Bergische Universität GH Wuppertal FB Physik
Prof.Jacobi
Physikalisches Praktikum für Elektrotechniker

Messung physikalischer Größen und ihre Auswertung
lineare Regressionsanalyse (Methode der kleinsten Quadrate)

Aufnahme der Meßfehler, Fehlerfortpflanzung:

Jede gemessene physikalische Größe ist mit einem Fehler behaftet.
Dabei unterscheiden wir unvermeidliche **statistische Fehler**, die sich z.B. durch Vielfachmessung mit dem gleichen Meßgerät ermitteln lassen und **systematische Fehler**, die von der Güte des Meßverfahrens abhängen und nur durch genaue Analyse der Meßapparatur erfaßt werden können.
Die systematischen Fehler ΔX_i einer Meßgröße X_i sind abzuschätzen, das Fehlerfortpflanzungsgesetz für die **systematischen Fehler** lautet

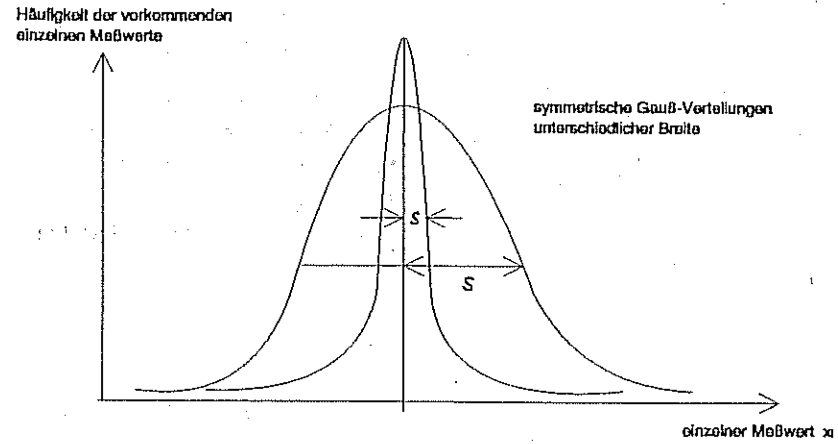
$$\Delta = \sum \left|\frac{\partial f_i}{\partial X_i}\right| \Delta X_i$$

Besteht die Funktion f_i z.B. aus zwei Meßgrößen x_i und y_i also $f(x_i, y_i)$, so würde das Fehlerfortpflanzungsgesetz für die systematischen Fehler lauten:

$$\Delta = \left|\frac{\partial f}{\partial x_i}\right| \cdot \Delta x + \left|\frac{\partial f}{\partial y_i}\right| \cdot \Delta y$$

Die systematischen Fehler können positives oder negatives Vorzeichen haben, d.h. sie können sich auch eventuell gegenseitig aufheben. Der systematische Gesamtfehler besitzt dann ein eindeutiges Vorzeichen. Der systematische Fehler der Funktion ist dann getrennt von dem statistischen Fehler anzugeben.

In den **statistischen Fehlern** können sich durchaus noch unbekannte systematische Fehler verborgen haben. Normalerweise offenbaren sich reine statistische Fehler, indem man die Meßwerte in einer Häufigkeitsverteilung darstellt und dann als Ergebnis eine **Gauß-Verteilung** erhält. Ist diese Gauß-Verteilung symmetrisch, so ist das Maximum dieser Verteilung der sogenannte **arithmetische Mittelwert**. Ist eine Messung zusätzlich mit unbekannten systematischen Fehlern behaftet, so erhält man beim Erstellen der Häufigkeitsverteilung keine symmetrische Verteilungsfunktion. Aus ihrer Unsymmetrie kann man Rückschlüsse auf die Größe noch unbekannter systematischer Fehler schließen. Dabei muß man sich aber das oben über systematische Fehler Gesagte vor Augen führen (mehrere systematische Fehler können sich addieren, aber auch gegenseitig aufheben). Desweiteren muß man auch bedenken, daß eine symmetrische Gauß-Verteilung infolge eines vorliegenden systematischen Fehlers auf der Meßwertachse auch parallel verschoben sein kann.



Der arithmetische Mittelwert $\bar{x}$ von einer Anzahl n einzelner Meßgrößen x_i wird berechnet nach der Gleichung: $\bar{x} = \frac{1}{n} \cdot \sum_i x_i$

Wenn eine schiefe Verteilung vorliegt, so ist das arithmetische Mittel nicht das Maximum der Verteilung !
Die Güte einer Messung zeigt sich in der Form (Breite) der Verteilungskurve:
Eine sehr schmale Verteilung bedeutet, daß die Meßwerte sehr nahe beieinander und im einzelnen auch nahe beim Mittelwert liegen, eine breite Verteilung sagt dagegen aus, daß die Einzelwerte weit um den Mittelwert streuen, also ein weniger gutes Meßverfahren vorliegt. Dabei sind, wie gesagt, noch keinerlei Aussagen über eventuell vorliegende systematische Fehler gemacht.
Die Güte wird mit dem halben Abstand der Wendepunkte der symmetrischen Gaußverteilung angegeben, man bezeichnet diese Größe auch als **Steuung** r oder **Standardabweichung** der Einzelwerte um den Mittelwert.

Es berechnet sich nach der Formel: $s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}$

wobei n die Anzahl der Einzelmessungen x_i und $\bar{x}$ das arithmetische Mittel ist.
Oft benutzt man für die Standardabweichung s auch den griechischen Buchstaben σ. Häufig wird auch das Quadrat der Streuung angegeben. Man nennt es **Varianz** $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$

Es sei hier darauf hingewiesen, daß die **Streuung** eine **Aussage über die Güte der Repräsentation der Einzelwerte durch den Mittelwert** macht, aber **nicht der Fehler des Mittelwertes** ist.
Der Fehler des Mittelwertes ergibt sich durch Anwenden des Fehlerfortpflanzungsgesetzes auf die Formel des Mittelwertes. Im Falle des arithmetischen Mittels führt das zur Gleichung für den **Fehler des arithmetischen Mittels** $s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n \cdot (n-1)}}$

Eine solche Fehlerangabe hat die gleiche Einheit wie der Meßwert selbst, man spricht daher vom absoluten statistischen Fehler der Meßgröße.
Wenn man einen systematischen Fehler in der Einheit der Meßgröße angibt, so ist das dann auch ein absoluter systematischer Fehler.
Man kann natürlich den absoluten Fehler auch auf die Meßgröße normieren und erhält dann den relativen Fehler der Meßgröße. Dieser ist einheitenlos. Man kan ihn in % durch Multiplikation mit 100 angeben:

$$s_{rel.} = \frac{s_{absolut}}{x}$$

Gewichtete Mittelwerte $\bar{\bar{x}}$:

Die Formel für die Wichtung der Mittelwerte lautet: $\bar{\bar{x}} = \frac{\sum_{i=1}^{n} p_i \cdot \bar{x}_i}{\sum_{i=1}^{n} p_i}$

Mißt man **mit gleichem Meßgerät**, nur jeweils **mit unterschiedlichen Häufigkeiten** n_i ,so erhält die Gleichung mit $p_i \equiv n_i$ das Aussehen

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{n} n_i \cdot \bar{x}_i}{\sum_{i=1}^{n} n_i}$$

Wenn die Streuungen σ für jeden Mittelwert andere sind, gilt wegen $p_i \equiv \frac{1}{\sigma_i^2}$

nun die Gleichung $\bar{\bar{x}} = \frac{\sum_{i=1}^{n} \frac{\bar{x}_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$

Das Fehlerfortpflanzugsgesetz FFG:
Ergibt sich die Notwendigkeit, mehrere fehlerbehaftete physikalische Größen in einer Größengleichung zu vereinen, so muß man das Fehlerfortpflanzungsgesetz anwenden.
Gesucht ist also der Mittelwert der Funktion und ihr Fehler, wenn die fehlerbehafteten Meßwerte, von der die Funktion abhängt, gegeben sind.
Im folgenden soll das **FFG** abgeleitet werden.
Eine Funktion, die z.B. von zwei Variablen abhängt, sei in einem dreidimensionalen Koordinatensystem dargestellt. Der jeweilige Meßwert $(\bar{x}_i, \bar{y}_i)$ befindet sich auf der Kurve $f(x_i, y_i)$. Ein Fehler des Funktionswertes stellt sich dabei so dar, als hätten wir eine geringfügige Abweichung von diesem Meßwert. Wie kann man diese Abweichung nun rechnerisch erfassen?
Verfolgen wir einmal eine geringfügige Abweichung in einer Richtung und sagen, daß die Abweichung in x- Richtung durch die Größe $(x+h)$ und die Abweichung in y- Richtung durch die Größe $(y+k)$ beschrieben ist. Die Änderung des Funktionswertes $f((x+h),(y+k))$ ergibt sich dann in erster Näherung über die Beziehung:

$$f((x+h),(y+k)) = f(x,y) + \left(\frac{\partial f(x,y)}{\partial x} \cdot h + \frac{\partial f(x,y)}{\partial y} \cdot k \right) + \ldots$$

Die Schreibweise $\frac{\partial f(x,y)}{\partial x}$ bedeutet, daß die Änderung der Funktion nach x gebildet und dabei y konstant gehalten wird (man nennt diese Form der Ableitung partielle Ableitung). Im folgenden wollen wir dafür eine abgekürzte Schreibweise benutzen: $\frac{\partial f(x,y)}{\partial x} = f_x$ bzw. $\frac{\partial f(x,y)}{\partial y} = f_y$
Da wir i x-Werte haben und h der jeweilige Abstand des i-ten x-Wertes vom Mittelwert $\bar{x}$ ist, schreiben wir $h_i = (x_i - \bar{x})$
Wenn wir nun den Mittelwert der Funktion berechnen wollen, müssen wir die Anzahl der x- Meßwerte und die Anzahl der y - Meßwerte kennen. Wir wollen die Anzahl der x-Meßwerte mit $i = r$ bezeichnen und ihre Summation über i mit der Summe $\sum_{i=1}^{r}$ beschreiben.
Entsprechendes gilt für die j y-Werte. Hier ist k der jeweilige Abstand des j-ten y-Wertes vom Mittelwert $\bar{y}$ und es sei $k_j = (y_j - \bar{y})$.
Die Anzahl der y-Meßwerte sei s und ihre Summation gehe über j mit der Summe $\sum_{j=1}^{s}$
Damit wird der Mittelwert der Funktion an der Stelle $(x+h),(y+k)$ zu:

$$\bar{f}((x+h),(y+k)) \approx \frac{1}{r \cdot s} \cdot \sum_{i=1}^{r} \sum_{j=1}^{s} \left(f(\bar{x},\bar{y}) + f_x \cdot h_i + f_y \cdot k_j \right)$$

Die Ausführung der Summation ergibt:

$\sum_{i=1}^{r}\sum_{j=1}^{s} f(\bar{x},\bar{y}) = r \cdot s \cdot f(\bar{x},\bar{y})$ (Doppelsumme über eine Konstante)

$\sum_{i=1}^{r}\sum_{j=1}^{s} f_x \cdot h_i = s \cdot \sum_{i=1}^{r} h_i \cdot f_x = 0$ (Mittelwertbedingung $\sum_i h_i = 0$)

$\sum_{i=1}^{r}\sum_{j=1}^{s} f_y \cdot k_j = r \cdot \sum_{j=1}^{s} k_j \cdot f_y = 0$ (Mittelwertbedingung $\sum_j k_j = 0$)

Die Ergebnisse zusammengefaßt ergeben: $\bar{f} = \frac{1}{r \cdot s} \cdot r \cdot s \cdot f(\bar{x},\bar{y})$

Der Mittelwert der Funktion ist gleich dem Funktionswert der Mittelwerte $\overline{f}(x,y) = f(\bar{x},\bar{y})$

Z.B. wird die physikalische Größengleichung für die Beschleunigung $a = \frac{2 \cdot s}{t^2}$, wenn Weg und Zeit gemessen werden und damit fehlerbehaftetet sind, zu $\bar{a} = \frac{2 \cdot \bar{s}}{\bar{t}^2}$ (mit den Mittelwerten von Weg $\bar{s}$ und Zeit $\bar{t}$)

Wie Sie sehen, multiplizieren wir Mittelwerte, die mit Fehlern behaftet sind. Wie sich die Fehler z.B. der Größe $\bar{a}$ berechnen sagt uns das **Fehlerfortpflanzugsgesetz FFG.**

Dabei erinnern wir uns an die Formel für die Standardabweichung für eine fehlerbehaftete Größe $s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}$

Im Zähler unter der Wurzel steht in der Summe die Abweichung des Einzelwertes vom Mittelwert. Wenn wir das entsprechend für unsere Funktion aufschreiben, wird $x_i \Rightarrow f_{ij}$ und $\bar{x} \Rightarrow \bar{f}$. Die Abweichung sei dann $v_{ij} = (f_{ij} - \bar{f})$ und die Quadratsumme der Abweichungen unter der Wurzel wird zu einer Doppelquadratsumme: $\sum_{i=1}^{r}\sum_{j=1}^{s} v_{ij}^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} (f_{ij} - \bar{f})^2$

Die Anzahl der Meßwerte n wird dann zu $r \cdot s$
(r x-Meßwerte **und** s y-Meßwerte)

Der Fehler der Funktion σ_f ist dann $\sigma_f = \sqrt{\frac{\sum_{i=1}^{r}\sum_{j=1}^{s} (f_{ij} - \bar{f})^2}{r \cdot s}}$

Mit $f_{ij} = f(\bar{x},\bar{y}) + f_x \cdot h_i + f_y \cdot k_j$ und $\bar{f} = f(\bar{x},\bar{y})$ erhalten wir für

$\sum_{i=1}^{r}\sum_{j=1}^{s} v_{ij}^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} (f(\bar{x},\bar{y}) + f_x \cdot h_i + f_y \cdot k_j - f(\bar{x},\bar{y}))^2$

Ausgerechnet ergibt das:

$$\sum_{i=1}^{r}\sum_{j=1}^{s} v_{ij}^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} (f_x \cdot h_i + f_y \cdot k_j)^2$$

$$\sum_{i=1}^{r}\sum_{j=1}^{s} v_{ij}^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} (f_x^2 \cdot h_i^2 + \underbrace{2 h_i k_j f_x f_y}_{=0} + f_y^2 \cdot k_j^2) = s \cdot \sum_{i=1}^{r} h_i^2 \cdot f_x^2 + r \cdot \sum_{j=1}^{s} k_j^2 \cdot f_y^2$$

Gehen wir mit allen unseren Anteilen in die Gleichung $\sigma_f = \sqrt{\frac{\sum_{i=1}^{r}\sum_{j=1}^{s} (f_{ij} - \bar{f})^2}{r \cdot s}}$

ein, so erhalten wir $\sigma_f = \sqrt{\frac{\sum_{i=1}^{r}\sum_{j=1}^{s} v_{ij}^2}{r \cdot s}} = \sqrt{\frac{s \cdot \sum_{i=1}^{r} h_i^2 \cdot f_x^2 + r \cdot \sum_{j=1}^{s} k_j^2 \cdot f_y^2}{r \cdot s}}$

da $\frac{\sum_{i=1}^{r} h_i^2}{r} = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2}{r} = \sigma_x^2$ und $\frac{\sum_{j=1}^{s} k_j^2}{s} = \frac{\sum_{j=1}^{s} (y_j - \bar{y})^2}{s} = \sigma_y^2$

die Varianzen der Meßwerte $\bar{x}$ und $\bar{y}$ sind (wir hatten sie zu Anfang mit dem Buchstaben s_x^2 bzw. s_y^2 bezeichnet), ergibt sich die Formel des **FFG** für zwei Variable zu

$$\sigma_f = \sqrt{\sigma_x^2 \cdot f_x^2 + \sigma_y^2 \cdot f_y^2}$$

(Die σ_x und σ_y sind die Fehler der Meßgrößen $\bar{x}$ und $\bar{y}$ und f_x bzw f_y die partiellen Ableitungen der Funktion). Bei Funktionen von mehr als zwei Variablen ist die Formel des FFG entsprechend $\sigma_f = \sqrt{\sum_{i=1}^{n} \sigma_i^2 \cdot f_i^2}$ zu interpretieren, wobei dann der Index i von 1 bis n die Anzahl der fehlerbehafteten Meßvariablen darstellen soll.

Danach erhalten wir z.B. den Fehler der Funktion $f(\bar{x},\bar{y})$ mit den fehlerbehafteten Meßwerten $\bar{x} \pm s_x$ *und* $\bar{y} \pm s_y$

$$s_f = \sqrt{s_x^2 \cdot \left(\frac{\partial f}{\partial \bar{x}}\right)^2 + s_y^2 \cdot \left(\frac{\partial f}{\partial \bar{y}}\right)^2}$$

Dabei stehen unter der Wurzel jeweils die Quadrate der absoluten Fehler der einzelnen Meßgrößen, die mit den Quadraten der partiellen Ableitung der Funktion nach dem Mittelwert der betreffenden Meßgröße multipliziert werden.

Wir wollen hier kurz das FFG zur Bestimmung des Fehlers einer Summe oder Differenz bzw. eines Produktes oder Quotienten anwenden:

In einem Praktikumsversuch sind die zur Bestimmung der Fahrstrecke der Gleiter erforderlichen Ablesungen des Start und Zielpunktes auf der Luftkissenschiene fehlerbehaftet: $\bar{l} \pm s_l = (\bar{x}_2 - \bar{x}_1) \pm s_l$

$$s_l = \sqrt{s_{x_2}^2 + s_{x_1}^2}$$

Der mittlere Fehler einer Summe oder einer Differenz zweier fehlerbehafteter Größen ist gleich der pythagoreischen Summe der mittleren Fehler der Summanden.

Für die **Bestimmung des Fehlers eines Produktes oder Quotienten**, wählen wir die Gleichung für die Beschleunigung $\bar{a} = \frac{2 \cdot \bar{l}(t)}{\bar{t}^2}$ (für die Strecke s haben wir, um Verwechselungen mit der Standardabweichung zu vermeiden l eingesetzt) $\quad \bar{a} \pm s_a = \frac{2 \cdot \bar{l}}{(t)^2} \pm s_a$

$$s_{\bar{a}} = \sqrt{s_l^2 \cdot \left(\frac{2}{(\bar{t})^2}\right)^2 + s_t^2 \cdot \left(\frac{4 \cdot \bar{l}}{(\bar{t})^3}\right)^2}$$

denn die partiellen Ableitungen nach $\bar{l}$ bzw. nach $\bar{t}$ ergiben: $\frac{\partial \bar{a}(\bar{l},\bar{t})}{\partial \bar{l}} = \frac{2}{\bar{t}^2}$

und $\quad \frac{\partial \bar{a}(\bar{l},\bar{t})}{\partial \bar{t}} = -\frac{4\bar{l}}{\bar{t}^3}$

In diesem Falle vereinfacht sich die Gleichung indem wir zum relativer Fehler übergehen:

$$s_{rel.a} = \frac{s_{\bar{a}}}{\bar{a}} = \sqrt{\left(\frac{s_l}{\bar{l}}\right)^2 + 4\left(\frac{s_t}{\bar{t}}\right)^2}$$

Für alle Produkte und Quotienten gilt:
Der relative mittlere Fehler eines Produktes oder eines Quotienten zweier fehlerbehafteter Größen ist gleich der pythagoreischen Summe der relativen mittleren Fehler der Faktoren.

Bestimmen der besten Geraden mit Hilfe der linearen Regression

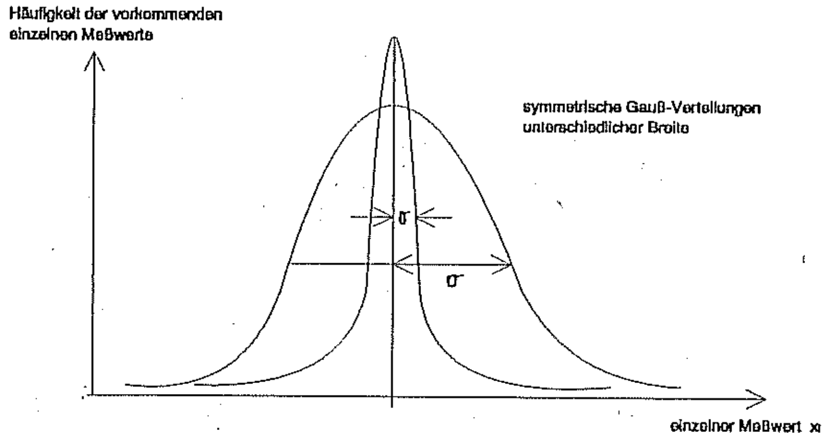
Wir gehen von der Geradengleichung $y_i = b \cdot x_i + c$ aus und nehmen an, daß dabei y_i die fehlerbehafteten Meßgrößen, während die x_i parametrisch vorgegeben sind. Natürlich sind in der Praxis auch die x_i fehlerbehaftet, aber in nur ganz wenigen Fällen findet sich die Ungenauigkeit des vorgegebenen Parameters nicht im Fehler irgendeiner Messgröße wieder:

Die mathematische Gleichung für die Gauß'sche Glockenkurve lautet:

$$P(x_i;\bar{x},\sigma) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x_i - \bar{x})^2}{2\sigma^2}}$$

$P(x_i;\bar{x},\sigma)$ nennen wir die Wahrscheinlichkeit oder Häufigkeit mit der die einzelnen Meßwerte x_i, die sich zum arithmetischen Mittelwert $\bar{x} = \frac{1}{n} \cdot \sum_i x_i$ ergeben. σ ist die Standardabweichung, die wir nach der Gleichung $\sigma = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}$ berechnen können. n ist die Anzahl der einzelnen Meßwerte x_i

Die grafische Darstellung der symmetrischen Gauß-Verteilungsfunktion (Gauß-Kurve) ergibt sich wie folgt. Die Standardabweichung findet sich in der halben Breite der Verteilungsfunktion wieder (halber Abstand der Wendepunkte).



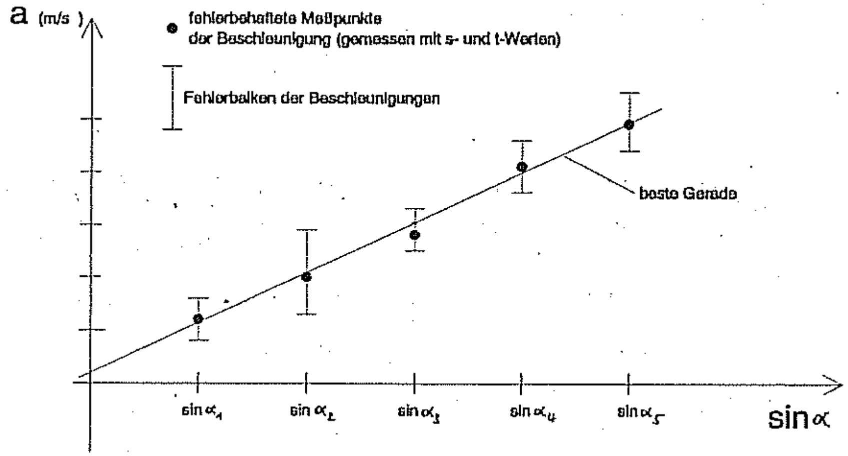
Die **bei der Verteilungsfunktion gesuchte Variable** ist der Maximalwert, der bei einer symmetrischen Verteilungsfunktion **der arithmetische Mittelwert** $\bar{x}$ ist, was sich beweisen läßt, indem man die eindimensionale Gauß-Funktion nach $\bar{x}$ differenziert und das Ergebnis Null setzt (Extremwertaufgabe!).

$$\frac{\partial P(x_i;\bar{x},\sigma)}{\partial \bar{x}} = -2 \cdot \frac{(x_i - \bar{x})}{2\sigma^2} \cdot \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x_i - \bar{x})^2}{2\sigma^2}} = 0$$

Wie man sieht, braucht man nur den Exponenten der e-Funktion zu differenzieren, denn alleine die Ableitung des Exponenten (die innere Ableitung) bleibt durch das Nullsetzen stehen:

$$\frac{\partial\left(-\frac{(x_i-\bar{x})^2}{2\sigma^2}\right)}{\partial \bar{x}} = -2\cdot\frac{(x_i-\bar{x})}{2\sigma^2} = 0 \quad \Rightarrow \quad x_{i\,max} = \bar{x}$$

Bei dem Verfahren der linearen Regression geht man in der gleichen Weise vor:
Man möchte für eine Anzahl von Meßwerten y_i, die zweidimensional verteilt sind - d.h. sie finden sich in einer grafischen Darstellung auf einer Geraden $y_i = b\cdot x_i + c$ wieder (z.B. wie in einem Praktikumsversuch, wo die Meßwerte y_i die Beschleunigungswerte a_i sind) - den besten Achsenabschnitt c und den besten Anstieg b ermitteln.



Nach den Gesetzen der Wahrscheinlichkeit findet man diese Werte; indem man das Maximum einer zwidimensionalen symmetrischen Gauß-Verteilungsfunktion sucht. Diese Verteilungsfunktion ist die Produktwahrscheinlichkeit der beiden Einzelwahrscheinlichkeiten für den Anstieg b und den Achsenabschnitt c:

$$P(b,c) = \prod_{j=1}^{2} P_i(y_i;b,c,\sigma) = \left(\prod_{i=1}^{n}\frac{1}{\sigma_i\sqrt{2\pi}}\right)\cdot e^{-\sum_{i=1}^{n}\frac{(y_i-(bx_i+c))^2}{2\sigma_i^2}}$$

Diese Funktion ist nun wie bei der Ermittlung des arithmetischen Fehlers, also bei der Suche des Maximums der eindimensionalen Verteilung, nach den Variablen b und c (bester Anstieg und bester Achsenabschnitt) partiell zu differenzieren und Null zu setzen (Maximum der zweidimensionalen

Verteilungsfunktion suchen). Hierbei erinnern wir uns gerne an das, was wir hierzu bei der eindimensionalen Verteilung gesagt haben: wir brauchen nur den Exponenten der e-Funktion partiell nach b und c zu differenzieren. Wir erhalten dabei zwei Gleichungen, die wir Null setzen:

$$\frac{\partial\left(-\sum_{i=1}^{n}\frac{(y_i-(b\cdot x_i+c))^2}{2\sigma_i^2}\right)}{\partial b} = \sum_{i=1}^{n}\frac{(y_i - b\cdot x_i - c)\cdot x_i}{\sigma_i^2} = 0 \quad \text{und}$$

$$\frac{\partial\left(-\sum_{i=1}^{n}\frac{(y_i-(b\cdot x_i+c))^2}{2\sigma_i^2}\right)}{\partial c} = \sum_{i=1}^{n}\frac{(y_i - b\cdot x_i - c)}{\sigma_i^2} = 0$$

Wir erhalten die beiden linearen Gleichungen für die Unbekannten b und c:

$$\sum_{i=1}^{n}\frac{x_i\cdot y_i}{\sigma_i^2} = b\cdot\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} + c\cdot\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} \quad \text{und}$$

$$\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2} = b\cdot\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} + c\cdot\sum_{i=1}^{n}\frac{1}{\sigma_i^2}$$

Wir lösen das lineare Gleichungssystem nach der Determinantenmethode (Cramer'sche Regel):
Für die Nennerdeterminante erhalten wir:

$$\Delta^* = \begin{vmatrix} \sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{n}\frac{1}{\sigma_i^2} \end{vmatrix} = \sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\cdot\sum_{i=1}^{n}\frac{1}{\sigma_i^2} - \left(\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right)^2$$

und für die Zählerdeterminante erhalten wir für die Berechnung von b

$$Z_b = \begin{vmatrix} \sum_{i=1}^{n}\frac{x_i\cdot y_i}{\sigma_i^2} & \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^{n}\frac{y_i}{\sigma_i^2} & \sum_{i=1}^{n}\frac{1}{\sigma_i^2} \end{vmatrix} = \sum_{i=1}^{n}\frac{x_i\cdot y_i}{\sigma_i^2}\cdot\sum_{i=1}^{n}\frac{1}{\sigma_i^2} - \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\cdot\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}$$

bzw. für die Zählerdeterminante für die Berechnung von c

$$Z_c = \begin{vmatrix} \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{n} \frac{x_i \cdot y_i}{\sigma_i^2} \\ \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} & \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} \end{vmatrix} = \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} \cdot \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i \cdot y_i}{\sigma_i^2} \cdot \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}$$

Als besten Anstieg erhalten wir

$$b = \frac{Z_b}{\Delta^*} \quad \text{also} \quad b = \frac{1}{\Delta^*} \cdot \left(\sum_i \frac{1}{(\sigma_i)^2} \cdot \sum_i \frac{x_i \cdot y_i}{(\sigma_i)^2} - \sum_i \frac{x_i}{(\sigma_i)^2} \cdot \sum_i \frac{y_i}{(\sigma_i)^2} \right)$$

und der beste Achsenabschnitt ist:

$$c = \frac{Z_c}{\Delta^*} \quad \text{also} \quad c = \frac{1}{\Delta^*} \cdot \left(\sum_i \frac{x_i^2}{\sigma_i^2} \cdot \sum_i \frac{y_i}{\sigma_i^2} - \sum_i \frac{x_i}{\sigma_i^2} \cdot \sum_i \frac{x_i \cdot y_i}{\sigma_i^2} \right)$$

mit $\Delta^* = \sum_i \frac{1}{\sigma_i^2} \cdot \sum_i \frac{x_i^2}{\sigma_i^2} - \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2$

Zu diesen Endgleichungen ist noch einmal folgendes zu sagen:
Die Größen x_i und y_i sind als Mittelwerte anzunehmen und σ_i^2 ist definitionsgemäß die Varianz (hier das Fehlerquadrat) der fehlerbehafteten Größe y_i .Wenn nun in einer Messung der Fall vorliegt, daß die Meßfehler (Streuungen der Mittelwerte y_i) für alle y_i gleich sind, so vereinfachen sich die Gleichungen zu:

$$b = \frac{1}{\Delta} \cdot \left(n \cdot \sum_i x_i \cdot y_i - \sum_i x_i \cdot \sum_i y_i \right) \text{ und } c = \frac{1}{\Delta} \cdot \left(\sum_i x_i^2 \cdot \sum_i y_i - \sum_i x_i \cdot \sum_i x_i \cdot y_i \right)$$

mit $\Delta = n \cdot \sum_i x_i^2 - \left(\sum_i x_i \right)^2$ denn $\sum_i \frac{1}{\sigma^2} = n \cdot \frac{1}{\sigma^2}$ wenn σ konstant ist

n ist dabei die Gesamtzahl aller Meßwerte y_i
Der Anstieg und der Achsenabschnitt sind natürlich fehlerbehaftet. Ihre Fehler berechnen sich durch Anwendung des FFG auf obige Gleichungen. Man erhält für den Fall, daß alle Meßwerte verschiedene Fehler haben

für den Fehler des Achsenabschnittes $s_c = \sum_{i=1}^{n} \left(\frac{\partial c}{\partial y_i} \right)^2 \cdot \sigma_i^2$

und ausgeführter Differentiation:

$$s_c^2 = \frac{1}{\Delta^{*2}} \cdot \left((\sum_i \frac{x_i^2}{(\sigma_i)^2})^2 \cdot \sum_i \frac{1}{(\sigma_i)^2} - (\sum_i \frac{x_i}{(\sigma_i)^2})^2 \cdot \sum_i \frac{x_i^2}{(\sigma_i)^2} \right)$$

und für den Fehler des Anstieges: $s_b = \sum_{i=1}^{n} \left(\frac{\partial b}{\partial y_i} \right)^2 \cdot \sigma_i^2$

und ausgeführter Differentiation:

$$s_b^2 = \frac{1}{\Delta^{*2}} \cdot \left((\sum_i \frac{1}{(\sigma_i)^2})^2 \cdot \sum_i \frac{x_i^2}{(\sigma_i)^2} - (\sum_i \frac{x_i}{(\sigma_i)^2})^2 \cdot \sum_i \frac{1}{(\sigma_i)^2} \right)$$

mit $\Delta^* = \sum_i \frac{1}{(\sigma_i)^2} \cdot \sum_i \frac{x_i^2}{(\sigma_i)^2} - \left(\sum_i \frac{x_i}{(\sigma_i)^2} \right)^2$

Für den Fall, daß alle Meßwerte den gleichen Fehler haben, ist $\sigma_i = \sigma$ und es ergibt sich für den Achsenabschnitt:

$$s_c^2 = \frac{1}{\Delta^2} \cdot \left(\frac{1}{\sigma^6} \cdot (N \cdot (\sum_{i=1}^{N} x_i^2)^2 - (\sum_{i=1}^{N} x_i)^2 \cdot \sum_{i=1}^{N} x_i^2) \right)$$

und für den Anstieg:

$$s_b^2 = \frac{1}{\Delta^2} \cdot \left(\frac{1}{\sigma^6} \cdot (N^2 \cdot \sum_{i=1}^{N} x_i^2 - N \cdot (\sum_{i=1}^{N} x_i)^2 \right)$$

mit N **als Anzahl der Meßwertpaare** (y_i, x_i)

und $\Delta = \frac{1}{\sigma^4} \cdot \left(N \cdot \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \right)$

Die Größe σ ist aber jetzt die Streuung der Meßwertepaare von der Geraden, die wir nach der Gleichung $\sigma^2 = \frac{1}{N-2} \cdot \sum_{i=1}^{N} (y_i - b \cdot x_i - c)^2$ berechnen. Der Vorfaktor $\frac{1}{N-2}$, trägt dem Rechnung, daß die Gerade durch 2 Meßpunkte bestimmt ist, daß also nur noch $N - 2$ Freiheitsgrade zur Disposition stehen. Wenn also $N = 2$ ist, so wird der Fehler ∞.

Rechnen wir die Fehler des Anstieges und des Achsenabschnittes aus:

$$s_c^2 = \frac{\sigma^2 \cdot \left(N \cdot (\sum_{i=1}^{N} x_i^2)^2 - (\sum_{i=1}^{N} x_i)^2 \cdot \sum_{i=1}^{N} x_i^2 \right)}{\frac{1}{\sigma^4} \cdot \left(N \cdot \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \right)^2} \quad \text{für den Achsenabschnitt und}$$

$$s_b^2 = \frac{\sigma^4 \cdot \left(N^2 \cdot \sum_{i=1}^{N} x_i^2 - N \cdot (\sum_{i=1}^{N} x_i)^2 \right)}{\frac{1}{\sigma^4} \cdot \left(N \cdot \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \right)^2} \quad \text{für den Anstieg}$$

Bearbeiten wir diese Gleichungen noch ein wenig, um sie zu vereinfachen:

Wenn wir in der Gleichung für s_c^2 die die Summe $\sum_{i=1}^{N} x_i^2$ ausklammem und kürzen ,ergibt sich $s_c^2 = \dfrac{\sigma^2 \cdot \sum_{i=1}^{N} x_i^2}{N \cdot \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$ und wenn wir in der Gleichung für s_b^2 im Zähler N ausklammem und kürzen, erhalten wir für

$$s_b^2 = \frac{\sigma^2 \cdot N}{N \cdot \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$

Nun stellt sich für uns noch die Frage, ob der Fehler von Anstieg und Achsenabschnitt vom Nullpunkt der Verteilungsfunktion, also von einer Koordinatenursprungs-Transformation, abhängt.
Legen wir einmal den Koordinatenursprung in das Maximum unserer Verteilungsfunktion. Damit sind die x_i links und rechts vom Koordinatenursprung völlig gleichverteilt, so daß damit $\sum_{i=1}^{N} x_i = 0$ wird. Wir erhalten für den Fehler des Achsenabschnittes $s_c^2 = \dfrac{\sigma^2 \cdot \sum_{i=1}^{N} x_i^2}{N \cdot \sum_{i=1}^{N} x_i^2} = \dfrac{\sigma^2}{N}$

σ^2 und N sind von der Lage des Koordinatensystems unabhäng. Dieser Fehler ist also eine von der Lage des Koordinatensystems unabhängige Größe. Wie sieht es aber mit dem Fehler des Anstieges aus? Wenn wir $\sum_{i=1}^{N} x_i = 0$ setzen, erhalten wir $s_b^2 = \dfrac{\sigma^2 \cdot N}{N \cdot \sum_{i=1}^{N} x_i^2} = \dfrac{\sigma^2}{\sum_{i=1}^{N} x_i^2}$ was, wie man sieht, von der Lage des Koordinatenursprunges nicht unabhängig ist.
Wir lösen dieses Problem, indem wir statt $\sum_{i=1}^{N} x_i^2$ nun $\sum_{i=1}^{N} (x_i - \overline{x})^2$ schreiben.

Damit erhalten wir für den Fehler des Anstieges $s_b^2 = \dfrac{\sigma^2}{\sum_{i=1}^{N} (x_i - \overline{x})^2}$

Eine etwas andere Methode, um den besten Anstieg und den besten Achsenabschnitt zu bestimmen:
Wenn man eine Gleichverteilung der Meßwertepaare um die beste Gerade $\overline{y} = b \cdot \overline{x} + c$ annimmt und der Fehler der Messung die Streuung σ der Meßwerte um diese Gerade ist, kann man auch bei der Herleitung der

Gleichungen für den besten Anstieg und den besten Achsenabschnitt mit Mittelwerten von x_i und y_i arbeiten:
Man bestimmt die Mittelwerte aus den N Wertepaaren gemäß

$$\overline{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i \quad \text{und} \quad \overline{y} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i$$

Nun bearbeiten wir unser Extremwertproblem, welches wir ganz am Anfang formuliert haben neu. Wir hatten den Exponenten der Gaußverteilungskurve nach b partiell abgeleitet und Null gesetzt. Das machen wir nun wieder genauso, nur ist jetzt σ eine Konstante und unsere Geradengleichung lautet nicht mehr $y_i = b \cdot x_i + c$ sondern $\overline{y} = b \cdot \overline{x} + c$. Die Ableitung von damals sei noch einmal in Erinnerung gerufen:

$$\frac{\partial\left(-\sum_{i=1}^{n} \frac{(y_i - (b \cdot x_i + c))^2}{2\sigma_i^2}\right)}{\partial b} = \sum_{i=1}^{n} \frac{(y_i - b \cdot x_i - c) \cdot x_i}{\sigma_i^2} = 0$$

hat jetzt die Form:

$$\frac{\partial\left(-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{N} (\overline{y}_i - (b \cdot \overline{x}_i + c))^2\right)}{\partial b} = \frac{1}{\sigma^2} \sum_{i=1}^{N} (\overline{y}_i - b \cdot \overline{x}_i - c) \cdot \overline{x}_i = 0 \qquad \text{bzw.}$$

$$\sum_{i=1}^{N} \overline{x}_i \overline{y}_i - b \cdot \sum_{i=1}^{N} \overline{x}_i^2 - c \cdot \sum_{i=1}^{N} \overline{x}_i = 0$$

Mit c aus der Gleichung $\overline{y} = b \cdot \overline{x} + c$

erhält man $\quad \sum_{i=1}^{N} \overline{x}_i \overline{y}_i - b \cdot \sum_{i=1}^{N} \overline{x}_i^2 - (\overline{y} - b \cdot \overline{x}) \cdot \sum_{i=1}^{N} \overline{x}_i = 0$

$$\sum_{i=1}^{N} \overline{x}_i \overline{y}_i - b \cdot \sum_{i=1}^{N} \overline{x}_i^2 - \overline{y} \cdot \sum_{i=1}^{N} \overline{x}_i - b \cdot \overline{x} \cdot \sum_{i=1}^{N} \overline{x}_i) = 0$$

In diese Gleichungen setzen wir für $\overline{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i$ und $\overline{y} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i$ und durch Umstellung nach b erhält man den besten Anstieg

$$b = \frac{\sum_{i=1}^{N} \overline{x}_i \overline{y}_i - \overline{y} \cdot \sum_{i=1}^{N} \overline{x}_i}{\sum_{i=1}^{N} \overline{x}_i^2 - \overline{x} \cdot \sum_{i=1}^{N} \overline{x}_i} \quad \text{und für den Achsenabschnitt} \quad c = \overline{y} - b \cdot \overline{x}$$

Bitte stören Sie sich nicht daran, daß unter den Summen Mittelwerte stehen. Die Rechenvorschriften der obigen Gleichungen sind eindeutig und ausführbar!

N ist die Anzahl der Wertepaare und diese haben sich bereits aus Mittelwertbildungen der Einzelmessungen ergeben. (Z.B. könnte sich das Wertepaar $(\bar{x}_1, \bar{y}_1)$ aus 5 Einzelmessungen $(x_1, y_1; x_2, y_2; x_3, y_3; x_4, y_4; x_5, y_5)$ ergeben haben).

Wenn Sie für Ihre **lineare Regressionsanalyse** eine **Ursprungsgerade** benötigen, beachten Sie bitte, daß Sie dafür nicht die oben abgeleiteten Formeln übernehmen dürfen. Bei den obigen Gleichungen stellt ein Wertepaar (0,0) einen fehlerbehafteten Meßpunkt dar, während bei einer Ursprungsgeraden der Koordinatenursprung fehlerfrei ist.
Wenn eine Ursprungsgerade gesucht wird, so müssen Sie die obige Regressionsanalyse mit der Geradengleichung $y_i = b \cdot x_i$ durchführen. Das Ergebnis ist dann die Anstiegsgleichung für die Ursprungsgerade:

$$b_{Ursprungsgerade} = \frac{\sum_{i=1}^{n} \frac{x_i \cdot y_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}}$$ für den Fall, daß alle Meßwerte y_i verschiedene Fehler σ_i haben und $b_{Ursprungsgerade} = \frac{\sum_i x_i \cdot y_i}{\sum_i x_i^2}$ wenn für alle y_i die Fehler gleich sind.

Den Fehler des Anstieges können Sie berechnen im ersten Fall mit der Formel:

$$s^2_{b_{Ursprungsgerade}} = \frac{1}{\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}}$$ und bei Gleichheit aller Fehler mit $s^2_{b_{Ursprungsgerade}} = \frac{S^2}{\sum_i x_i^2}$

S^2 ist dabei wieder die Streuung der Meßwerte y_i um die beste Gerade, sie ergibt sich für die Ursprungsgerade zu $\sigma^2 = \frac{1}{N-2} \cdot \sum_{i=1}^{N} (y_i - b \cdot x_i)^2$

Zum Schluß sei hier noch die Formel zur Berechnung des **Korrelationskoeffizienten** angegeben. Der Korrelationskoeffizient r_{xy} gibt den Grad des linearen Zusammenhanges zweier Größen an. Je näher man dabei an 1 herankommt, um so stärker ist der lineare Zusammenhang. Sinnvoll wäre es, wenn man direkt am Anfang den Korrelationskoeffizient berechnet, um zu sehen inwieweit eine lineare Regressionsanalyse überhaupt sinnvoll ist.

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \; ; \quad (-1 \le r_{xy} \le +1)$$

Wobei $\sigma_{xy} = \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \right) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$

als Kovarianz bezeichnet wird und $\sigma_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$ bzw.

$\sigma_y^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2$ die uns bekannten Varianzen sind. Damit ergibt sich der Korrelationskoeffizient zu $r_{xy} = \frac{\sum_i (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot \sum_i (y_i - \bar{y})^2}}$

$\bar{y} = \frac{1}{N} \cdot \sum_i y_i$ bzw. $\bar{x} = \frac{1}{N} \cdot \sum_i x_i$

Übrigens sollten die einzelnen Meßwerte y_i mit ihren Fehlerbalken so weit von der Geraden entfernt sein, daß einzelne Balken die beste Gerade nicht mehr beinhalten, dann ist es sinnvoll, in jedem Falle mit den Gleichungen zu arbeiten, in der die Sreuung σ die Fehler beschreibt (sowohl bei der Bestimmung der Geraden als auch bei der Berechnung der Fehler des Anstieges bzw.des Achsenabschnittes).

Fehlerrechnung

z. B.

- Beiblatt „Messung physikalischer Größen und ihre Auswertung"
- Walcher, W. : Praktikum der Physik, 8. Auflage
  Kap 1.2 : S. 28 – 39
  S. 45 (Schieblehre)
- Eichler : Das neue physikalische Grundpraktikum
  Kap 1 : S. 3 – 28
- Westphal : Physikalisches Praktikum
- viele weitere Bücher ...