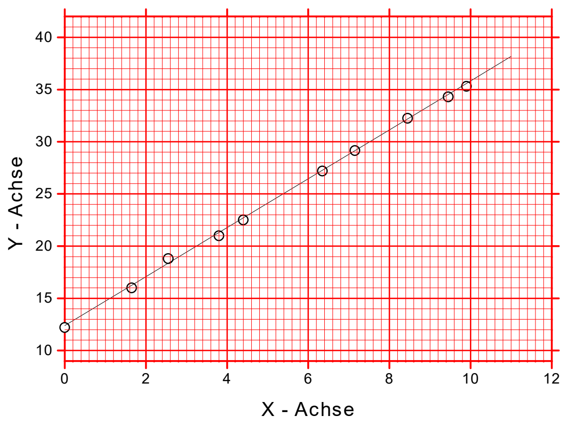
Graphische Auswerten:
Geraden

Ausgleichende Gerade

Auf wie viele Stellen soll die Steigung B angegeben werden?

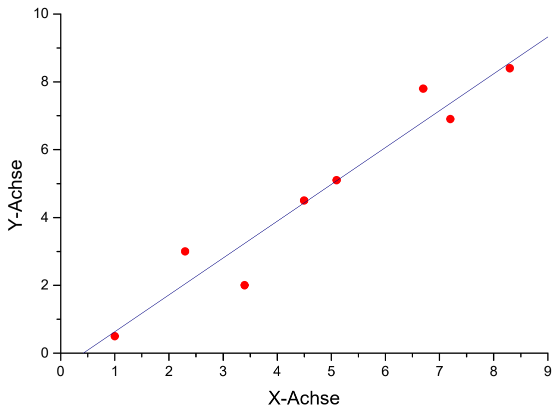
Im Folgenden werden drei Beispiele diskutiert, die zeigen, wie genau man die Steigung aus graphischen Darstellungen bestimmen kann:



a)
Die Daten streuen wenig.

Die Fehler der ein…
Datenpunkte …bekannt.

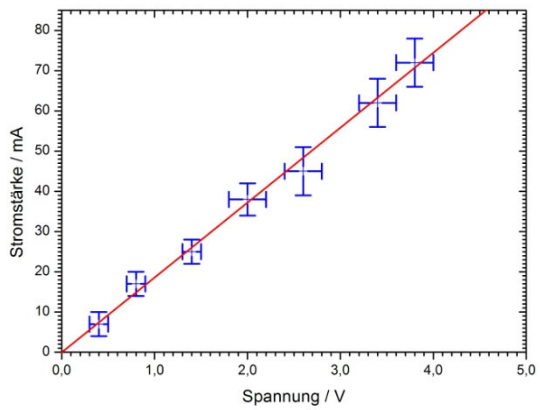
Dies wird sehr selten vorkommen.



b)
Die Daten streuen.

Die Fehler der ei…
Datenpunkt… …bekannt.

Dies ist im Praktikum in den meisten Fällen so.



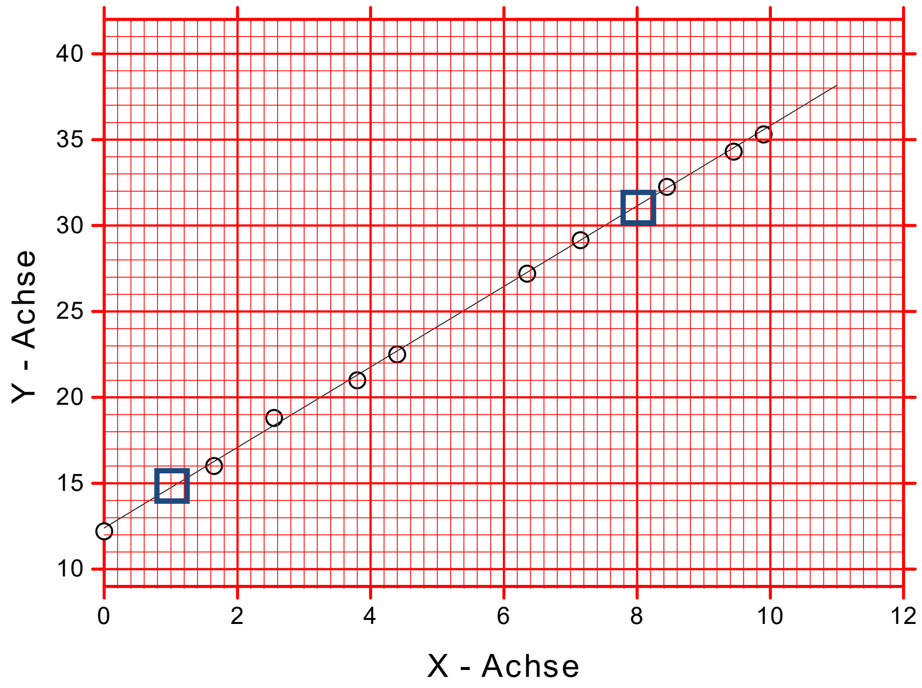
c)
Die Daten streuen…

Die Fehle…
Dat… …bekannt.

Dies wird in Veröffentlichungen und Übungsaufgaben vorkommen.

Ausgleichende Gerade: Fall a.)

- Steigung B und Achsenabschnitt A einer Geraden $Y = A + B\,X$ können graphisch leicht ermittelt werden.
- Hat man eine "mittlere, ausgleichende Gerade" durch die Punkteschar gelegt, dann entnimmt man der Zeichnung die Koordinaten zweier **weit auseinanderliegender** Punkte $P_a(X_a.Y_a)$ und $P_e(X_e.Y_e)$ und bestimmt die Steigung B zu:



$$B = \frac{Y_e - Y_a}{X_e - X_a}$$

$$B = \frac{31{,}2 - 14{,}8}{8{,}00 - 1{,}00} = \frac{16{,}4}{7{,}0} = 2{,}34286$$

$$A = 12{,}4$$

Auf wie viele Stellen sollen die Koeffizienten A und B angegeben werden ?

Ausgleichende Gerade: Fall a.)

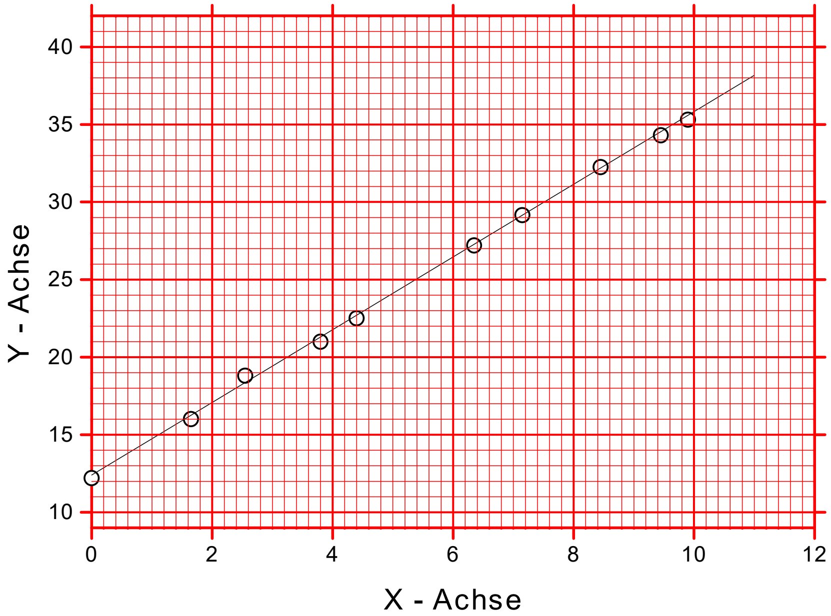
Auf wie viele Stellen soll die Steigung B angegeben werden?

Wenig streuende Daten; Fehler unbekannt.

Man gibt die Steigung auf soviel signifikante Stellen an,
dass bei der Berechnung eines *Y*-Wertes im Rahmen des Rundungsfehlers
der Steigung, die Ablesegenauigkeit widergespiegelt wird. (in diesem Beispiel etwa 0,1).

Beispiele für möglichst großes x : x = 9.
Aus der Graphik ist y zu 33,5 bestimmbar.
Ablesegenauigkeit etwa 0,1

(9; 33,5)



Ausgleichende Gerade: Fall a.)

Auf wie viele Stellen soll die Steigung B angegeben werden?

B=2,34286

Angabe von B auf zwei signifikante Stellen: B=2,3

y = 2,3 * 9 + 12,4 = 33,1

Schwankung der Steigung um einmal den Wert in der letzten signifikanten Stelle

y = 2,2 * 9 + 12,4 = 32,2 y = 2,4 * 9 + 12,4 = 34,0

Die Schwankung des y-Wertes, den man bei der Unsicherheit in B rechnerisch erhält.
ist größer als die Ablesegenauigkeit.

d.h. B ist in zu geringer Stellenzahl angegeben

Ausgleichende Gerade: Fall a.)

Auf wie viele Stellen soll die Steigung B angegeben werden?

B=2,34286

Angabe von B auf drei signifikante Stellen: B=2.34

y = 2,34 * 9 + 12,4 = 33,5

Schwankung der Steigung um einmal den Wert in der letzten signifikanten Stelle

y = 2,33 * 9 + 12,4 = 33,4 y = 2,35 * 9 + 12,4 = 33,6

Die Schwankung des y-Wertes, den man bei der Unsicherheit in B rechnerisch erhält.
ist etwa so groß wie die Ablesegenauigkeit.

d.h. B ist in vernünftiger Stellenzahl angegeben

Ausgleichende Gerade: Fall a.)

Auf wie viele Stellen soll die Steigung B angegeben werden?

B=2,34286

Angabe von B auf vier signifikante Stellen: B=2,343

y = 2,343 * 9 + 12,4 = 33,487

Schwankung der Steigung um einmal den Wert in der letzten signifikanten Stelle

y = 2,342 * 9 + 12,4 = 33,478   y = 2,344 * 9 + 12,4 = 33,496

Die Schwankung des y-Wertes, den man bei der Unsicherheit in B rechnerisch erhält.
ist kleiner als die Ablesegenauigkeit.

d.h. B ist mit zu vielen Stellen angegeben

Ablesegenauigkeit ist nicht gleich Messgenauigkeit

Wie groß ist der Fehler der Steigung bzw. des Achsenabschnittes ?

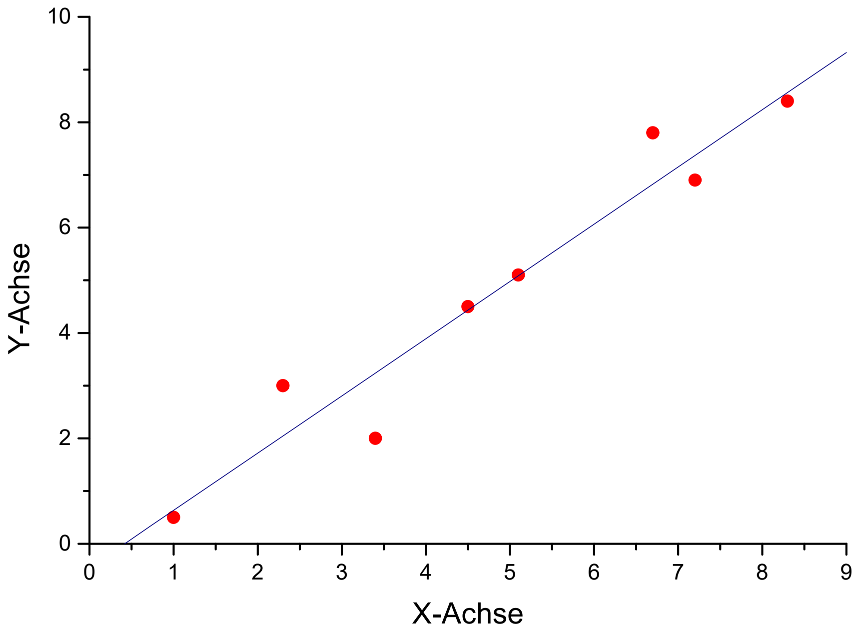
Ausgleichende Gerade: Fall b.)

Streuende Daten; Fehler unbekannt.

Messung

x	y
1,0	0,5
2,3	3,0
3,4	2,0
4,5	4,5
5,1	5,1
6,7	7,8
7,2	6,9
8,3	8,4

Graphische Darstellung

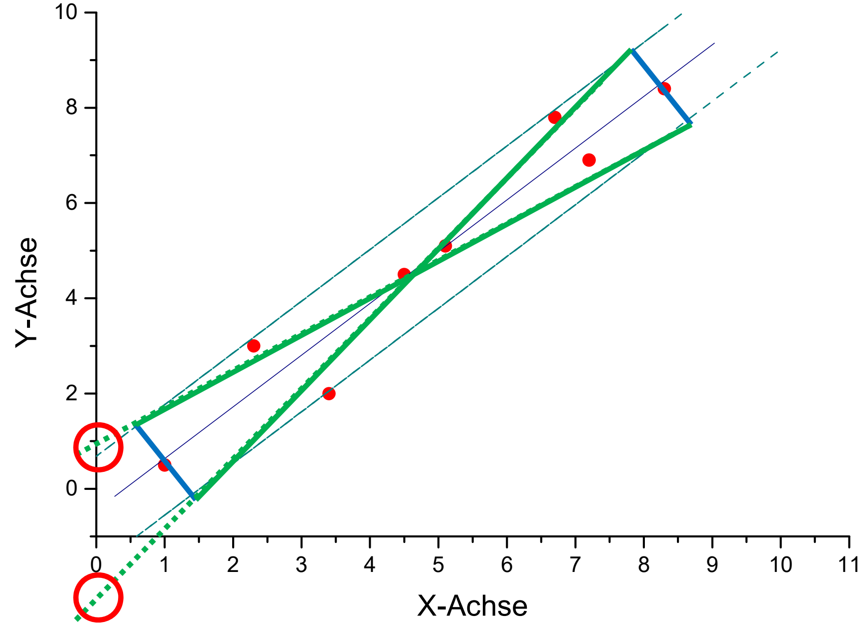


Anfertigung einer graphischen Darstellung inkl. ausgleichender Gerade

Wie groß ist der Fehler der Steigung bzw. des Achsenabschnittes ?

Ausgleichende Gerade: Fall b.)

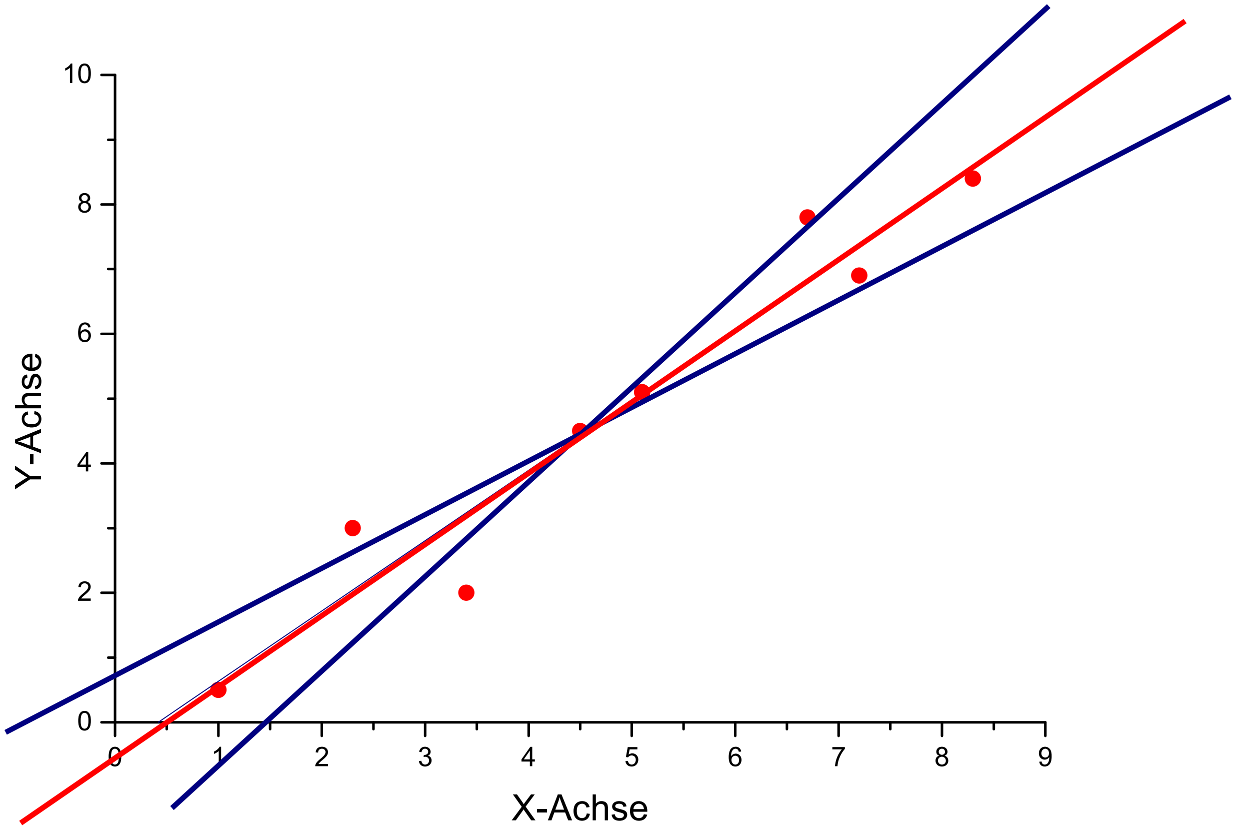
Streuende Daten; Fehler unbekannt.



- „Einzeichnen“ von zwei weiteren parallelen Geraden, die nach oben und unten verschoben sind (ca. 70% der Messpunkte sollten innerhalb liegen)
- „Einzeichnen“ eines Streubereichsrechteckes
- Die Diagonalen in diesem Rechteck liefern etwa den Fehler der Steigung sowie des Achsenabschnittes

Ausgleichende Gerade: Fall b.)

Streuende Daten; Fehler unbekannt.



Alle drei Geraden gehen durch den Schwerpunkt der Daten.

Dies ist in der Regel der Fall, wenn die Fehler aller Datenpunkte die gleiche Größe haben.

Ausgleichende Gerade: Fall c.)

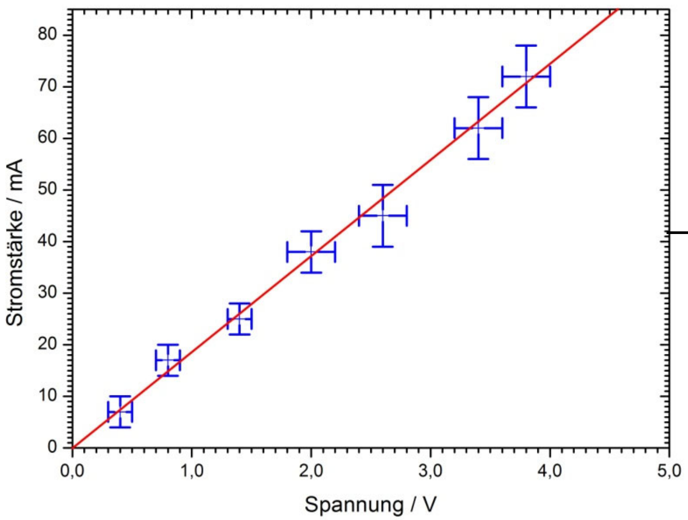
Streuende Daten; Fehler bekannt.

Messung der Kennlinie eines Ohmschen Widerstands

Messung

U / V	I / mA
$0{,}4 \pm 0{,}1$	7 ± 3
$0{,}8 \pm 0{,}1$	17 ± 3
$1{,}4 \pm 0{,}1$	25 ± 3
$2{,}0 \pm 0{,}2$	38 ± 4
$2{,}6 \pm 0{,}2$	45 ± 6
$3{,}4 \pm 0{,}2$	62 ± 6
$3{,}8 \pm 0{,}2$	72 ± 6

Graphische Darstellung



Auswertung

Optimale Gerade
durch den Nullpunkt

Anfangswert	Endwert
(0,00; 0,0)	(4,56; 84,83)

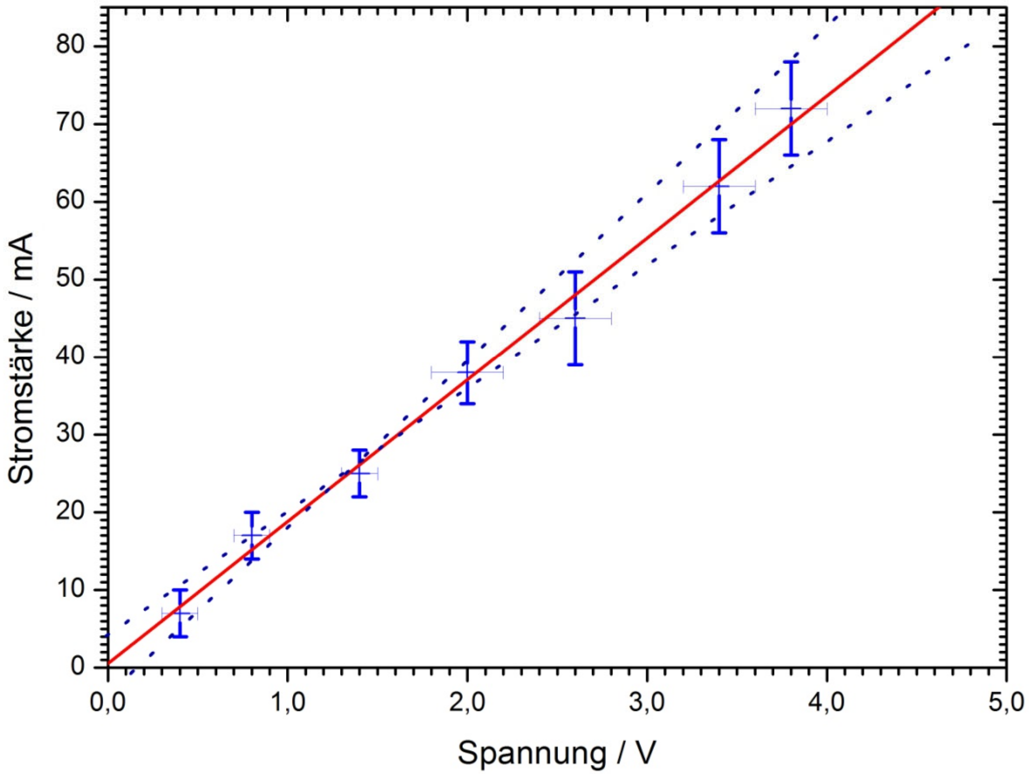
Widerstand 53,75 Ω

Der Nullpunkt wird aus physikalischen Gründen stark gewichtet.

<u>Es entscheidet immer die Messung</u>. Man kann nur prüfen, ob die Messung im Rahmen der Fehler mit der physikalischen Theorie im Einklang ist.

Ausgleichende Gerade: Fall c.)

Nimmt man realistischerweise an, dass auch kleine Spannungen und Ströme fehlerhaft gemessen werden, ist die starke Gewichtung des Nullpunktes nicht gerechtfertigt.



Auswertung

Optimale Gerade

Widerstand 54,92 Ω

Flache Gerade

Widerstand 63,81 Ω

Steile Gerade

Widerstand 44,82 Ω

$$R = (54{,}9 \pm 9{,}5)\ \Omega$$

Gewichtete Mittelwerte

Gewichtete Mittelwerte

Problem: Zusammenfassung von Messergebnissen

Zwei Arbeitsgruppen A und B mit unterschiedlichen Versuchsmethoden

$$X = X_A \pm \sigma_A$$
$$X = X_B \pm \sigma_B$$

Zusammenfassen zu einem Messwert.

Wenn x_A - x_B größer als die Summe der beiden **Standardabweichungen**,

dann einer der Messungen misstrauen.

Die Messergebnisse sind inkonsistent.

Annahme die Messwerte seien konsistent.

Welchen Wert besitzt der Bestwert?

Gewichtete Mittelwerte

Erster naheliegender Gedanke:

$$(x_A + x_B)/2$$

Ungeeignet, wenn die beiden Standardfehler unterschiedlich sind.
Bei obiger Mittelung haben beide Messergebnisse das gleiche Gewicht, obwohl bei einer der Messungen der Mittelwert genauer ist als bei der anderen

Annahme beide Messungen folgen einer Gauß-Verteilung

$$G_M(X_A) \propto \frac{1}{\sigma_A} e^{-\frac{(X_A - M)^2}{2\sigma_A^2}} \qquad G_M(X_B) \propto \frac{1}{\sigma_B} e^{-\frac{(X_B - M)^2}{2\sigma_B^2}}$$

Die Wahrscheinlichkeit, dass A den Messwert X_A **und** B den Wert X_B findet,

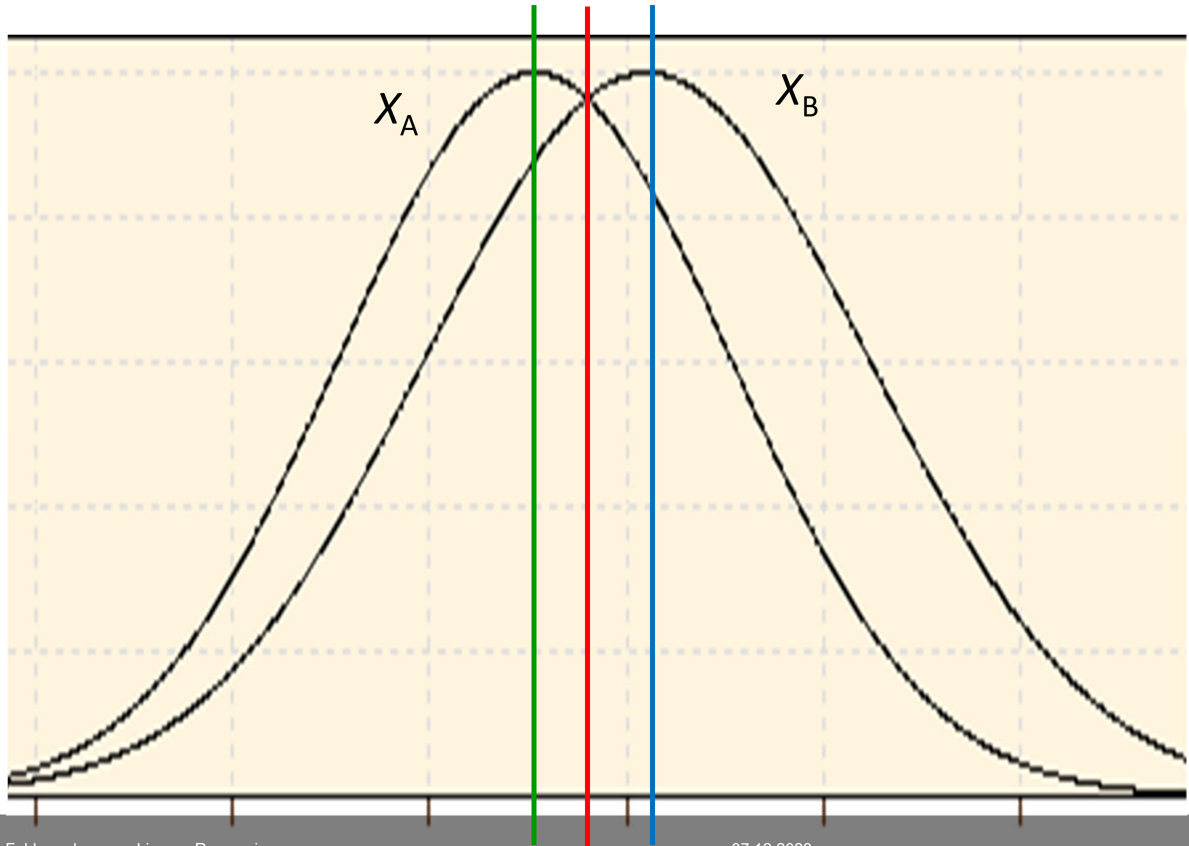
ist das Produkt der Wahrscheinlichkeiten.

Gewichtete Mittelwerte

Annahme beide Messungen folgen einer Gauß-Verteilung

$$G_M\left(X_A\right) \propto \frac{1}{\sigma_A} e^{-\frac{\left(X_A - M\right)^2}{2\sigma_A^2}} \qquad G_M\left(X_B\right) \propto \frac{1}{\sigma_B} e^{-\frac{\left(X_B - M\right)^2}{2\sigma_B^2}}$$

Die Wahrscheinlichkeit, dass A den Messwert X_A **und** B den Wert X_B findet, ist das Produkt der Wahrscheinlichkeiten.



Gewichtete Mittelwerte

Aus: $G_M(X_A) \propto \frac{1}{\sigma_A} e^{-\frac{(X_A - M)^2}{2\sigma_A^2}}$ und $G_M(X_B) \propto \frac{1}{\sigma_B} e^{-\frac{(X_B - M)^2}{2\sigma_B^2}}$

erhalten wir $G_{Mges} = G_M(X_A) \cdot G_M(X_B) \propto \frac{1}{\sigma_A \sigma_B} e^{-\frac{\chi^2}{2}}$

mit $\chi^2 = \left(\frac{X_A - M}{\sigma_A}\right)^2 + \left(\frac{X_B - M}{\sigma_B}\right)^2$

Die Abkürzung "CHI-Quadrat" ist die Summe der Quadrate der Abweichungen der beiden Messwerte von M,

jeweils dividiert durch die entsprechende Unsicherheit.

M ist der zu suchende Bestwert, dessen Wahrscheinlichkeit maximal ist,
oder was identisch ist, dessen χ^2 minimal ist.

Minimierung der "Summe der Quadrate" oder
Methode der kleinsten Quadrate

Gewichtete Mittelwerte

$$\chi^2 = \left(\frac{X_A - M}{\sigma_A}\right)^2 + \left(\frac{X_B - M}{\sigma_B}\right)^2$$

Differenzieren nach *M* und Ableitung gleich Null setzen.

$$\frac{\partial}{\partial M}\chi^2 = -2\frac{X_A - M}{\sigma_A^2} - 2\frac{X_B - M}{\sigma_B^2} = 0$$

Auflösen nach M ergibt den Bestwert:

$$M = x_{Best} = \frac{\frac{X_A}{\sigma_A^2} + \frac{X_B}{\sigma_B^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2}}$$

Übersichtlichere Schreibweise mit den Abkürzungen:

$$X_{Best} = \frac{w_A X_A + w_B X_B}{w_A + w_B} \qquad w_A = \frac{1}{\sigma_A^2} \quad und \quad w_B = \frac{1}{\sigma_B^2}$$

Allgemeine Formulierung

$$X_{Best} = \frac{\sum_{i=1}^{N} X_i w_i}{\sum_{i=1}^{N} w_i}$$

Gewichtete Mittelwerte - Beispiel

Messung des Ohm´schen Widerstandes durch 3 Studenten

(Ergebnisse in Ohm)

$\mathbf{R_1 = 11 \pm 1, \quad R_2 = 12 \pm 1, \quad R_3 = 10 \pm 3}$

$w_1 = 1$, $w_2 = 1$ und $w_3 = 1/9$

$$R_{Best} = \frac{\sum_{i=1}^{3} w_i R_i}{\sum_{i=1}^{3} w_i} = \frac{(1 \cdot 11) + (1 \cdot 12) + \left(\frac{1}{9} \cdot 10\right)}{\left(1 + 1 + \frac{1}{9}\right)} = 11.42\,\Omega$$

Wie groß ist der Fehler von R_{Best} ?

Gewichtete Mittelwerte - Beispiel

$$R_{Best} = \frac{w_1 R_1 + w_2 R_2 + w_3 R_3}{w_1 + w_2 + w_3}$$

$$\frac{\partial R_{Best}}{\partial R_i} = \frac{w_i}{w_1 + w_2 + w_3} \quad und \quad \sigma_{R_i}^2 = \frac{1}{w_i}$$

$$\sigma_{R_{Best}} = \sqrt{\left(\frac{\partial R_{Best}}{\partial R_1}\right)^2 \sigma_{R_1}^2 + \left(\frac{\partial R_{Best}}{\partial R_2}\right)^2 \sigma_{R_2}^2 + \left(\frac{\partial R_{Best}}{\partial R_3}\right)^2 \sigma_{R_3}^2}$$

$$= \sqrt{\left(\frac{w_1}{w_1+w_2+w_3}\right)^2 \frac{1}{w_1} + \left(\frac{w_2}{w_1+w_2+w_3}\right)^2 \frac{1}{w_2} + \left(\frac{w_3}{w_1+w_2+w_3}\right)^2 \frac{1}{w_3}}$$

$$= \sqrt{\frac{w_1 + w_2 + w_3}{(w_1 + w_2 + w_3)^2}}$$

Allgemeine Formulierung:

$$\boxed{\sigma_{Best} = \frac{1}{\sqrt{\sum w_i}}}$$

In unserem Beispiel: $\sigma_{Best} = (1+1+1/9)^{-1/2} = 0.69\ \Omega$

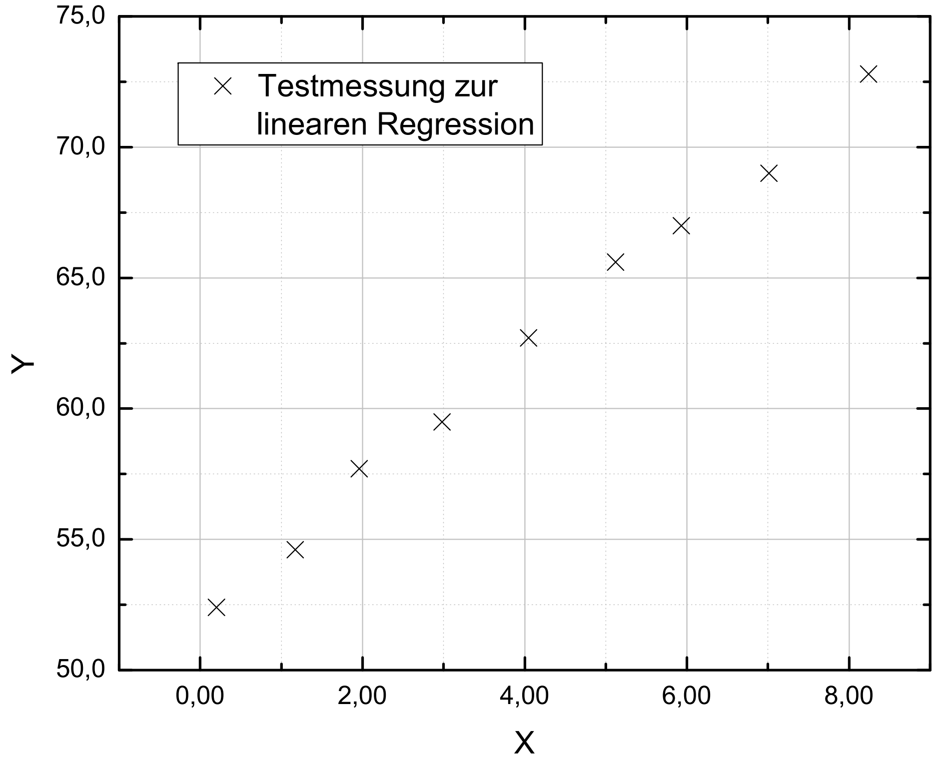
$$R = (11.42 \pm 0.69)\ \Omega$$

Lineare Regression

Fragestellung

Wir führen eine Messung durch und erhalten folgende Werte:

X_i	Y_i
0,20	52,4
1,17	54,6
1,96	57,7
2,98	59,5
4,05	62,7
5,12	65,6
5,93	67,0
7,01	69,0
8,24	72,8



1.) Wie können wir objektiv beurteilen, inwieweit die Daten unsere Annahme erfüllen und wirklich einer bestimmten Funktion (hier einer Gerade) folgen?

2.) Angenommen die Beziehung zwischen *x* und *y* ist linear, welche Gerade passt am besten zu den Messwerten?

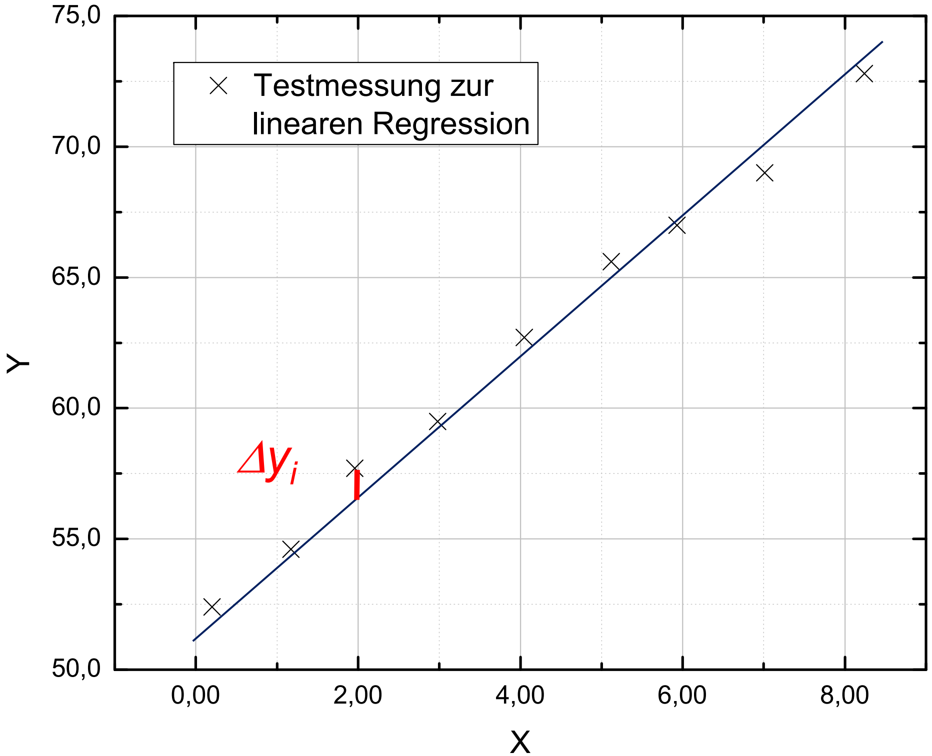
Vorüberlegungen

Wenn die Messwerte (X_i, Y_i) keinerlei Unsicherheit hätten,
dann läge jeder Punkt exakt auf der Geraden

Annahmen: Die Fehler in *X* sind wesentlich kleiner als die Fehler in *Y*.
X ist fehlerfrei und *Y* ist fehlerbehaftet.

Die Daten sind beschrieben durch den funktionellen Zusammenhang:

$$y = a + bx.$$



Die Abweichung Δy_i jedes Datenpunktes von der bestangepassten Geraden ist dann:

$$\Delta y_i = y_i - y(x_i) = y_i - a - bx_i.$$

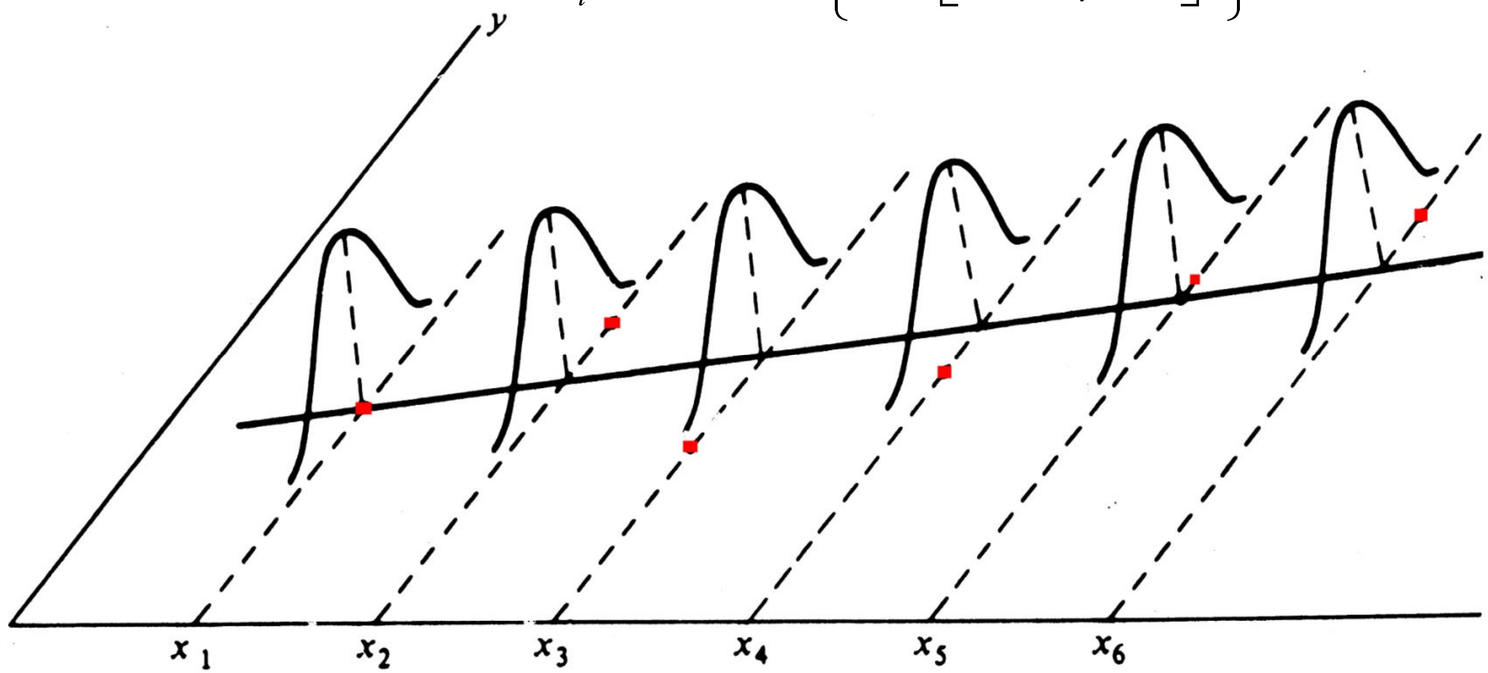
Vorüberlegungen

$$y \quad = \quad a + bx.$$

Für jedes X_i gibt es eine Wahrscheinlichkeit P_i den Messwert Y_i zu erhalten.

Normalverteilung der Messwerte:

$$P_i \quad = \quad \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right\}$$



Vorüberlegungen

$$P_i = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right\}$$

Die Wahrscheinlichkeit unseren Datensatz bei *N* Messpunkten genau so zu beobachten ist dann gegeben durch:

$$P(a_0, b_0) = \prod_{i=1}^{N} P_i = \prod_{i=1}^{N}\left(\frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right\}\right)$$

$$= \prod_{i=1}^{N}\left(\frac{1}{\sigma_i \sqrt{2\pi}}\right) \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right\}$$

Die beste Gerade liegt dann vor, wenn die Wahrscheinlichkeit maximal wird. Das ist der Fall, wenn der Exponent

$$\chi^2 = \sum_{i=1}^{N}\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2 = \sum_{i=1}^{N}\left[\frac{y_i - a - bx_i}{\sigma_i}\right]^2$$

minimal wird.

Lineare Regression - Bestwerte

Lineare Regression

oder

Anpassung einer Geraden nach der Methode der kleinsten Quadrate

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - y(x_i)}{\sigma_i} \right]^2 = \sum_{i=1}^{N} \left[\frac{y_i - a - bx_i}{\sigma_i} \right]^2$$

Für χ^2 minimal ist die Übereinstimmung am besten.

$$\frac{\partial}{\partial a} \chi^2 = 0;$$

$$\frac{\partial}{\partial b} \chi^2 = 0;$$

Lineare Regression - Bestwerte

$$\frac{\partial}{\partial a}\chi^2 = \frac{\partial}{\partial a}\sum_{i=1}^{N}\left[\frac{1}{\sigma_i^{\,2}}(y_i - a - bx_i)^2\right] = -2\sum_{i=1}^{N}\left[\frac{1}{\sigma_i^{\,2}}(y_i - a - bx_i)\right] = 0;$$

$$\frac{\partial}{\partial b}\chi^2 = \frac{\partial}{\partial b}\sum_{i=1}^{N}\left[\frac{1}{\sigma_i^{\,2}}(y_i - a - bx_i)^2\right] = -2\sum_{i=1}^{N}\left[\frac{x_i}{\sigma_i^{\,2}}(y_i - a - bx_i)\right] = 0;$$

An dieser Stelle ergibt sich ein wesentliches Ergebnis:

Falls alle Datenpunkte den gleichen Fehler haben, gilt:

$$b\sum x_i + a \cdot N = \sum y_i$$

$$b\frac{\sum x_i}{N} + a = \frac{\sum y_i}{N}$$

$$b \cdot \overline{x} + a = \overline{y}$$

Die beste Gerade geht durch den Mittelwert der x_i und y_i

Man spricht vom Schwerpunkt der Daten

Lineare Regression - Bestwerte

Zurück zu unserer ursprünglichen Frage. Wir suchen weiter die besten a und b:

$$\frac{\partial}{\partial a}\chi^2 = \frac{\partial}{\partial a}\sum_{i=1}^{N}\left[\frac{1}{\sigma_i^{\,2}}(y_i - a - bx_i)^2\right] = -2\sum_{i=1}^{N}\left[\frac{1}{\sigma_i^{\,2}}(y_i - a - bx_i)\right] = 0;$$

$$\frac{\partial}{\partial b}\chi^2 = \frac{\partial}{\partial b}\sum_{i=1}^{N}\left[\frac{1}{\sigma_i^{\,2}}(y_i - a - bx_i)^2\right] = -2\sum_{i=1}^{N}\left[\frac{x_i}{\sigma_i^{\,2}}(y_i - a - bx_i)\right] = 0;$$

Ein wenig geordnet gibt das:

$$\sum\frac{y_i}{\sigma_i^{\,2}} = a\sum\frac{1}{\sigma_i^{\,2}} + b\sum\frac{x_i}{\sigma_i^{\,2}};$$

$$\sum\frac{x_i y_i}{\sigma_i^{\,2}} = a\sum\frac{x_i}{\sigma_i^{\,2}} + b\sum\frac{x_i^{\,2}}{\sigma_i^{\,2}};$$

Der Übersichtlichkeit halber lassen wir ab hier die Summengrenzen weg.
(weiter über alle Datenpunkte von 1 bis N)

Lineare Regression - Bestwerte

$$\sum \frac{y_i}{\sigma_i^2} = a \sum \frac{1}{\sigma_i^2} + b \sum \frac{x_i}{\sigma_i^2};$$

$$\sum \frac{x_i y_i}{\sigma_i^2} = a \sum \frac{x_i}{\sigma_i^2} + b \sum \frac{x_i^2}{\sigma_i^2};$$

Mit der Determinantenmethode gibt das:

$$a = \frac{1}{\Delta} \begin{vmatrix} \sum \frac{y_i}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i y_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta} \left(\sum \frac{y_i}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} \right);$$

$$b = \frac{1}{\Delta} \begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{y_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i y_i}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta} \left(\sum \frac{1}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} \right);$$

$$\Delta = \begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2} \right)^2 .$$

Lineare Regression - Bestwerte

$$a = \frac{1}{\Delta}\begin{vmatrix} \sum \frac{y_i}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i y_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum \frac{y_i}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} \right);$$

$$b = \frac{1}{\Delta}\begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{y_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i y_i}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum \frac{1}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} \right);$$

$$\Delta = \begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2} \right)^2$$

Falls jeder Datenpunkt den gleichen Fehler aufweist, lässt sich dieses Gleichungssystem vereinfachen.

Das wird häufig, aber bei weitem nicht immer möglich sein.

Lineare Regression - Bestwerte

In unserem Beispiel habe jeder Datenpunkt den selben durch die Messung bedingten Fehler. Dann ergibt sich:

$$a = \frac{1}{\Delta^s}\begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix} = \frac{1}{\Delta}\left(\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i\right);$$

$$b = \frac{1}{\Delta^s}\begin{vmatrix} N & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix} = \frac{1}{\Delta}\left(N \sum x_i y_i - \sum x_i \sum y_i\right);$$

$$\Delta^s = \begin{vmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} = N \sum x_i^2 - \left(\sum x_i\right)^2.$$

Wir benötigen folgende Größen:

$$\boxed{\sum x_i^2 \quad , \quad \sum (x_i \cdot y_i) \quad , \quad \sum x_i \quad , \quad \sum y_i}$$

Lineare Regression - Bestwerte

Berechnen einer Ausgleichsgeraden: *x*-fehlerfrei, *y*-fehlerbehaftet

$$y \quad = \quad a + bx.$$

	x_i	y_i	y	$(y - y_i)^2$	x_i^2	$x_i \cdot y_i$
	0,20	**52,4**			0,040	10,480
	1,17	**54,6**				
	1,96	**57,7**				
	2,98	**59,5**				
	4,05	**62,7**				
	5,12	**65,6**				
	5,93	**67,0**				
	7,01	**69,0**				
	8,24	**72,8**			67,8976	599,872
Summe:	36,66	561,3			208,9504	2435,443

Lineare Regression - Bestwerte

x_i	y_i	$x_i \cdot x_i$	$x_i \cdot y_i$
0,20	52,4	0,0400	10,480
1,17	54,6	1,3689	63,882
1,96	57,7	3,8416	113,092
2,98	59,5	8,8804	177,310
4,05	62,7	16,4025	253,935
5,12	65,6	26,2144	335,935
5,93	67,0	35,1649	397,310
7,01	69,0	49,1401	483,690
8,24	72,8	67,8976	599,872
36,66	**561,3**	**208,9504**	**2435,443**

$$a = \frac{1}{\Delta^s}\begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix} = \frac{1}{\Delta}\left(\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i\right);$$

$$b = \frac{1}{\Delta^s}\begin{vmatrix} N & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix} = \frac{1}{\Delta}\left(N\sum x_i y_i - \sum x_i \sum y_i\right);$$

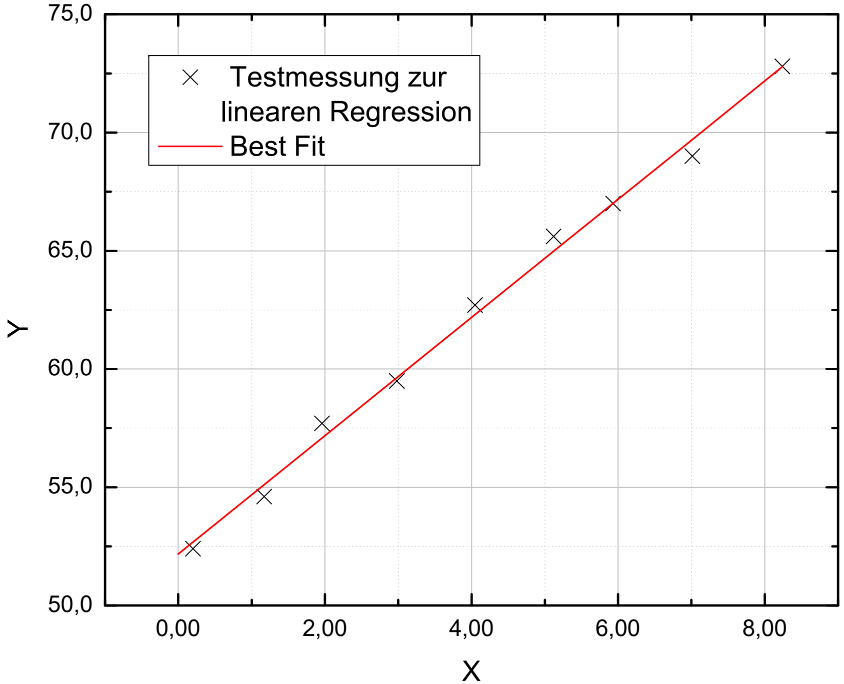
$$\Delta^s = \begin{vmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} = N\sum x_i^2 - \left(\sum x_i\right)^2 .$$

Lineare Regression - Bestwerte

$$b = \frac{9 \cdot 2435,443 - 36,66 \cdot 561,3}{9 \cdot 208,9504 - (36,66)^2} \qquad a = \frac{208,9504 \cdot 561,3 - 36,66 \cdot 2435,443}{9 \cdot 208,9504 - (36,66)^2}$$

$$= \frac{21918,987 - 20577,258}{1880,5536 - 1343,9556} \qquad = \frac{117283,8595 - 89283,34038}{1880,5536 - 1343,9556}$$

$$= \frac{1341,729}{536,598} = 2,500436081 \qquad = \frac{28000,51912}{536,598} = 52,18155699$$



$y = 2,500\,x + 52,18$

Damit haben wir die beste Gerade, die unsere Daten beschreibt.

ABER: Wie genau dürfen wir die Koeffizienten sinnvollerweise angeben?

Wie groß ist ihr Fehler?

Lineare Regression - Fehler

Klassisches Fehlerfortpflanzungsproblem: Wir haben eine Größe die sich aus mehreren fehlerbehafteten Einzelgrößen zusammensetzt.

In diesem Fall sind es unsere Koeffizienten, deren Fehler wir suchen.

Weitere Annahme: Unsere einzelnen Messpunkte sind statistisch voneinander unabhängig.

Dann können wir eine ganz normale Fehlerrechnung nach Gauß durchführen.

Es gilt:

$$\sigma_z^2 = \sum_k \left[\sigma_k^2 \left(\frac{\partial z}{\partial y_k} \right)^2 \right]$$

Die Annahme, dass unsere Messpunkte voneinander statistisch unabhängig sind, muss ebenfalls überprüft werden.

Dazu später mehr.

Lineare Regression - Fehler

$$\sigma_z^2 = \sum_k \left[\sigma_k^2 \left(\frac{\partial z}{\partial y_k}\right)^2\right];$$

Wir benötigen zunächst die partiellen Ableitungen:

$$a = \frac{1}{\Delta}\left(\sum \frac{y_i}{\sigma_i^2}\sum \frac{x_i^2}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2}\sum \frac{x_i y_i}{\sigma_i^2}\right);$$

$$b = \frac{1}{\Delta}\left(\sum \frac{1}{\sigma_i^2}\sum \frac{x_i y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2}\sum \frac{y_i}{\sigma_i^2}\right);$$

$$\Delta = \sum \frac{1}{\sigma_i^2}\sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2}\right)^2.$$

$$\frac{\partial a}{\partial y_k} = \frac{1}{\Delta}\left(\frac{1}{\sigma_k^2}\sum \frac{x_i^2}{\sigma_i^2} - \frac{x_k}{\sigma_k^2}\sum \frac{x_i}{\sigma_i^2}\right);$$

$$\frac{\partial b}{\partial y_k} = \frac{1}{\Delta}\left(\frac{x_k}{\sigma_k^2}\sum \frac{1}{\sigma_i^2} - \frac{1}{\sigma_k^2}\sum \frac{x_i}{\sigma_i^2}\right);$$

Lineare Regression - Fehler

$$\sigma_z^2 = \sum_k \left[\sigma_k^2 \left(\frac{\partial z}{\partial y_k}\right)^2\right]; \quad \frac{\partial a}{\partial y_k} = \frac{1}{\Delta}\left(\frac{1}{\sigma_k^2}\sum\frac{x_i^2}{\sigma_i^2} - \frac{x_k}{\sigma_k^2}\sum\frac{x_i}{\sigma_i^2}\right); \quad \Delta = \sum\frac{1}{\sigma_i^2}\sum\frac{x_i^2}{\sigma_i^2} - \left(\sum\frac{x_i}{\sigma_i^2}\right)^2.$$

Damit bestimmen wir zunächst den Fehler in a :

$$\sigma_a^2 = \sum_{k=1}^{N}\frac{\sigma_k^2}{\Delta^2}\left[\frac{1}{\sigma_k^4}\left(\sum\frac{x_i^2}{\sigma_i^2}\right)^2 - 2\frac{x_k}{\sigma_k^4}\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{x_i}{\sigma_i^2} + \frac{x_k^2}{\sigma_k^4}\left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]$$

$$= \frac{1}{\Delta^2}\left[\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\left(\sum\frac{x_i^2}{\sigma_i^2}\right)^2 - 2\sum_{k=1}^{N}\frac{x_k}{\sigma_k^2}\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{x_i}{\sigma_i^2} + \sum_{k=1}^{N}\frac{x_k^2}{\sigma_k^2}\left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]$$

$$= \frac{1}{\Delta^2}\left(\sum\frac{x_i^2}{\sigma_i^2}\right)\left[\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\sum\frac{x_i^2}{\sigma_i^2} - 2\sum_{k=1}^{N}\frac{x_k}{\sigma_k^2}\sum\frac{x_i}{\sigma_i^2} + \left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]$$

Beide Summen laufen von eins bis N.

$$= \frac{1}{\Delta^2}\left(\sum\frac{x_i^2}{\sigma_i^2}\right)\left[\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\sum\frac{x_i^2}{\sigma_i^2} - \left(\sum\frac{x_i}{\sigma_i^2}\right)^2\right]$$

$$= \frac{1}{\Delta}\left(\sum\frac{x_i^2}{\sigma_i^2}\right);$$

Lineare Regression - Fehler

$$\sigma_z^2 = \sum_k \left[\sigma_k^2 \left(\frac{\partial z}{\partial y_k}\right)^2\right]; \quad \frac{\partial b}{\partial y_k} = \frac{1}{\Delta}\left(\frac{x_k}{\sigma_k^2}\sum\frac{1}{\sigma_i^{\,2}} - \frac{1}{\sigma_k^2}\sum\frac{x_i}{\sigma_i^{\,2}}\right); \quad \underline{\Delta = \sum\frac{1}{\sigma_i^{\,2}}\sum\frac{x_i^{\,2}}{\sigma_i^{\,2}} - \left(\sum\frac{x_i}{\sigma_i^{\,2}}\right)^2}.$$

Analog bestimmen wir den Fehler in b :

$$\sigma_b^2 = \sum_{k=1}^{N}\frac{\sigma_k^2}{\Delta^2}\left[\frac{x_k^{\,2}}{\sigma_k^4}\left(\sum\frac{1}{\sigma_i^{\,2}}\right)^2 - 2\frac{x_k}{\sigma_k^4}\sum\frac{1}{\sigma_i^{\,2}}\sum\frac{x_i}{\sigma_i^{\,2}} + \frac{1}{\sigma_k^4}\left(\sum\frac{x_i}{\sigma_i^{\,2}}\right)^2\right]$$

$$= \frac{1}{\Delta^2}\left[\sum_{k=1}^{N}\frac{x_k^{\,2}}{\sigma_k^2}\left(\sum\frac{1}{\sigma_i^{\,2}}\right)^2 - 2\sum_{k=1}^{N}\frac{x_k}{\sigma_k^2}\sum\frac{1}{\sigma_i^{\,2}}\sum\frac{x_i}{\sigma_i^{\,2}} + \sum_{k=1}^{N}\frac{1}{\sigma_k^2}\left(\sum\frac{x_i}{\sigma_i^{\,2}}\right)^2\right]$$

$$= \frac{1}{\Delta^2}\left(\sum\frac{1}{\sigma_i^{\,2}}\right)\left[\sum_{k=1}^{N}\frac{x_k^2}{\sigma_k^2}\sum\frac{1}{\sigma_i^{\,2}} - 2\sum_{k=1}^{N}\frac{x_k}{\sigma_k^2}\sum\frac{x_i}{\sigma_i^{\,2}} + \left(\sum\frac{x_i}{\sigma_i^{\,2}}\right)^2\right]$$

Beide Summen laufen von eins bis N.

$$= \frac{1}{\Delta^2}\left(\sum\frac{1}{\sigma_i^{\,2}}\right)\left[\underline{\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\sum\frac{x_i^{\,2}}{\sigma_i^{\,2}} - \left(\sum\frac{x_i}{\sigma_i^{\,2}}\right)^2}\right]$$

$$= \frac{1}{\Delta}\left(\sum\frac{1}{\sigma_i^{\,2}}\right);$$

Lineare Regression - Fehler

Wir erhalten also allgemein folgende Fehler, die beide nicht von den y_i abhängen:

$$\sigma_a^2 = \frac{1}{\Delta}\left(\sum \frac{x_i^{\,2}}{\sigma_i^{\,2}}\right); \qquad \sigma_b^2 = \frac{1}{\Delta}\left(\sum \frac{1}{\sigma_i^{\,2}}\right);$$

$$\text{mit} \quad \Delta = \sum \frac{1}{\sigma_i^{\,2}} \sum \frac{x_i^{\,2}}{\sigma_i^{\,2}} - \left(\sum \frac{x_i}{\sigma_i^{\,2}}\right)^2 .$$

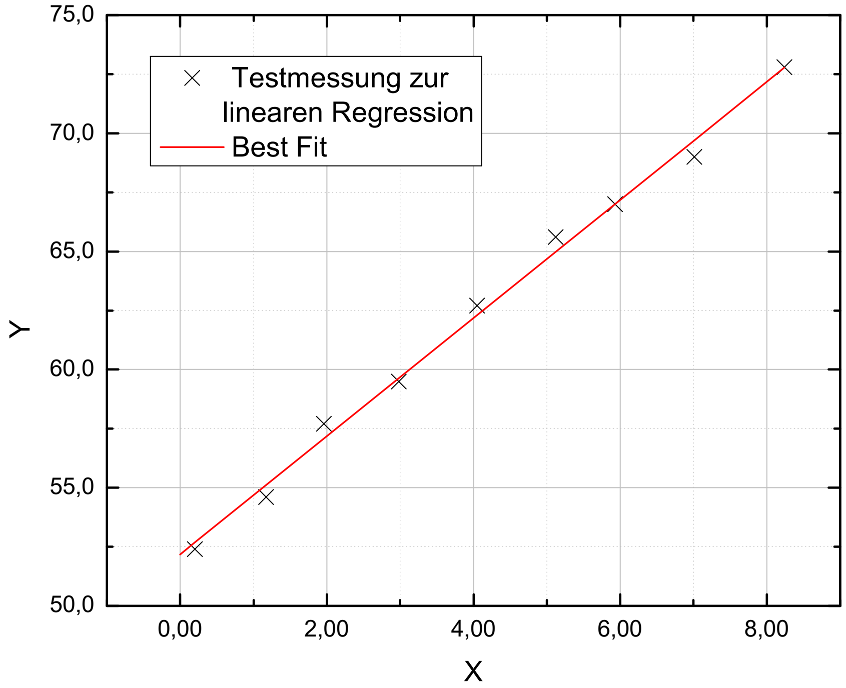
Falls jeder Datenpunkt den gleichen Fehler aufweist, lässt sich dies abermals vereinfachen.

$$\sigma_a^2 = \frac{\sigma^2}{\Delta^s} \sum x_i^{\,2}; \qquad \sigma_b^2 = N \frac{\sigma^2}{\Delta^s};$$

$$\text{mit} \quad \Delta^s = N \sum x_i^{\,2} - \left(\sum x_i\right)^2 .$$

Lineare Regression - Fehler

Zurück zu unserem Datensatz. Hier hatten wir gar keine Fehler angegeben. Ist die Angabe der Koeffizienten dann exakt?



Sicher nicht, die Daten streuen!

Die Streuung der Daten ist immer das Ergebnis der resultierenden Fehler. Wir können aus der Streuung einen <u>mittleren</u> Fehler schätzen!

Lineare Regression - Fehler

Es kann allgemein gezeigt werden, dass die Stichprobenvarianz für eine erwartungstreue Schätzung gegeben ist durch:

$$s^2 \equiv \frac{1}{N-c} \sum_i \left(y_i - \overline{y} \right)^2$$

Dabei ist N die Anzahl der beobachteten Klassen und c Anzahl der aus den Daten berechneten und/oder in der Rechnung verwendeten Parameter c.

Mit anderen Worten: im Nenner vor der Summe stehen die Freiheitsgrade

In unserem konkreten Fall passen wir die Daten an für:

$$y \quad = \quad a + bx.$$

Und damit:

$$s^2 = \frac{1}{N-2} \sum_i \left(y_i - a - bx_i \right)^2$$

Lineare Regression - Fehler

Konkret an unserem Anfangsbeispiel

Mittlere quadratische Abweichung

$$s^2 = \frac{1}{N-2}\sum_i (y_i - a - bx_i)^2$$

X_i	Y_i
0,20	52,4
1,17	54,6
1,96	57,7
2,98	59,5
4,05	62,7
5,12	65,6
5,93	67,0
7,01	69,0
8,24	72,8

$$y = 2{,}500\, x + 52{,}18$$

$$s^2 = \frac{1.7726}{7} = 0{,}2532$$

$$s = \pm\, 0{,}50$$

Wie geht dieser Fehler s in den Fehler von a bzw. b ein ?

Lineare Regression - Fehler

$$s^2 = 0,2532$$

$$\sigma_a^2 = \frac{\sigma^2}{\Delta^s} \sum x_i^{\,2}; \qquad \sigma_b^2 = N \frac{\sigma^2}{\Delta^s};$$

$$\text{mit} \quad \Delta^s = N \sum x_i^{\,2} - \left(\sum x_i\right)^2 .$$

Ergibt in unserem konkreten Fall:

$\boldsymbol{b}$ = 2,500 ± 0,065 (2,565 ; 2,435)

$\boldsymbol{a}$ = 52,18 ± 0,31 (52,49 ; 51,87)

Siehe dazu auch die heutige Übungsaufgabe!

Lineare Regression - Darstellung

Endergebnis:

$B = 2{,}500 \pm 0{,}065 \quad (2{,}565 \,;\, 2{,}435)$

$A = 52{,}18 \pm 0{,}31 \quad (52{,}49 \,;\, 51{,}87)$

Angabe der Geradengleichung ?

Zeichnen der Geraden ?

$$Y = (2{,}500 \pm 0{,}065)\, X + (52{,}18 \pm 0{,}31)$$

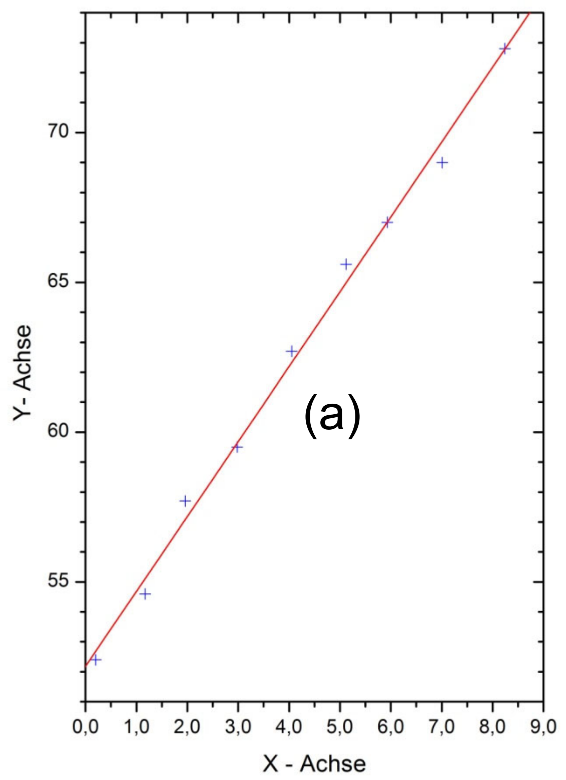
$$Y = 2{,}565\, X + 52{,}49$$

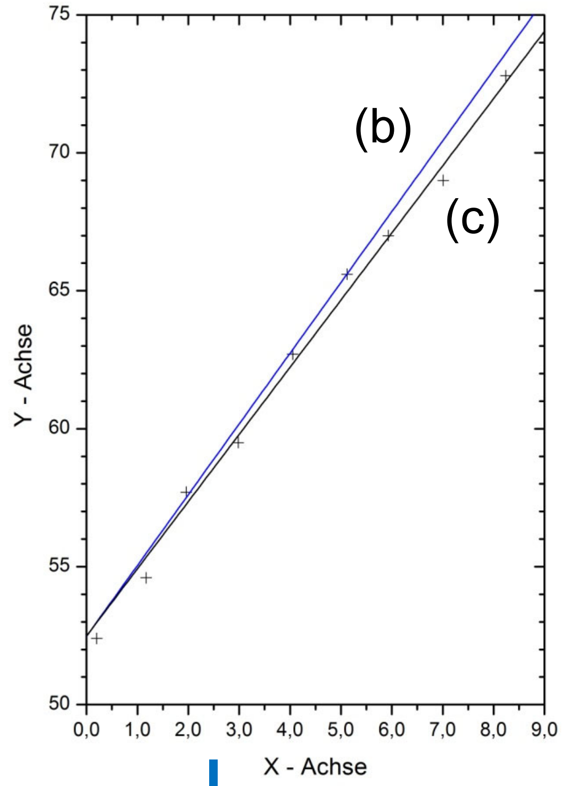
$$Y = 2{,}435\, X + 51{,}87$$

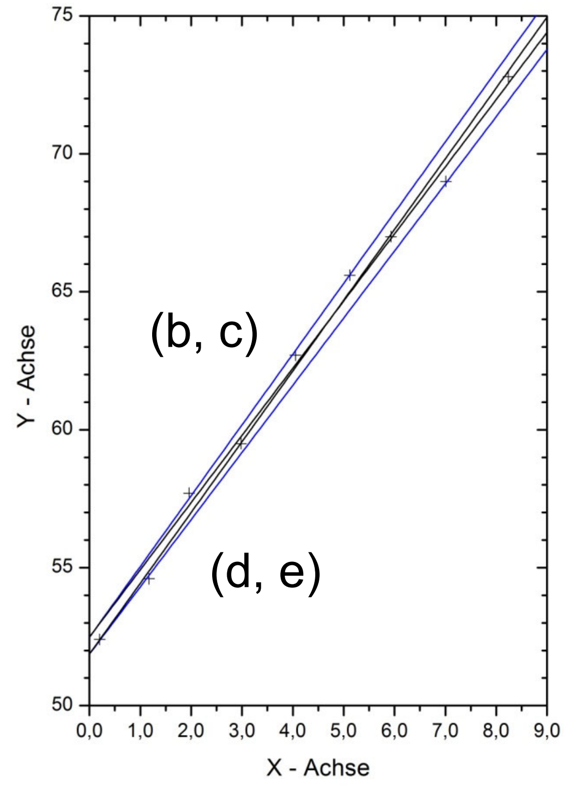
$$Y = 2{,}565\, X + 51{,}87$$

$$Y = 2{,}435\, X + 52{,}49$$

Inkorrekte Darstellung der „Fehlergeraden“







Optimale Gerade (a)

$Y = 2{,}500\ X + 52{,}18$

Geradengleichung mit Fehlerangabe

$Y = (2{,}500 \pm 0{,}065)\ X + (52{,}18 \pm 0{,}31)$

$Y = 2{,}565\ X + 52{,}49$ *(b)*

$Y = 2{,}435\ X + 51{,}87$ *(d)*

$Y = 2{,}565\ X + 51{,}87$ *(e)*

$Y = 2{,}435\ X + 52{,}49$ *(c)*

Kleiner Fehler bei *x*=0, großer bei *x*=9

Lineare Regression - Darstellung

Auszeichnung des Schwerpunktes:

$$b \sum x_i + a \cdot N = \sum y_i$$

$$b \frac{\sum x_i}{N} + a = \frac{\sum y_i}{N}$$

$$b \cdot \overline{x} + a = \overline{y}$$

vgl. folgende Graphen

$$\overline{x} = 4.073 \qquad \overline{y} = 62.367$$

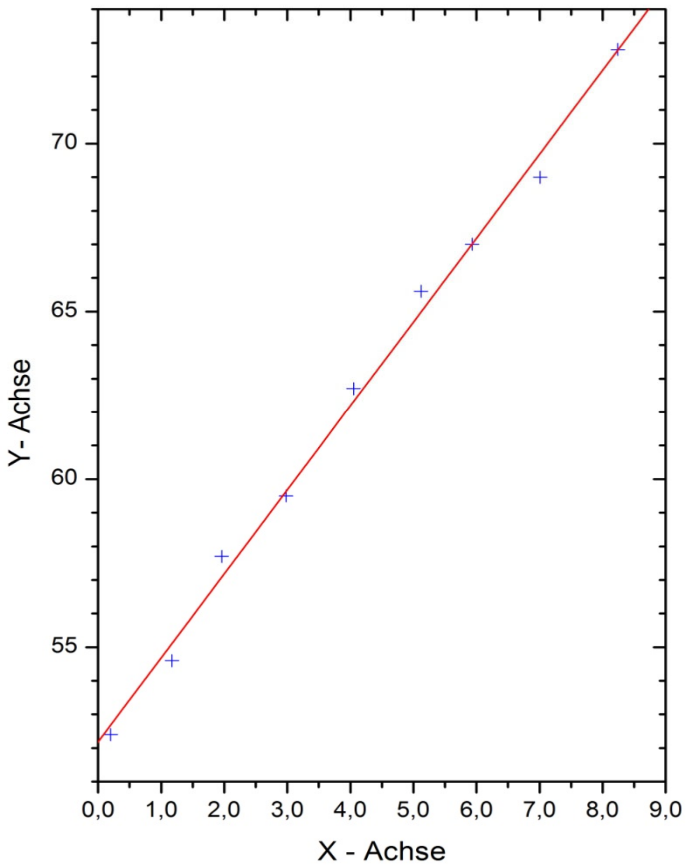
$$y = b\,x + a\,; \qquad \overline{y} = b\,\overline{x} + a$$

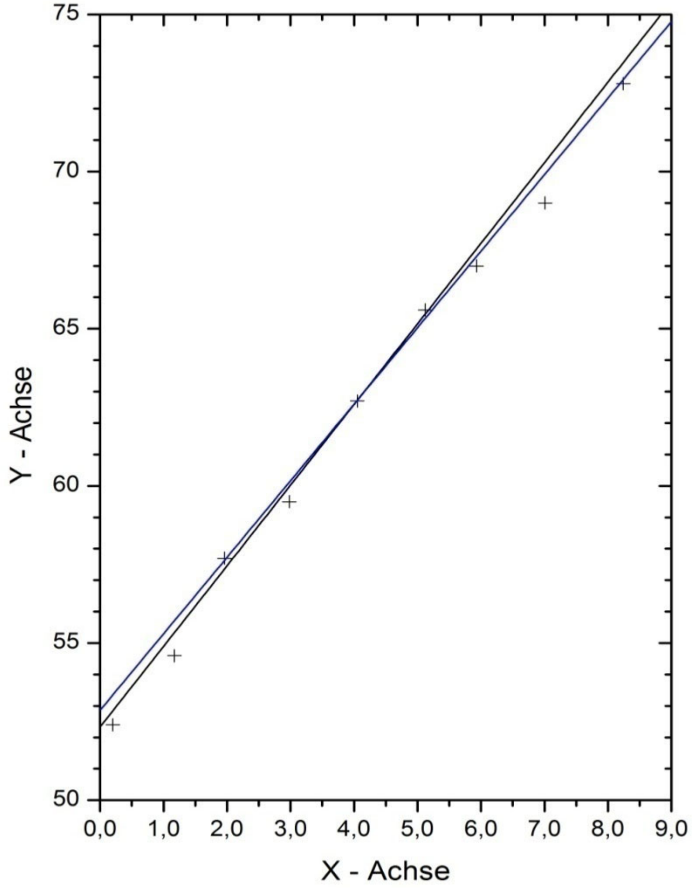
$$y = b\left(x - \overline{x}\right) + \overline{y}$$

$$y = \left(b \pm \Delta b\right) \cdot \left(x - \overline{x}\right) + \overline{y} \pm \Delta C$$

$$\Delta C = \sqrt{\overline{x}^2 \left(\Delta b\right)^2 + \left(\Delta a\right)^2} = 0.41$$

Lineare Regression – Auszeichnung des Schwerpunktes





$$\overline{x} = 4.073 \qquad \overline{y} = 62.367$$

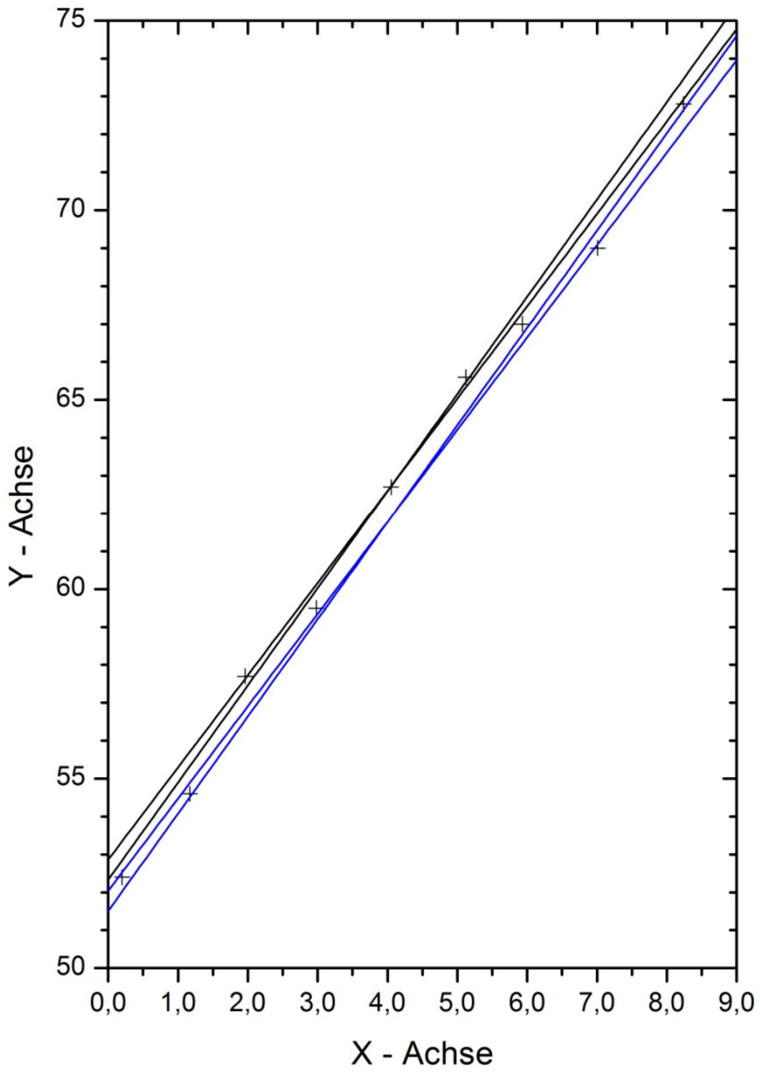
$$y = bx + a; \qquad \overline{y} = b\overline{x} + a$$

$$y = b\left(x - \overline{x}\right) + \overline{y}$$

$$y = b\left(x - \overline{x}\right) + \overline{y}$$

$$y = \left(b \pm \Delta b\right) \cdot \left(x - \overline{x}\right) + \overline{y}$$

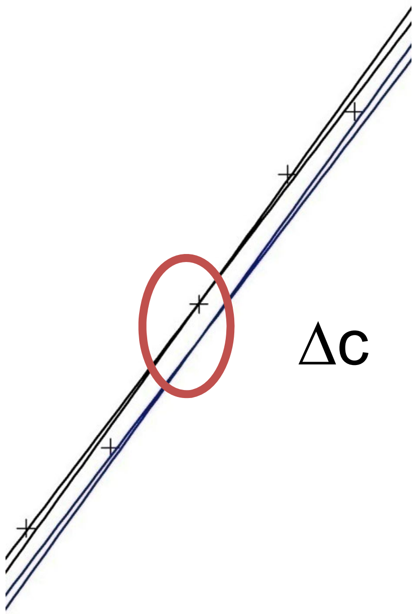
Lineare Regression – Auszeichnung des Schwerpunktes



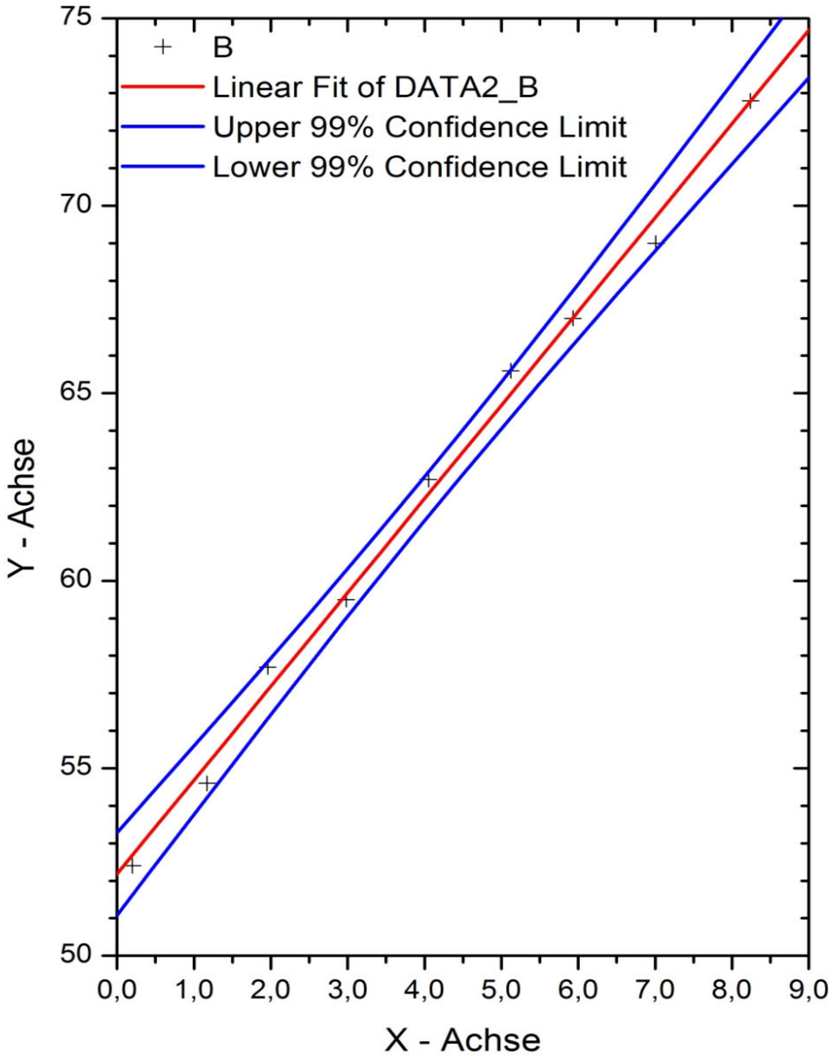
$$y = (b \pm \Delta b) \cdot (x - \overline{x}) + \overline{y} \pm \Delta C$$

$$\Delta C = \sqrt{\overline{x}^2 (\Delta b)^2 + (\Delta a)^2} = 0.41$$

Die Gerade dreht um den Schwerpunkt der Messungen, wobei der Fehler der Y-Werte den Fehler des Drehpunktes bestimmt.



Lineare Regression - Darstellung



Moderne Software zeigt nur die äußeren Teile der Geraden

Die Geraden drehen um den Schwerpunkt der Messung, wobei der Fehler des Y-Wertes den Fehler des Drehpunktes bestimmt.

Die beiden Kurven geben das Vertrauensintervall an, in dem die korrekte Gerade zu finden ist.

Neben A und B liefert der Taschenrechner auch einen so genannten Korrelationskoeffizienten r.