) = -n\bar{x}\bar{y} + n\bar{x}\bar{y} - n\bar{x}\bar{y} + \sum_{i=1}^{n} x_i y_i$. Die Steigung a berechnet sich also zu

$$a = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

. b lässt sich dann aus a berechnen: $b = \bar{y} - a\bar{x}$ (s. weiter oben).

Die Lösung der Aufgabe 3 zeigt, dass F für dieses a tatsächlich minimal wird.

Definiert man die Kovarianz $c_{xy} := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ und zeigt man, dass

$c_{xy} = \frac{1}{n-1}(\sum_{i=1}^{n} x_i y_i - n\,\overline{x}\,\overline{y})$ und $s_x^2 = \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2)$ (siehe Aufgabe 2), so gilt für die

Steigung a der gesuchten Geraden: $a = \frac{\text{Kovarianz}}{\text{Varianz der } x-\text{Werte}} = \frac{c_{xy}}{s_x^2}$.

Beispiel:	x-Werte:	1	2	3	4	(n = 4)
	y-Werte:	1	2	4	5	

Auf spezielle Darstellung zur Berechnung von Hand wird verzichtet, da später die gewünschten Grössen ohnehin vom Taschenrechner geliefert werden.

Mittelwert $\underline{\overline{x}} = \frac{1}{4}(1+2+3+4) = \underline{2.5}$

Mittelwert $\underline{\overline{y}} = \frac{1}{4}(1+2+4+5) = \underline{3}$

Varianz der x-Werte $\underline{s_x^2} = \frac{1}{3}((1 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2) = \underline{\frac{5}{3}}$

Varianz der y-Werte $\underline{s_y^2} = \frac{1}{3}((1 - 3)^2 + (2 - 3)^2 + (4 - 3)^2 + (5 - 3)^2) = \underline{\frac{10}{3}}$

Kovarianz $\underline{c_{xy}} = \frac{1}{3}((-1.5) \cdot (-2) + (-0.5) \cdot (-1) + 0.5 \cdot 1 + 1.5 \cdot 2) = \underline{\frac{7}{3}}$

Steigung $\underline{a} = \frac{c_{xy}}{s_x^2} = \frac{\frac{7}{3}}{\frac{5}{3}} = \frac{7}{5} = \underline{1.4}$, Achsenabschnitt $b = 3 - 1.4 \cdot 2.5 = \underline{-0.5}$

Die Gleichung der Regressionsgeraden g heisst also $\underline{y = f(x) = 1.4x - 0.5}$

Aufgaben zu Teil 1

Nr. 1:

Gegeben sind die folgenden vier Punkte:

x-Werte	0	2	3	5
y-Werte	8	3	1	-2

Bestimme (ohne Mithilfe der speziellen Funktionen des Taschenrechners) die Mittelwerte $\overline{x}$ und $\overline{y}$, die Varianzen s_x^2 und s_y^2, die Kovarianz c_{xy} und die Gleichung der Ausgleichsgeraden g: y = f(x) = ax + b.

Nr. 2:

Beweise mit Hilfe der Rechenregeln für Summen und der Definition der Mittelwerte:

a) Für die Varianz s_x^2 der x-Werte gilt: $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}(\sum_{i=1}^{n}x_i^2 - n\,\overline{x}^2)$

b) Für die Kovarianz c_{xy} gilt: $c_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n-1}(\sum_{i=1}^{n}x_i y_i - n\,\overline{x}\,\overline{y})$

Nr. 3:

Im Theorieteil haben wir aus der Setzung F'(a) = 0 für a den Wert $a = \frac{c_{xy}}{s_x^2}$ erhalten.

Zeige mit Hilfe der zweiten Ableitung F''(a), dass F für dieses a tatsächlich minimal wird.

Lösungen

1. $\overline{x} = 2.5$ $\quad\overline{y} = 2.5$ $\quad s_x^2 = 4.333...$ $\quad s_y^2 = 17.666...$ $\quad c_{xy} = -8.666...$
   g: y = -2x + 7.5

2. a) $\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2) = \sum_{i=1}^{n}x_i^2 - 2\overline{x}\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\overline{x}^2 =$

$$\sum_{i=1}^{n}x_i^2 - 2\overline{x}\,n\,\overline{x} + n\,\overline{x}^2 = \sum_{i=1}^{n}x_i^2 - n\,\overline{x}^2$$

b) $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n}(x_i y_i - x_i\overline{y} - \overline{x}y_i + \overline{x}\overline{y}) = ... = \sum_{i=1}^{n}x_i y_i - n\,\overline{x}\,\overline{y}$

3. $F''(a) = 2\sum_{i=1}^{n}(x_i^2 - 2\overline{x}\,x_i + \overline{x}^2) = 2\sum_{i=1}^{n}(x_i - \overline{x})^2 = 2(n-1)\,s_x^2 > 0$ (sogar für alle a)

Arbeitsblatt 2 : Korrelationsrechnung

Aus dem Arbeitsblatt 1 ist das Verfahren, um für jede beliebige Punktwolke (x_i / y_i) , i = 1,.. ,n eine Ausgleichsgerade mit Gleichung y = ax + b zu bestimmen, bekannt. Die **Korrelationsrechnung** gibt nun ein Mass dafür an, ob die Annahme eines **linearen** Zusammenhangs zwischen den y - und den x-Werten sinnvoll ist. Dazu wird der sog. Korrelationskoeffizient r_{xy} definiert.

Im Arbeitsblatt 1 wurde die **Fehlerquadratsumme** F definiert:

$$F := \sum_{i=1}^{n}(f(x_i) - y_i)^2 \text{ bzw. } F(a,b) = \sum_{i=1}^{n}(ax_i + b - y_i)^2 .$$

Da $\overline{y} = f(\overline{x}) = a\overline{x} + b$, so ist $b = \overline{y} - a\overline{x}$.

F ist also nur noch abhängig von einer Variablen a:

$$F(a) = \sum_{i=1}^{n}(ax_i + \overline{y} - a\overline{x} - y_i)^2 = \sum_{i=1}^{n}(a(x_i - \overline{x}) + (\overline{y} - y_i))^2 =$$

$$\sum_{i=1}^{n}(a^2(x_i - \overline{x})^2 + 2a(x_i - \overline{x})(\overline{y} - y_i) + (\overline{y} - y_i)^2)$$

Im Arbeitsblatt 1 wurde die Kovarianz c_{xy} als $c_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ definiert.

Für die Steigung a der Geraden g folgte dann dort:

$$a = \frac{\text{Kovarianz}}{\text{Varianz der } x-\text{Werte}} = \frac{c_{xy}}{s_x^2} .$$

Setzt man nun a in die Fehlerquadratsumme F ein, so gilt:

$$F = \sum_{i=1}^{n}(\frac{c_{xy}^2}{s_x^4}(x_i - \overline{x})^2 + 2\frac{c_{xy}}{s_x^2}(x_i - \overline{x})(\overline{y} - y_i) + (\overline{y} - y_i)^2) .$$

Berücksichtigt man $(n-1)\, s_x^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2$, $(n-1)\, s_y^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2$,

$(n-1)\, c_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ und die Rechenregeln für Summen {konstante Faktoren nach vorne!}, so gilt:

$$F =$$
$$\frac{c_{xy}^2}{s_x^4}(n-1)\, s_x^2 - 2\frac{c_{xy}}{s_x^2}(n-1)c_{xy} + (n-1)\, s_y^2 = (n-1)(s_y^2 - \frac{c_{xy}^2}{s_x^2}) = (n-1)\, s_y^2\, (1 - \frac{c_{xy}^2}{s_x^2 s_y^2})$$

Wir definieren nun den **Korrelationskoeffizienten** $\mathbf{r_{xy}}$ als $r_{xy} := \frac{c_{xy}}{s_x s_y}$.

Eigenschaften von r_{xy} :

1. $-1 \leq r_{xy} \leq 1$ denn: Fehlerquadratsumme $F \geq 0$, also $\frac{c_{xy}^2}{s_x^2 s_y^2} \leq 1$.

2. $r_{xy} = a \cdot \frac{s_x}{s_y}$ denn: $a = \frac{c_{xy}}{s_x^2}$

3. $r_{xy} > 0 \Leftrightarrow c_{xy} > 0 \Leftrightarrow a > 0 \Leftrightarrow$ Regressionsgerade g hat positive Steigung.

   $r_{xy} < 0 \Leftrightarrow c_{xy} < 0 \Leftrightarrow a < 0 \Leftrightarrow$ Regressionsgerade g hat negative Steigung.

4. Je näher $|r_{xy}|$ bei 1, desto besser die Linearität zwischen den x- und y-Werten, denn je näher $|r_{xy}|$ bei 1, desto kleiner wird die Fehlerquadratsumme F.

5. Für $|r_{xy}| = 1$ liegen alle Punkte $(x_i \ / \ y_i)$ auf der Geraden g, denn F wird Null.

Das folgende Beispiel ist bereits von Arbeitsblatt 1 bekannt. Berechnet wird nun zusätzlich der Korrelationskoeffizient.

Beispiel:	x-Werte:	1	2	3	4	(n = 4)
	y-Werte:	1	2	4	5	

Auf spezielle Darstellung zur Berechnung von Hand wird verzichtet, da später die gewünschten Grössen ohnehin vom Taschenrechner geliefert werden.

Mittelwert $\underline{\bar{x}} = \frac{1}{4}(1+2+3+4) = \underline{2.5}$

Mittelwert $\underline{\bar{y}} = \frac{1}{4}(1+2+4+5) = \underline{3}$

Varianz der x-Werte $\underline{s_x^2} = \frac{1}{3}((1 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2) = \underline{\frac{5}{3}}$

Varianz der y-Werte $\underline{s_y^2} = \frac{1}{3}((1 - 3)^2 + (2 - 3)^2 + (4 - 3)^2 + (5 - 3)^2) = \underline{\frac{10}{3}}$

Kovarianz $\underline{c_{xy}} = \frac{1}{3}((-1.5) \cdot (-2) + (-0.5) \cdot (-1) + 0.5 \cdot 1 + 1.5 \cdot 2) = \underline{\frac{7}{3}}$

Steigung $\underline{a} = \frac{c_{xy}}{s_x^2} = \frac{\frac{7}{3}}{\frac{5}{3}} = \frac{7}{5} = \underline{1.4}$, Achsenabschnitt $b = 3 - 1.4 \cdot 2.5 = \underline{-0.5}$

Die Gleichung der Regressionsgeraden g heisst also $\underline{y = f(x) = 1.4x - 0.5}$

Korrelationskoeffizient $r_{xy} = a\,\frac{s_x}{s_y} = 1.4 \cdot \frac{\sqrt{5}}{\sqrt{10}} = \frac{1.4}{\sqrt{2}} \approx 0.9899$, also gute Korrelation.

Aufgaben zu Teil 2

Nr. 1:

Gegeben sind die folgenden vier Punkte:

x-Werte	0	2	3	5
y-Werte	8	3	1	-2

Bestimme (ohne Mithilfe der speziellen Funktionen des Taschenrechners) die Mittelwerte $\bar{x}$ und $\bar{y}$, die Varianzen s_x^2 und s_y^2, die Kovarianz c_{xy} und die Gleichung der Ausgleichsgeraden g: y = f(x) = ax + b.
Gib den Korrelationskoeffizienten r_{xy} an.

Nr. 2:

Gegeben sind die folgenden vier Punkte:

x-Werte	1	2	3	4
y-Werte	1	-1	0	1

a) Zeichne die vier Punkte in ein Koordinatensystem.

b) Bestimme die Mittelwerte $\bar{x}$ und $\bar{y}$, die Varianzen s_x^2 und s_y^2, die Kovarianz c_{xy} und die Gleichung der Ausgleichsgeraden g: y = f(x) = ax + b. Zeichne die Gerade g auch ins Koordinatensystem.
Gib den Korrelationskoeffizienten r_{xy} an.

c) Ist die Annahme eines **linearen** Zusammenhangs zwischen den x- und den y-Werten sinnvoll?

Lösungen

1. $\bar{x} = 2.5$ $\bar{y} = 2.5$ $s_x^2 = 4.333...$ $s_y^2 = 17.666...$ $c_{xy} = -8.666...$
   g: y = -2x + 7.5
   $r_{xy} = -0.9905$ (sehr gute Korrelation)

2. a) -
   b) $\bar{x} = 2.5$ $\bar{y} = 0.25$ $s_x^2 = 1.666...$ $s_y^2 = 0.91666...$ $c_{xy} = 0.1666...$
   g: y = 0.1x (b = 0)
   $r_{xy} = 0.1348$ (praktisch keine Korrelation)

   c) Die Annahme eines linearen Zusammenhangs ist zu verwerfen!

Arbeitsblatt 3 : Regression mit dem TI89

(Dieser Teil – auch die Aufgaben und die Lösungen - stammt von Walter Burgherr, Mathematiklehrer an der Kantonsschule Reussbühl. Er wurde von mir auf den TI89 angepasst)

Im Arbeitsblatt 3 bearbeiten wir Probleme der linearen Regression mit dem Taschenrechner. Dieser bietet ausserdem Möglichkeiten an, die Punkte (x_i / y_i) durch Graphen anderer Funktionstypen (quadratische Fkt., kubische Fkt., biquadratische Fkt., Potenzfkt., Exponentialfkt., logarithmische Fkt.) zu approximieren (erweiterte Regression).

Lernziele:

1. Du kannst Datenreihen mit Hilfe des Matrix-Editors in den TI-89 eingeben und editieren.
2. Aus den Listen der x- und y-Werte kannst du die Koeffizienten der Regressionsgeraden a und b, sowie den Korrelationskoeffizienten r ermitteln.
3. Du weisst, wie eine optimal annähernde Potenz- , Exponential- oder Logarithmusfunktion durch Zurückführen auf lineare Regression gewonnen wird.
4. Du kannst die entsprechenden Funktionsparameter mit dem Rechner bestimmen.
5. Du erkennst, welche der angebotenen Funktionstypen den Zusammenhang zwischen x- und y-Werten am besten wiedergibt.

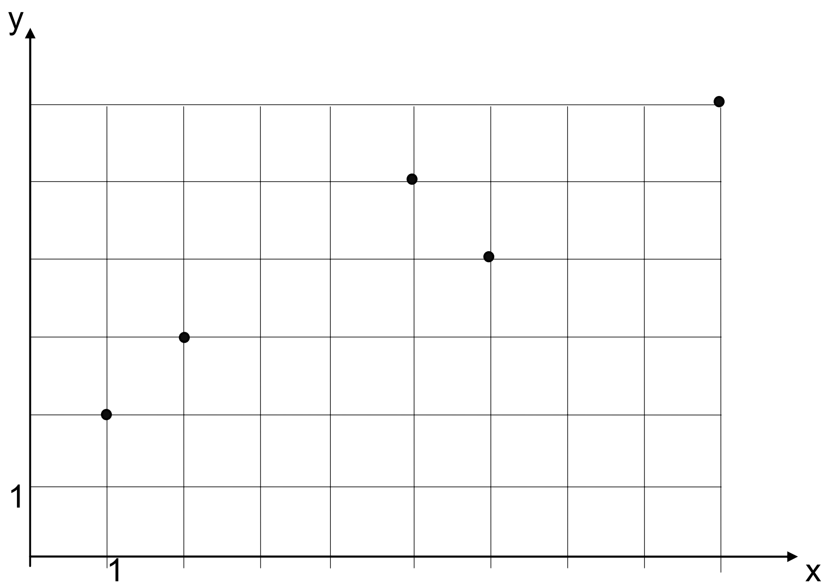
Vorgehen: Bearbeite für Dich das Arbeitsblatt 3: Regression mit dem Taschenrechner TI-89.
Lernkontrolle: Lösen der weiter unten angegebenen Aufgaben.

Lineare Regression

Beispiel: Es sind die folgenden fünf Punkte gegeben und im Koordinatensystem dargestellt.

x:	1	2	5	6	9
y:	2	3	5	4	6

Zeichne eine Gerade (provisorische Regressionsgerade), die deiner Meinung nach optimal an die Punkte angepasst ist (d.h. im Mittel möglichst nahe an die Punkte herankommt) und lies Steigung a^* und Achsenabschnitt b^* ab.



a* = a = r = a' = r' =
b* = b = b' =

Gib die x-Werte wie folgt in den TI89 ein:

wähle das Menu APPS (Applications)
wähle den Matrix-Editor
wähle New und dann Type Data
Gib einen Variablennamen ein, z.B. Re1
Gib nun die x Werte in Spalte c1 und die y-Werte in Spalte c2 ein

Wähle mit F5 das Untermenu CALC und aktiviere mit ENTER die Berechnung <LinReg>
Gib für x c1 und für y c2 ein.

Lies die Koeffizienten a, b der Regressionsgeraden y = ax + b und die Korrelation r ab und vergleiche mit a^* und b^*.
Zeichne die korrekte Regressionsgerade in die Figur ein.

Ändere den Punkt $P_3(5 / 5)$, indem du in der Spalte c_2 für y_3 den Wert 4 eingibst. Bestimme wiederum a', b' und r'. Wie haben sich Regressionsgerade und Korrelation geändert?

Merke: Die Korrelation liegt zwischen -1 und 1: $-1 \leq r \leq 1$
Je näher die Korrelation r bei 1 oder bei -1 liegt, desto besser beschreibt die lineare Funktion den Zusammenhang zwischen x- und y-Werten.

Regression mit Potenzfunktion $y = a\,x^b$

Logarithmiert man die Funktionsgleichung $y = a\,x^b$,
(Basis beliebig), so ergibt sich $\log y = \log a + b \log x$
Falls Punkte (x_i / y_i) auf der Potenzkurve liegen, so liegen die entsprechenden Punkte $(\log x_i / \log y_i)$ auf der Geraden mit Steigung b und Achsenabschnitt log a.
Die Menuwahl CALC/<PwrReg> veranlasst, dass der Rechner für die Logarithmen der gegebenen x- und y-Werte eine lineare Regression durchführt und daraus die Parameter der Potenzkurve a und b bestimmt.
Für die Korrelation r (der Logarithmen): siehe p.10 (*)

Regression mit Exponentialfunktion $y = a\, b^x$

Logarithmiert man die Funktionsgleichung $y = a\, b^x$

so ergibt sich $\log y = \log a + x \log b$

Falls Punkte (x_i / y_i) auf der Exponentialkurve liegen, so liegen die Punkte $(x_i / \log y_i)$ auf der Geraden mit Steigung log b und Achsenabschnitt log a.

Die Menuwahl CALC/<ExpReg> veranlasst, dass der Rechner für die gegebenen x-Werte und die Logarithmen der y-Werte eine lineare Regression durchführt und die Parameter der Exponentialkurve a und b angibt. Für die Korrelation r: siehe (*).

Regression mit Logarithmusfunktion $y = a + b \ln x$

Falls Punkte (x_i / y_i) auf einer solchen Logarithmuskurve liegen, so liegen die Punkte $(\ln x_i / y_i)$ auf der Geraden mit Steigung b und dem Achsenabschnitt a.

Die Menuwahl CALC/<LnReg> veranlasst, dass der Rechner für die natürlichen Logarithmen der x-Werte und die y-Werte eine lineare Regression durchführt und die Parameter der Logarithmuskurve a und b angibt.

(*) Der TI89 gibt den Korrelationskoeffizienten r nur bei linearer Regression an. Will man die entsprechenden r-Werte auch bei PwrReg, ExpReg und LnReg, so hat man die x-Werte, bzw. y-Werte zu logarithmieren und entsprechend den Modellen eine lineare Regression durchzuführen. Am einfachsten erstellt man dazu im Matrix-Editor eine Kolonne c3=ln(c1) und eine weitere Kolonne c4=ln(c2).

Polynomische Regression

Durch Punkte (x_i / y_i) wird eine möglichst gut approximierende quadratische, kubische oder Polynomfunktion höheren Grades gelegt. Das Kriterium für eine optimale Kurve ist dasselbe wie bei der linearen Regression (Arbeitsblatt 1). Die Koeffizienten werden mit Methoden der höheren Mathematik bestimmt. Der Rechner TI-89 liefert quadratische, kubische und Polynome vierten Grades. Menuwahl: <QuadReg>, <CubicReg> und <QuartReg>

Aufgabe: Führe mit den (ursprünglichen) Zahlen des obigen Beispiels logarithmische, Exponential- und Potenz-Regression durch und notiere die Ergebnisse

a_{li} =	a_{lo} =	a_{ex} =	a_{po} =
b_{li} =	b_{lo} =	b_{ex} =	b_{po} =
r_{li} =	r_{lo} =	r_{ex} =	r_{po} =

Welches Regressionsmodell hat die absolut grösste Korrelation?
Notiere die bestangepasste dieser Funktionsgleichungen

$y = f(x) =$

Gib ebenso die bestapproximierenden Polynome an

$y = P_2 (x) =$

$y = P_3 (x) =$

$y = P_4 (x) =$

Welches davon ist das beste? Erklärung!

Aufgaben zu Teil 3

Nr. 1:

Bei der adiabatischen Kompression eines fast idealen Gases werden Druck und Temperatur in einer Messreihe wie folgt ermittelt:

Druck p (x-Werte, in bar)	0.78	1.52	3.14	4.56	7.18	9.64
Temperatur T (y-Werte, in ° K)	269	350	465	539	644	723

Bestimme die Parameter a, b und die Korrelation r für lineare, logarithmische, exponentielle und Potenz-Regression.

$T = a * p + b$	$T = a + b * \ln p$	$T = a * b^p$	$T = a * p^b$
a_{li} =	a_{lo} =	a_{ex} =	a_{po} =
b_{li} =	b_{lo} =	b_{ex} =	b_{po} =
r_{li} =	r_{lo} =	r_{ex} =	r_{po} =

(siehe oben)

Welches ist das Modell mit der am nächsten bei 1 liegenden Korrelation?
Notiere die Funktionsgleichung der am besten angepassten Funktion.

$T = f(p) =$

Für ein ideales Gas gilt die Beziehung: $T^{\kappa} = \text{const} * p^{\kappa-1}$ (DMK/DPK Formeln und Tafeln, p. 154)
Bestimme κ (Kappa) aus der obigen Funktionsgleichung; um welches Gas könnte es sich handeln ? (Formelsammlung S. 173)

Nr. 2:

Der radioaktive Zerfall von Radium 224 erfolgt exponentiell. Eine Messung ergibt die folgenden Werte:

Zeit t (in Tagen)	0.322	1.184	2.321	3.165	5.430
Masse m (in Gramm)	2 .133	1.811	1.457	1.242	0.808

Bestimme die zugehörige Exponentialfunktion, die die Messfehler ausgleicht, durch exponentielle Regression. Notiere die Parameter a, b und die Korrelation r.

a =	a’ =
b =	b’ =
r =	r’ =

Berechne weiter die Werte $\ln y_i$ und führe für die Punkte $(x_i / \ln y_i)$ eine lineare Regression durch; Vergleiche die Parameter a’, b’, r’ mit den zuerst gefundenen.

Lösungen

1. $T = a * p^b = a * p^{(\kappa-1)/\kappa}$ mit a = 296.7 und b = 0.3931, r = 0.9999993 und
   $\kappa = 1 / (1-b) = 1.648$ (Argon)

2. $m = a * b^t$ mit a = 2.267 b = 0.8269 r = -0.9999988
   $\ln m = a' * t + b'$ mit a’= -0.19006 b’= 0.81846 r’= -0.9999988
   Der Vergleich liefert $a = e^{b'}$ $b = e^{a'}$ $r = r'$
   Das Vorzeichen von r ist durch die Steigung der Regressionsgeraden gegeben; diese fällt.