

Rechnet QtiPlot falsch ?

oder

Einige Anmerkungen zur Methode der linearen Regression

P. Schäfer
Institut für Physik, Humboldt-Universität zu Berlin,
Newtonstraße 15, 12489 Berlin

9. Dezember 2017

1 Das Problem

Bei der Durchsicht der von den Studenten im Physikalischen Grundpraktikum angefertigten Protokolle fällt auf, dass die dort mit der Methode der linearen Regression erhaltenen Ergebnisse zum Teil erheblich von den mit Mathematica 10.0 (2014) berechneten Kontrollwerten abweichen.

Das Problem soll an einem einfachen Beispiel aus dem Einführungskurs verdeutlicht werden. Beim Versuch F4 (siehe hierzu: Einführungspraktikum (2007) Seite 9...11) wird eine Feder schrittweise mit Massestücken von ca. 50 g belastet. Die Position x einer an der Feder befestigten Marke wird mit Hilfe einer Spiegelskala gemessen (erste Messreihe). Nach Erreichen der Höchstbelastung mit 8 Massestücken wird die Feder wieder schrittweise entlastet und dabei die Position der Marke erneut abgelesen (zweite Messreihe). Die Unsicherheit der so bestimmten Positionen x wird sowohl durch den systematischen Restfehler der verwendeten Spiegelskala als auch durch Fehler beim Ablesen von der Spiegelskala bestimmt. Der Mittelwert der Massen \bar{m} aller im Versuch verwendeten Massestücke beträgt 50.22 g bei einer Standardabweichung von 0,36 g. Auch dieser Wert beeinflusst die Streuung der gemessenen Positionen.

Aus dem Anstieg der Geraden

$$x = x_0 - \frac{\bar{m}g}{k}i \quad (1)$$

kann dann die gesuchte Federkonstante bestimmt werden. Hierbei bedeuten x_0 die unbekannte Position der verwendeten Markierung bei unbelasteter Feder, g

die Erdbeschleunigung, k die gesuchte Federkonstante und i die Anzahl der Maststücke, mit denen die Feder belastet wird. Das negative Vorzeichen resultiert aus der Tatsache, dass sich der Nullpunkt der verwendeten Spiegelskala unten befindet, und daher die Werte von x mit zunehmender Belastung abnehmen. Der Wert x_0 für $i = 0$ kann nicht abgelesen werden, da die Achsen der unbelasteten Federn mehr oder weniger stark gebogen sind. Eine typische Messdatenreihe, aufgenommen am Messplatz 2 am 7.1.2015, ist in Tabelle 1 aufgelistet.

i	Belastung	Entlastung
1	186.0	186.0
2	175.5	176.0
3	165.0	165.5
4	155.0	156.0
5	145.5	146.0
6	135.5	136.0
7	126.0	125.0
8	116.0	115.5

Tabelle 1: Beispiel einer Messdatenreihe des Versuchs F4

Betrachtet man diese Daten als 16 Wertepaare i, x , dann können sie mit der Methode der linearen Regression, so wie sie zum Beispiel in W. Schenk u.a. (2012) (Seiten 9 und 17...18), V. Nollau (1975) Seiten 129...151, L. Fahrmeier u.a. (2009) Seiten 59...63 und 90...95 oder im Einführungsskript (2007) Seite 42 Gleichungen 49, 50 und Seite 42 Gleichung für s_y (nach Gleichung 51) beschrieben ist, ausgewertet werden. In diesem Fall erhält man unabhängig von der verwendeten Software Mathematica (2014) oder QtiPlot 0.9.8.9-9 (2014) identische Resultate (siehe Tabelle 2).

Da mit dem systematischen Restfehler der verwendeten Spiegelskala $u_x = 200\mu m + 5 \cdot 10^{-4} x$ (Einführungspraktikum (2007) Seite 11) eine Abschätzung des Fehlers der Position x vorliegt und dieser für die einzelnen Messpunkte unterschiedlich ist, erscheint die Anwendung der gewichteten linearen Regression sinnvoll. Diese Methode wird z.B. von Nollau (1975) (Seiten 152...156) oder Fahrmeier (2009) (Seiten 124...127) ausführlich beschrieben. Bei der instrumentellen Gewichtung werden die Größen $c \frac{1}{s_i^2}$ als Gewichte verwendet. s_i^2 ist dabei die Varianz, das Fehlerquadrat, der jeweiligen Position x_i und c eine frei wählbare Konstante, die oftmals gleich 1 gesetzt wird.

Die Ergebnisse der obigen Messreihe mit der Methode der instrumentell gewichteten linearen Regression mit den beiden zuvor schon genutzten Auswerteprogrammen sind ebenfalls in der Tabelle 2 aufgelistet. Zusätzlich erfolgte eine

Auswertung der Daten mit den Formeln aus Nollau (1975) und Fahrmeir (2009) sowie mit den im Einführungsskript (2007) auf Seite 41,42 angegebenen Gleichungen 43 und 45. Letztere lässt sich aus der von U. Wolff (2014) auf den Seite 117/118 hergeleiteten Gleichung 9.20 für $\alpha = 2$ ableiten.

lineare Regression				
Software	x_0		Anstieg A	
	Wert	Fehler	Wert	Fehler
Mathematica	195.723	0.225292	-10.0149	0.0446145
QtiPlot [†]	195.723	0.225292	-10.0149	0.0446145
Einführungsskript	195.723	0.225292	-10.0149	0.0446145
gewichtete lineare Regression (instrumentelle Gewichtung)				
Software	x_0		Anstieg A	
	Wert	Fehler	Wert	Fehler
Mathematica	195.707	0.233646	-10.0114	0.044768
Nollau	195.707	0.233646	-10.0114	0.044768
QtiPlot [†]	195.707	0.156635	-10.0114	0.030012
Einführungsskript	195.707	0.156635	-10.0114	0.030012
Mathematica [‡]	195.707	0.156635	-10.0114	0.030012

[†] LinearFit mit und ohne Option „Scale Errors with sqrt(Chi2/doF)“

[‡] LinearModelFit mit Option „VarianceEstimatorFunction \rightarrow (1&)“

Tabelle 2: Ergebnisse der Auswertung der Messdaten mit verschiedenen Methoden und verschiedener Software. Die zugehörigen Ausgaben der genutzten Software sind im Anhang enthalten.

Sieht man sich die Ergebnisse genauer an, fällt sofort auf, dass die von Mathematica und QtiPlot berechneten Standardabweichungen der Parameter x_0 und A nicht übereinstimmen. Demgegenüber liefern sowohl Mathematica und die Berechnung mit den Gleichung von (z.B. Fahrmeir 2009, Nollau 1975) als auch QtiPlot und die Gleichungen aus dem Einführungsskript jeweils zueinander passende Ergebnisse. Verwendet man bei der Auswertung mit Mathematica die Option „VarianceEstimatorFunction \rightarrow (1&)“, so stimmt das Ergebnis sowohl mit den von QtiPlot ausgegebenen Werten als auch mit den mit Hilfe der Gleichungen aus dem Einführungsskript berechneten Werten überein. Damit deutet sich hier ein grundsätzliches Problem an, dass über die eingangs aufgeworfene Frage, ob die Software QtiPlot richtig rechnet, weit hinausgeht und auf die Frage hinausläuft, warum stehen die Formeln aus dem Einführungsskript im Widerspruch zu gängigen Lehrbüchern (z.B. Fahrmeir 2009, Nollau 1975).

Ein erster Hinweis darauf findet sich bei F. James (2006) auf Seite 3.

„Unfortunately, statisticians do not agree on basic principles. They can crudely be divided into two schools: Bayesian and frequentists (or classical).

The Bayesian approach is closer to everyday reasoning, where probability is interpreted as a degree of belief that something will happen, or that a parameter will have a given value.

The frequentist approach is closer to scientific reasoning, where probability means the relative frequency of something happening. This makes it more objective, since it can be determined independently of the observer, but restricts its application to repeatable phenomena. “

Liegt die Ursache für die unterschiedlichen Formeln und Ergebnisse in den Unterschieden zwischen diesen beiden Richtungen begründet? Dazu sollen die jeweiligen Aussagen näher betrachtet werden, wobei eine Beschränkung auf das allgemeine lineare Modell erfolgt. Zu den Grundlagen, zu Herleitungen und Beweisen sowie zu weiterführenden Fragen sei auf die angegebene Literatur verwiesen.

2 Das allgemeine lineare Modell

Den Ausgangspunkt bildet eine Reihe von n Messwerten,

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

die in Abhängigkeit von r nicht stochastischen Einflussgrößen erhalten wurden. Die n Vektoren $\mathbf{x}_i = (x_1, x_2, \dots, x_r)_i$ müssen dabei nicht alle unterschiedliche Werte besitzen. Die Abhängigkeit der Messwerte von den Einflussgrößen kann über

$$y_i = f(\mathbf{x}_i, \theta) + \epsilon_i \quad i = 1, \dots, n \quad (2)$$

beschrieben werden. Die Spaltenvektoren

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix} \quad \text{und} \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

enthalten die p unbekannt Parameter des Modells beziehungsweise eine unabhängige Zufallsgröße, die die Abweichungen der Messwerte y_i von den aus dem Modell berechneten Werten beschreibt. Für den Erwartungswert E und die Varianz Var dieser Größe soll gelten:

$$E[\epsilon] = 0 \quad \text{und} \quad \text{Var}[\epsilon] = \sigma^2 \quad (3)$$

wobei σ^2 im allgemeinen unbekannt ist. Weitere Annahmen über die Verteilung der ϵ_i , insbesondere die Voraussetzung einer Normalverteilung, sind an dieser Stelle nicht notwendig. ¹ Damit gilt für den Erwartungswert ² von y_i :

$$E[y_i] = f(\mathbf{x}_i, \theta) \quad i = 1, \dots, n \quad (4)$$

Wenn sich die Funktion $f(\mathbf{x}_i, \theta)$ als Linearkombination von p beliebigen, nicht notwendigerweise linearen Funktionen $f_j(\mathbf{x}_i)$ darstellen lässt, kann man dafür schreiben:

$$E[y_i] = \sum_{j=1}^p \theta_j f_j(\mathbf{x}_i) \quad i = 1, \dots, n \quad (5)$$

In diesem Fall liegt ein in den Parametern θ lineares Modell vor. Aus diesem lassen sich verschiedene Spezialfälle ableiten. Wenn zum Beispiel die Messwerte

nur von einer Einflussgröße $\mathbf{x}_i = (x_1)_i$ linear abhängen, kann $f_1(\mathbf{x}_i) = 1$ sowie $f_2(\mathbf{x}_i) = x_i$ festgesetzt werden. In diesem Fall ergibt sich daraus das Modell der linearen Regression

$$E[y_i] = \theta_1 + \theta_2 x_i \quad i = 1, \dots, n \quad (6)$$

mit den unbekanntem Parametern θ_1 und θ_2 . In ähnlicher Weise lassen sich auch das Modell für eine Gerade durch den Koordinatenursprung oder für die Anpassung mit einem Polynom herleiten.

Durch die Wahl der p Funktionen $f_j(\mathbf{x}_i)$ und der n Messstellen \mathbf{x}_i sind die Werte

$$a_{i,j} = f_j(\mathbf{x}_i) \quad i = 1, \dots, n \quad j = 1, \dots, p$$

fest vorgeben. Diese $a_{i,j}$ bilden die $n \times p$ Design-Matrix

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,p} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,p} \end{pmatrix}$$

des linearen Modells, das damit als einfache Matrixgleichung geschrieben werden kann.

$$\mathbf{y} = \mathbf{A} \boldsymbol{\theta} + \boldsymbol{\epsilon} \quad \text{beziehungsweise} \quad E[\mathbf{y}] = \mathbf{A} \boldsymbol{\theta} \quad (7)$$

Bisher wurde vorausgesetzt, dass alle Messergebnisse die gleiche Genauigkeit besitzen. Dies ist oftmals nicht der Fall und die einzelnen Messwerte sollen daher mit unterschiedlicher Gewichtung in das Modell einfließen. Zu diesem Zweck kann zu jedem Messwert ein entsprechendes Gewicht p_i festgelegt werden. Damit lassen sich die relativen Unterschiede in der Genauigkeit der einzelnen Messungen in einfacher Weise beschreiben. Diese Werte bilden die Gewichtsmatrix $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_n)$. Setzt man für

$$\begin{aligned} \mathbf{y}^* &= \mathbf{P}^{1/2} \mathbf{y} \\ \mathbf{A}^* &= \mathbf{P}^{1/2} \mathbf{A} \\ \boldsymbol{\epsilon}^* &= \mathbf{P}^{1/2} \boldsymbol{\epsilon} \end{aligned} \quad (8)$$

so ergibt

$$\mathbf{y}^* = \mathbf{A}^* \boldsymbol{\theta} + \boldsymbol{\epsilon}^* \quad (9)$$

wieder ein lineares Modell (Gleichung 7).

Liegen Informationen über die Varianz σ_i der Fehler der einzelnen Messwerte y_i vor, so können diese zur Gewichtung benutzt werden. Dazu geht man von der

Vorstellung aus, dass die einzelnen Messwerte y_i jeweils Mittelwerte einer Stichprobe des Umfangs n_i repräsentieren. Die Varianz dieser Mittelwerte folgt aus der Varianz σ der zugrunde liegenden Grundgesamtheit entsprechend $\sigma_i^2 = \sigma^2/n_i$. Setzt man als Gewichte den Umfang der jeweiligen Stichproben ein, wie dies auch von Fahrmeir (2009) für gruppierte Daten beschrieben wird, so folgt daraus:

$$p_i = n_i = \frac{\sigma^2}{\sigma_i^2}$$

Da oftmals die Varianz σ^2 der Grundgesamtheit vorab nicht bekannt ist, werden die Gewichte nur proportional zu n_i angesetzt.

$$p_i \propto \frac{\sigma^2}{\sigma_i^2} \quad p_i = c \frac{1}{\sigma_i^2} \quad (10)$$

wobei c in diesem Fall, eine willkürlich festgelegte Konstante³ ist, die, wie später zu sehen sein wird, keinen Einfluss auf die Schätzungen von $\widehat{\theta}$ (Gleichung 12) und $\widehat{\text{Cov}}[\widehat{\theta}]$ (Gleichung 14) hat.

Die Diagonalmatrix $\text{diag}(\sigma_1^2, \sigma_2^2 \dots \sigma_n^2)$ entspricht der $n \times n$ Kovarianzmatrix $\text{Cov}[\epsilon]$ der bisher als statistisch unabhängig angenommenen Abweichungen ϵ_i der Messwerte y_i von den aus dem Modell berechneten Werten. Diese Annahme ist im weiteren nicht mehr erforderlich. Die Kovarianzmatrix der Messabweichungen Σ ist allgemein gegeben durch:

$$\Sigma = \text{Cov}[\epsilon] = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \cdots & \sigma_{n,n} \end{pmatrix}$$

wobei

$$\sigma_{i,i} = \sigma_i^2 = \text{Var}[\epsilon_i] \quad \text{und} \quad \sigma_{i,j} = \text{Cov}[\epsilon_i, \epsilon_j] \quad \text{mit} \quad i \neq j$$

sind. Die Kovarianzmatrix Σ ist symmetrisch und positiv semidefinit. Ist sie positiv definit, kann ihre Inverse, auch als Präzisionsmatrix bezeichnet, zur Festlegung der Gewichtsmatrix genutzt werden.

$$\mathbf{P} = \sigma^2 \Sigma^{-1}$$

σ^2 ist hierbei der Varianzfaktor, oftmals auch als Varianz der Gewichtseinheit bezeichnet, der festlegt welchem Messwert das Gewicht Eins gegeben wird. Bei unbekanntem σ wird die Gewichtsmatrix als proportional zur Präzisionsmatrix angesetzt.

$$\mathbf{P} \propto \Sigma^{-1} \quad \mathbf{P} = c \Sigma^{-1} \quad (11)$$

Für die Konstante c gilt das oben gesagte.

Setzt man für die Gewichtsmatrix die Einheitsmatrix ein, $\mathbf{P} = \mathbf{I}$, so ergibt sich daraus das einfache lineare Modell (Gleichung 7). Eine getrennte Behandlung dieses Spezialfalles ist im weiteren nicht erforderlich.

2.1 Klassische Statistik

Gesucht sind die Schätzwerte für $\hat{\theta}$, $\hat{\sigma}^2$ und $\widehat{\text{Cov}}[\hat{\theta}]$, die den wahren Werten der unbekannt Parameter möglichst nahe kommen. Bei der Methode der kleinsten Quadrate bestimmt das Minimum der Summe der quadrierten Abweichung

$$\begin{aligned} Q(\theta) &= \sum_{i=1}^n \sum_{j=1}^n (y_i - \mathbb{E}[y_i]) p_{i,j} (y_j - \mathbb{E}[y_j]) \\ &= \boldsymbol{\epsilon}^T \mathbf{P} \boldsymbol{\epsilon} \\ &= (\mathbf{y} - \mathbf{A} \theta)^T \mathbf{P} (\mathbf{y} - \mathbf{A} \theta) \\ &= \mathbf{y}^T \mathbf{P} \mathbf{y} - 2\theta^T \mathbf{A}^T \mathbf{P} \mathbf{y} + \theta^T \mathbf{A}^T \mathbf{P} \mathbf{A} \theta \end{aligned}$$

die Schätzwerte der Parameter. Aus

$$\frac{\partial Q(\theta)}{\partial \theta} = -2\mathbf{A}^T \mathbf{P} \mathbf{y} + 2\mathbf{A}^T \mathbf{P} \mathbf{A} \theta = \mathbf{0}$$

folgt unmittelbar:

$$\hat{\theta} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} \quad (12)$$

Der Schätzwert für die Varianz der Messabweichungen kann aus dem Minimum von $Q(\theta)$ bestimmt werden.

$$\hat{\sigma}^2 = \frac{1}{n-p} Q(\theta)_{min} = \frac{1}{n-p} (\mathbf{y} - \mathbf{A} \hat{\theta})^T \mathbf{P} (\mathbf{y} - \mathbf{A} \hat{\theta}) \quad (13)$$

Für die Kovarianzmatrix von $\hat{\theta}$ ergibt sich damit: ⁴

$$\widehat{\text{Cov}}[\hat{\theta}] = \hat{\sigma}^2 (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \quad (14)$$

Bisher wurden keine Annahmen über die Verteilung der Messdaten zu Grunde gelegt. Die optimalen Eigenschaften der Methode der kleinsten Quadrate folgen einzig aus der Linearität des Modells. Wird für die Messabweichungen ϵ eine Normalverteilung angenommen, führt die Maximum-Likelihood-Schätzung von $\hat{\theta}$ beziehungsweise die Restringierte Maximum-Likelihood-Schätzung ⁵ von $\hat{\sigma}^2$ zu dem gleichen Ergebnis. Unter dieser Voraussetzung sind auch die geschätzten Werte $\hat{\theta} \sim N(\theta, \sigma^2 (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1})$ normalverteilt.

Das Konfidenzintervall (Fehlerintervall) eines geschätzten Parameters $\hat{\theta}_j$ wird durch die Wahrscheinlichkeit P bestimmt, mit der dieser im Intervall $d_{min} < \theta_j < d_{max}$ liegt

$$P(d_{min} < \theta_j < d_{max}) = 1 - \alpha$$

wobei $1 - \alpha$ das Signifikanzniveau ⁶ angibt. Zur Festlegung der Intervallgrenzen d wird der Test der Hypothese $H_0: \theta_j = d$ verwendet. Dieser basiert auf der Teststatistik

$$t = \frac{\hat{\theta}_j - d}{s_j}$$

mit

$$s_j = \widehat{\text{Var}}[\hat{\theta}_j]^{1/2} = \left(\widehat{\text{Cov}}[\hat{\theta}]_{j,j} \right)^{1/2} = \left(\hat{\sigma}^2 (\mathbf{A}^T \mathbf{P} \mathbf{A})_{j,j}^{-1} \right)^{1/2}$$

Die Hypothese H_0 wird abgelehnt, wenn sich

$$|t| > t_{n-p}(1 - \alpha/2)$$

ergibt. Hierbei bedeutet $t_{n-p}(1 - \alpha/2)$ das $(1 - \alpha/2)$ -Quantil der Standard t-Verteilung ⁷ mit $n - p$ Freiheitsgraden. Die Grenzen des Konfidenzintervalls zum Signifikanzniveau $(1 - \alpha)$ werden somit bestimmt durch

$$d_{min,max} = \hat{\theta}_j \mp t_{n-p}(1 - \alpha/2) s_j \quad (15)$$

Ähnliche Gleichungen kann man in vielen Lehrbüchern zur klassischen Statistik, z.B. in Nollau (1975) (152 ff), James (2006) (183 ff), Fahrmeir (2009) (124 ff), B. R. Martin (2012) (145 ff) oder J. Wakefield (2013) (214 ff) finden, wobei teilweise anstelle der Gewichtsmatrix \mathbf{P} direkt die inverse Kovarianzmatrix Σ^{-1} benutzt oder das erste Element des Parametervektors θ mit θ_0 bezeichnet wird.

2.2 Bayes Statistik

Die Bayes Statistik verfolgt einen grundsätzlich anderen Ansatz als die Klassische Statistik. Die Klassische Statistik versucht nur aus den gegebenen Daten, den Messwerten, z.B. mit der Methode der kleinsten Quadrate die wahrscheinlichsten Werte der unbekannt Parameter zu schätzen, deren Erwartungswerte als fest angenommenen werden. Demgegenüber geht die Bayes-Statistik davon aus, das auch die Parameter selbst Zufallsgrößen sein können und bestimmt nicht nur die Parameterwerte, für die die Datenwahrscheinlichkeit am größten ist, sondern auch deren Wahrscheinlichkeitsverteilung, aus der sich die Unsicherheit über den wahren Wert der Parameter ableiten lässt. Dazu arbeitet die Bayes-Statistik mit bedingten Wahrscheinlichkeiten. Mit Hilfe des Bayes-Theorem

$$p(\theta | \mathbf{y}) = \frac{p(\mathbf{y} | \theta) p(\theta)}{p(\mathbf{y})} \quad (16)$$

wird die gesuchte bedingte Wahrscheinlichkeitsverteilung $p(\theta | \mathbf{y})$ der Parameter, deren Posteriori-Wahrscheinlichkeit bestimmt. ⁸ Neben der Likelihood-Funktion $p(\mathbf{y} | \theta)$ sind Vorkenntnisse über die Wahrscheinlichkeitsverteilung der unbekannt Parameter $p(\theta)$, über die sogenannte Priori-Verteilung notwendig. Je nach Festlegung beinhaltet diese eine mehr oder weniger große Subjektivität. Auch wenn die statistischen Aussagen ausschließlich aus der Posteriori-Verteilung abgeleitet werden, bleibt diese subjektive Komponente erhalten. Im Gegensatz dazu arbeitet die klassische Statistik nur mit der Likelihood-Funktion. Da der Ausdruck $p(\mathbf{y})$ nur eine Normierungskonstante ist, reicht es aus, Gleichung 16 in der Form

$$p(\theta | \mathbf{y}) \propto p(\mathbf{y} | \theta) \times p(\theta) \quad (17)$$

zu verwenden.

In der Schreibweise der Bayes-Statistik lautet das allgemeine lineare Modell:

$$\begin{aligned} \mathbf{y} &= \mathbf{A} \theta + \epsilon \quad \text{beziehungsweise} \quad \mathbb{E}[\mathbf{y} | \theta] = \mathbf{A} \theta \\ \text{mit} \quad \mathbb{E}[\epsilon | \theta] &= 0 \\ \text{und} \quad \text{Cov}[\epsilon | \theta, \sigma^2] &= \text{Cov}[\mathbf{y} | \sigma^2] = \mathbf{\Sigma} = \sigma^2 \mathbf{P}^{-1} \end{aligned} \quad (18)$$

wobei im weiteren für die Verteilung der Messabweichungen eine mehrdimensionale Normalverteilung vorausgesetzt wird. ⁹ Unter diesen Voraussetzungen ergibt sich die Likelihood-Funktion zu:

$$p(\mathbf{y} | \theta, \sigma) = \frac{1}{(2\pi)^{n/2} |(\sigma^2 \mathbf{P}^{-1})|^{1/2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{A} \theta)^T \mathbf{P} (\mathbf{y} - \mathbf{A} \theta)\right)$$

Da über die Parameter θ und σ^2 keine Vorinformationen vorliegen, wird im weiteren von einem nichtinformativen Priori-Verteilung ausgegangen. Für θ wird angenommen, dass alle möglichen Werte gleich wahrscheinlich sind.

$$p(\theta) \propto 1 \quad \text{für} \quad -\infty < \theta_i < \infty, \quad i = 1, \dots, n$$

Da σ^2 nur Werte im Intervall $0 < \sigma^2 < \infty$ annehmen kann, wird von einer Gleichverteilung für $\ln \sigma$ ausgegangen. Dies führt auf die nichtinformativ Priori-Verteilung für die Varianz: ¹⁴

$$p(\sigma) \propto \sigma^{-2} \quad \text{für} \quad 0 < \sigma < \infty$$

Somit ergibt sich für die Priori-Verteilung $p(\theta, \sigma^2)$:

$$p(\theta, \sigma^2) = p(\theta) \times p(\sigma^2) \propto \sigma^{-2} \quad (19)$$

Diese Verteilungsfunktion ist nicht normierbar und wird deswegen auch als uneigentliche Priori-Verteilung bezeichnet. Da die verwendete Likelihood-Funktion normierbar ist, folgt aus der Verknüpfung nach dem Bayes-Theorem eine Posteriori-Verteilung, die normierbar ist.

$$p(\theta, \sigma \mid \mathbf{y}) \propto (\sigma^2)^{-(n+2)/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{A}\theta)^T \mathbf{P}(\mathbf{y} - \mathbf{A}\theta)\right)$$

Der Exponenten lässt sich geeignet umformen.¹⁰

$$(\mathbf{y} - \mathbf{A}\theta)^T \mathbf{P}(\mathbf{y} - \mathbf{A}\theta) = (n-p)s^2 + (\theta - \mu_0)^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\theta - \mu_0)$$

Darin bedeuten

$$\mu_0 = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y}$$

und

$$s^2 = \frac{1}{n-p} (\mathbf{y} - \mathbf{A}\mu_0)^T \mathbf{P}(\mathbf{y} - \mathbf{A}\mu_0)$$

Damit ergibt sich eine Normal-inverse Gammaverteilung¹¹ $\text{NIG}(\mathbf{m}, \mathbf{M}, a, b)$ mit den Parametern

$$\begin{aligned} \mathbf{m} &\rightarrow \mu_0, & \mathbf{M} &\rightarrow (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \\ a &\rightarrow \frac{n-p}{2}, & b &\rightarrow \frac{n-p}{2} s^2 \end{aligned}$$

und der Dichte

$$p(\theta, \sigma \mid \mathbf{y}) \propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{1}{2\sigma^2} [s^2(n-p) + (\theta - \mu_0)^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\theta - \mu_0)]\right\} \quad (20)$$

Um daraus die Verteilung $p(\theta \mid \mathbf{y})$ zu erhalten, wird über σ^2 integriert.

$$p(\theta \mid \mathbf{y}) \propto \int (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{1}{2\sigma^2} [s^2(n-p) + (\theta - \mu_0)^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\theta - \mu_0)]\right\} d\sigma^2$$

Der Integrand ist der Kern einer inversen Gamma-Verteilung¹² $\text{InvGa}(a, b)$ für σ^2 mit den Parametern

$$\begin{aligned} a &\rightarrow n/2 \\ b &\rightarrow \frac{s^2(n-p) + (\theta - \mu_0)^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\theta - \mu_0)}{2} \end{aligned}$$

Aus dem Normierungsparameter der inversen Gamma-Verteilung folgt:

$$\begin{aligned} p(\theta | \mathbf{y}) &\propto \Gamma\left(\frac{n}{2}\right) \left[\frac{s^2(n-p) + (\theta - \mu_0)^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\theta - \mu_0)}{2} \right]^{-\frac{n}{2}} \\ &\propto \left[1 + \frac{(\theta - \mu_0)^T (s^2(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1})^{-1} (\theta - \mu_0)}{n-p} \right]^{\frac{n-p+p}{2}} \end{aligned}$$

Dieser Ausdruck entspricht dem Kern einer p -dimensionalen t-Verteilung¹³ $T_p(\mu, \mathbf{D}, d)$ mit :

$$\begin{aligned} \mu &\rightarrow \mu_0 \\ \mathbf{D} &\rightarrow s^2(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \\ d &\rightarrow n-p \end{aligned}$$

Für den Parametervektor θ ergibt sich damit die p -dimensionalen t-Verteilung

$$\theta | \mathbf{y} \sim T_p(\mu_0, s^2(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1}, n-p) \quad (21)$$

als Posteriori-Randverteilung. Zu dem gleichen Ergebnis gelangt man auch mit anderen nichtinformativen Priori-Verteilungen^{14, 15}. Aus dem Erwartungswert der mehrdimensionalen t-Verteilung

$$E[\theta | \mathbf{y}] = \mu_0$$

ergibt sich die Bayesschätzung $\hat{\theta}_B$ zu

$$\hat{\theta}_B = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} \quad (22)$$

Diese stimmt mit dem Ergebnis der Klassischen Statistik (12) überein.

Aus der Posteriori-Randverteilung (21) lässt sich der Konfidenzbereich für jeden Teilvektor des Parametervektor θ und damit auch das Konfidenzintervall des einzelnen Parameters θ_j ableiten. Es ergibt sich dann eine eindimensionale t-Verteilung mit $n-p$ Freiheitsgraden, die sich mit der Substitution

$$x = \frac{\mu_{0j} - \theta_j}{s_j}$$

und dem entsprechenden Hauptdiagonalelement der Dispersionsmatrix \mathbf{D}

$$s_j = (s^2(\mathbf{A}^T \mathbf{P} \mathbf{A})_{j,j}^{-1})^{1/2}$$

auf eine Standard t-Verteilung mit $n-p$ Freiheitsgraden zurückführen lässt. Die sich daraus ergebenden Konfidenzintervalle

$$\mu_{0j} - t_{n-p}(1-\alpha/2) s_j \leq \theta_j \leq \mu_{0j} + t_{n-p}(1-\alpha/2) s_j \quad (23)$$

stimmen mit denen der Klassischen Statistik (15) überein.

Aus der Posteriori-Verteilung (20) folgt als Randverteilung für die Varianz σ^2 eine inverse Gammaverteilung¹² $\text{InvGa}(a, b)$ mit den Parametern

$$\begin{aligned} a &\rightarrow \frac{n-p}{2} \\ b &\rightarrow \frac{n-p}{2} s^2 \end{aligned}$$

Aus dem Erwartungswert der inversen Gammaverteilung ergibt sich damit die Bayesschätzung des Varianzfaktors $\hat{\sigma}_B^2$ zu

$$\begin{aligned} \hat{\sigma}_B^2 &= \text{E}[\sigma^2 \mid \mathbf{y}] \\ &= \frac{n-p}{n-p-2} s^2 \\ &= \frac{1}{n-p-2} (\mathbf{y} - \mathbf{A} \mu_0)^T \mathbf{P} (\mathbf{y} - \mathbf{A} \mu_0) \end{aligned}$$

Das gleiche Ergebnis kann auch auf anderen Weise gefunden werden¹⁶. Die Bayesschätzung $\hat{\sigma}_B^2$ ist um den Faktor $(n-p)/(n-p-2)$ größer als die Schätzung $\hat{\sigma}^2$ der klassischen Statistik(13).

2.3 Fazit

Die Klassische Statistik liefert im Rahmen des Allgemeinen Linearen Modells für den Parametervektor und für die Fehlerintervalle das gleiche Ergebnis wie die Bayes Statistik bei Verwendung einer nicht-informativen Priori-Verteilung. Für große Werte von n gilt das auch für den Varianzfaktor σ^2 . Der Varianzfaktor, der sich aus der Klassischen Statistik ergibt, stellt die untere Grenze für den Varianzfaktor dar, den man über den entsprechenden Ansatz der Bayes-Statistik erhält.

Ein Vergleich mit Tabelle 2 zeigt, dass die mit QtiPlot und den Formeln im Skript Einführungsskript berechneten Werte für die Fehlerintervalle der Parameter signifikant kleiner sind als die nach der Klassischen Statistik berechneten Größen. Die grundlegend unterschiedliche Vorgehensweise der beiden Richtungen in der Statistik kann nicht die Ursache für die aufgetretenen Differenzen sein.

3 Fehlerfortpflanzung

Im Einführungsskript findet man auf Seite 42 die Aussage: „ Die Standardabweichungen für die beiden Parameter ... erhält man durch Anwendung des Fortpflanzungsgesetzes für die Unsicherheiten“. Als Quelle kann hier das im Vorwort erwähnte Buch P.R. Bevington, D.K. Robinson (2003) vermutet werden. Dort findet sich auf Seite 109 die Herleitung der im Einführungsskript verwendeten Formeln. Eine ähnliche Aussage wie im Einführungsskript ist auch bei Wolff (2014) auf Seite 118 zu finden, ohne dass dazu eine Begründung oder eine Quellenangabe erfolgt.

Was bedeutet die Anwendung der Fehlerfortpflanzung auf das Allgemeine Lineare Modell? Dazu wird zuerst eine skalare Größe θ betrachtet, die eine beliebige Funktion des n dimensionalen Vektors \mathbf{y} ist.

$$\theta \equiv \theta(\mathbf{y}) = \theta(y_1, y_2, \dots, y_n)$$

Wenn die Fehler $\epsilon_i = y_i - \bar{y}_i$ mit $\bar{y}_i = E[y_i]$ klein sind, kann die Taylorentwicklung von $\theta(\mathbf{y})$ in der Umgebung des Punktes $\mathbf{y} = \bar{\mathbf{y}}$ nach der ersten Ordnung abgebrochen werden.

$$\theta(\mathbf{y}) = \theta(\bar{\mathbf{y}}) + \sum_{i=1}^n (y_i - \bar{y}_i) \left. \frac{\partial \theta(\mathbf{y})}{\partial y_i} \right|_{\mathbf{y}=\bar{\mathbf{y}}} \quad (24)$$

Für die Varianz von $\theta(\mathbf{y})$ gilt:

$$\text{Var}[\theta(\mathbf{y})] = E \left[\left(\theta(\mathbf{y}) - E[\theta(\mathbf{y})] \right)^2 \right] \simeq E \left[\left(\theta(\mathbf{y}) - \theta(\bar{\mathbf{y}}) \right)^2 \right]$$

Wird die Taylorentwicklung 24 in diese Gleichung eingesetzt, ergibt sich:

$$\begin{aligned} \text{Var}[\theta(\mathbf{y})] &= E \left[\left(\sum_{i=1}^n (y_i - \bar{y}_i) \left. \frac{\partial \theta(\mathbf{y})}{\partial y_i} \right|_{\mathbf{y}=\bar{\mathbf{y}}} \right)^2 \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \left. \frac{\partial \theta(\mathbf{y})}{\partial y_i} \right|_{\mathbf{y}=\bar{\mathbf{y}}} E \left[(y_i - \bar{y}_i)(y_j - \bar{y}_j) \right] \left. \frac{\partial \theta(\mathbf{y})}{\partial y_j} \right|_{\mathbf{y}=\bar{\mathbf{y}}} \end{aligned}$$

Dabei sind die

$$E \left[(y_i - \bar{y}_i)(y_j - \bar{y}_j) \right] = \sigma_{i,j} = \text{Cov}[y_i, y_j]$$

die Elemente der Kovarianzmatrix Σ des Vektors \mathbf{y} . Wird außerdem

$$(\Delta \theta)^2 = \text{Var}[\theta(\mathbf{y})]$$

gesetzt, erhalten wir das allgemeine Fehlerfortpflanzungsgesetz

$$(\Delta \theta)^2 = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial \theta(\mathbf{y})}{\partial y_i} \Big|_{\mathbf{y}=\bar{\mathbf{y}}} \sigma_{i,j} \frac{\partial \theta(\mathbf{y})}{\partial y_j} \Big|_{\mathbf{y}=\bar{\mathbf{y}}} \quad (25)$$

Für den Fall, dass die Werte des Vektors \mathbf{y} nicht korreliert sind, wenn für $i \neq j$ gilt $\text{Cov}[y_i, y_j] = 0$, vereinfacht sich die Gleichung 25 zu der häufig verwendeten Form:

$$(\Delta \theta)^2 = \sum_{i=1}^n \left(\frac{\partial \theta(\mathbf{y})}{\partial y_i} \Big|_{\mathbf{y}=\bar{\mathbf{y}}} \Delta y_i \right)^2$$

mit $(\Delta y_i)^2 = \sigma_{i,i} = \text{Var}[y_i]$.

Das allgemeine Fehlerfortpflanzungsgesetz 25 für eine Größe θ lässt sich auf den p -dimensionalen Vektor θ erweitern. Dessen Elemente θ_k sind alle Funktionen des selben Vektors \mathbf{y} und daher korreliert. Für die Elemente der Kovarianzmatrix $\text{Cov}[\theta]$ gilt dann:

$$\text{Cov}[\theta_k, \theta_l] = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial \theta_k}{\partial y_i} \Big|_{\mathbf{y}=\bar{\mathbf{y}}} \sigma_{i,j} \frac{\partial \theta_l}{\partial y_j} \Big|_{\mathbf{y}=\bar{\mathbf{y}}} \quad (26)$$

Wird eine $p \times n$ Matrix \mathbf{G} mit den Elementen

$$g_{k,i} = \frac{\partial \theta_k}{\partial y_i} \Big|_{\mathbf{y}=\bar{\mathbf{y}}} \quad \text{mit } k = 1 \dots p \quad \text{und } i = 1 \dots n$$

eingeführt, kann Gleichung 26 als Matrixgleichung geschrieben werden.

$$\text{Cov}[\theta] = \mathbf{G} \Sigma \mathbf{G}^T \quad (27)$$

Bezeichnet man die $p \times n$ Matrix $(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}$ aus Gleichung 12 mit \mathbf{H} , dann kann die Gleichung 12 auch geschrieben werden als:

$$\begin{aligned} \hat{\theta} &= \mathbf{H} \mathbf{y} \\ \hat{\theta}_k &= \sum_{i=1}^n h_{k,i} y_i \end{aligned}$$

Aus dieser Darstellung wird sofort klar, dass für die einzelne Elemente der Matrix \mathbf{H} gilt:

$$h_{k,i} = \frac{\partial \theta_k}{\partial y_i}$$

Damit kann aus den Gleichungen 27 und 11 die Beziehung

$$\begin{aligned} \text{Cov}[\theta] &= \mathbf{H} \Sigma \mathbf{H}^T \\ &= c(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \end{aligned} \quad (28)$$

abgeleitet ¹⁷ werden. Die so bestimmte Kovarianzmatrix enthält die, bei der Aufstellung des Allgemeinen Linearen Modells willkürlich festlegbare Konstante c . Alle aus der Kovarianzmatrix abgeleiteten Fehlerintervalle sind daher von der Festlegung dieser Konstanten abhängig. Nur wenn die Kovarianzmatrix Σ vollständig bekannt ist, oder aus den vorliegenden Daten bestimmt werden kann, so wie es von M. Grabe (2011) im Kapitel 21 beschrieben wird, führt die Festlegung $c = 1$ beziehungsweise $\mathbf{P} = \Sigma^{-1}$ zu richtigen Ergebnissen. Nur in diesem Fall gilt:

$$\text{Cov}[\theta] = (\mathbf{A}^T \Sigma^{-1} \mathbf{A})^{-1} \quad (29)$$

Auf diesen Umstand macht auch Bevington (2003) auf Seite 108 aufmerksam. Er schlägt vor, die bestimmten Varianzen, die Quadrate der Fehlerintervalle, zu korrigieren, sie mit dem Wert von $\chi^2/(n - p)$ zu multiplizieren, wenn $\chi^2/(n - p)$ zu stark von 1 abweicht. Genau dann ergibt sich die von der klassischen Statistik (Gleichung 14) berechnete Kovarianzmatrix. Auf die zwingende Notwendigkeit der Berücksichtigung des Varianzfaktors $\hat{\sigma}^2$ weist auch Martin (2012) ausdrücklich im Abschnitt 8.1.2 hin. Wird dies unterlassen, dann führt die Anwendung des Fehlerfortpflanzungsgesetzes auf das Allgemeine Lineare Modell zu falschen Ergebnissen.

Damit ist die Ursache der Widersprüche in Tabelle 2 geklärt. Die Formeln aus dem Einführungsskript (2007) liefern das gleiche falsche Ergebnis wie die Mathematica Routine „LinearModelFit“ mit der Option „VarianceEstimatorFunction \rightarrow (1&)“. In beiden Fällen wird der Varianzfaktor auf den Wert 1 festgelegt. Ob das auch bei QtiPlot der Fall ist, wird im weiteren noch untersucht.

Da in der im Einführungsskript (2007) auf Seite 42 angegebenen Gleichung 45 der Varianzfaktor nicht berücksichtigt wird und dort auch die obigen Hinweise in keiner Weise erwähnt werden, ist diese Gleichung und die ihr zu Grunde liegende Vorgehensweise als falsch zu betrachten.

4 Rechnet QtiPlot falsch ?

Um diese Frage zu beantworten, ist es notwendig den Quellcode des Programms genauer zu analysieren, da die zugehörige Dokumentation keinerlei Informationen zu den genutzten Algorithmen, zu den verwendeten statistischen Verfahren und den dahinterliegenden statistischen Theorien enthält. QtiPlot (2014) nutzt die Routinen der GNU Scientific Library. Diese sind ausführlich in GNU Scientific Library, Reference Manual (2011) beschrieben. Dort bezieht man sich auf R.M. Barnett u.a. (1996), der älteren Version von J. Beringer u.a. (2012).

Wie die Quellcodeanalyse zeigt, werden für die Fitverfahren, die auf dem Allgemeinen Linearen Modell basieren, mehrere verschiedene Bibliotheksfunktionen aufgerufen. Für den Fall ungewichteter Daten verwendet QtiPlot die Bibliotheksfunktionen `int gsl_fit_linear (...)` für den linearen Fit beziehungsweise die Funktion `int gsl_fit_mul (...)` für den Fit einer Geraden durch den Nullpunkt. Beide Routinen geben nach GSL Reference Manual (2011) (Seite 404) die Elemente der „Varianz-Kovarianz-Matrix“ der Parameter zurück. Diese Matrix entspricht der Gleichung 14. Für den Fall gewichteter Daten kommen die Bibliotheksfunktionen `int gsl_fit_wlinear (...)` beziehungsweise die Funktion `int gsl_fit_wmul (...)` (GSL Reference Manual (2011) Seite 404 folgende) zur Anwendung. Diese berechnen die Kovarianz-Matrix der Parameter. Diese entspricht nach Barnett (1996) (Seite 161, Gleichung. 28.28) beziehungsweise Behringer (2012) (Seite 392 Gleichung 36.17) der Gleichung 29. Um daraus die „Varianz-Kovarianz-Matrix“ zu erhalten, ist die Multiplikation mit $\chi^2/(n-p) = \hat{\sigma}^2$ notwendig. Der Wert von χ^2 wird von allen vier Bibliotheksfunktionen berechnet. Dass zwischen gewichteten und ungewichteten Daten unterschieden wird, obwohl dies vom zugrundeliegenden Allgemeinen Linearen Modell nicht notwendig wäre, hat rein numerische Gründe. Es erspart an mehreren Stellen die Ausführung der Multiplikation mit der Gewichtsmatrix, die in diesem Falle die Einheitsmatrix ist. Eine Folge der unterschiedlichen Bedeutung der von den verwendeten Funktionen zurückgegebenen Werte ist, dass die Fehler im Fall gewichteter Daten noch mit `sqrt(chiquadrat/doF)` multipliziert werden müssen, während dies im Fall von ungewichteten Daten bereits erfolgt ist. Die Quelltextanalyse ergab, dass die Option „Gewichtung mit `sqrt(chiquadrat/doF)`“ bei allen auf dem linearen Modell basierenden Fitverfahren wirkungslos ist. Die Berücksichtigung des Varianzfaktors $\hat{\sigma}^2$ bei der Berechnung der Fehlerintervalle muss im Nachgang von Hand durch den Anwender von QtiPlot erfolgen.

Die Frage, ob QtiPlot numerisch falsch rechnet, kann man nun eindeutig mit nein beantworten. Dies gilt zumindest für alle Fitroutinen, die auf dem Allgemeinen Linearen Modell basieren und die obengenannten Funktionen der GNU Scientific Library verwenden. Die richtige Interpretation der berechneten Werte durch die Anwender des Programms ist jedoch auf Grund der mangelhaften Dokumenta-

tion der Bedeutung der ausgegebenen Werte und der verwendeten statistischen Verfahren nicht möglich.

5 Schlussfolgerungen

Wie sich gezeigt hat, widersprechen die im Einführungsskript angegebenen Formeln, den Aussagen der beiden dominierenden Richtungen der Statistik. Sie stützen sich auf eine nur unter ganz strengen Rahmenbedingungen anwendbare Vorgehensweise. Die dafür notwendigen Voraussetzungen bei der Datenerhebung werden jedoch nicht behandelt. An diesem Punkt ist das Einführungsskript als fehlerhaft anzusehen.

QtiPlot rechnet numerisch richtig. Durch die ungenügende Dokumentation ist der Anwender des Programms nicht in der Lage die gelieferten numerischen Werte im Sinne der zu Grunde liegenden statistischen Methoden, unabhängig davon ob diese auf der Klassischen Statistik oder Bayes-Statistik beruhen, richtig zu interpretieren. Vor diesem Hintergrund ist nicht nur von der Verwendung des noch weniger dokumentierten Fit-Wizard von QtiPlot abzuraten, sondern generell von der Anwendung von QtiPlot und anderer aus QtiPlot hervorgegangenen Programmen wie SciDAVis. Für die Belange der Ausbildung von Studenten im Rahmen des Praktikums sind gut dokumentierte, frei zugängliche Werkzeuge wie GnuPlot wesentlich sinnvoller.

Anmerkungen

¹ Zu den notwendigen Voraussetzungen für ein lineares Modell gibt es unterschiedliche Auffassungen:

Martin (2012) Seite 145: Insbesondere ist die Annahme nicht notwendig, dass die Fehler der Messwerte normalverteilt sind.

James (2006) Seite 186: Es werden keine Annahmen über die Verteilung der Daten \mathbf{y} gemacht.

Nollau (1975) setzt schon bei der Herleitung des linearen Modells für die ϵ_i unabhängige normalverteilte Zufallsgrößen voraus.

² Im Unterschied zu Martin (2012) (Seite 145), Fahrmeir (2009) (Seite 62) und Nollau (1975) (Seite 134) verwendet James (2006) in diesem Zusammenhang im Kapitel 8.4 (Seite 183ff) anstelle von $E[y_i]$ den Ausdruck $E[y_i, \theta]$.

³ Bevington (2003) wählt im Kapitel 11 für die Gewichte (Gleichung (11.2) Seite 194, Gleichung (11.32) Seite 203 sowie auf Seite 215) den Ausdruck:

$$p_i = \frac{1/\sigma_i^2}{(1/n) \sum (1/\sigma_i^2)} = \frac{n}{\sum (1/\sigma_i^2)} \frac{1}{\sigma_i^2}$$

Daraus folgt für die Konstante c :

$$c = \frac{n}{\sum (1/\sigma_i^2)}$$

Dies entspricht einer Normalisierung der Gewichtungsfaktoren auf deren Mittelwert.

⁴ Nach Fahrmeir (2009) (Seite 463) gilt für die Linearkombination eines Zufallsvektors \mathbf{X} mit einer geeignet dimensionierten Matrix \mathbf{A} und einem geeignet dimensionierten Vektor \mathbf{b} :

$$\begin{aligned} E[\mathbf{A}\mathbf{X} + \mathbf{b}] &= \mathbf{A} E[\mathbf{X}] + \mathbf{b} \\ \text{Cov}[\mathbf{A}\mathbf{X} + \mathbf{b}] &= \mathbf{A} \text{Cov}[\mathbf{X}] \mathbf{A}^T \end{aligned}$$

Damit ergibt sich mit $\Sigma = \text{Cov}[\mathbf{y}]$:

$$\begin{aligned} \text{Cov}[\hat{\theta}] &= \text{Cov}[(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y}] \\ &= (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \Sigma ((\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P})^T \end{aligned}$$

Mit $\Sigma = \hat{\sigma}^2 \mathbf{P}^{-1}$ folgt daraus:

$$\begin{aligned} \text{Cov}[\hat{\theta}] &= \hat{\sigma}^2 (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{P} \mathbf{A}) (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \\ &= \hat{\sigma}^2 (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \end{aligned}$$

da die Matrizen \mathbf{P} und $\mathbf{A}^T \mathbf{P} \mathbf{A}$ symmetrisch sind.

⁵ Die Maximum-Likelihood-Schätzung

$$\hat{\sigma}^2 = \frac{1}{n} \hat{\epsilon}^T \mathbf{P} \hat{\epsilon}$$

ist nicht erwartungstreu. Im allgemeinen wird sie deswegen durch die Restringierte Maximum-Likelihood-Schätzung

$$\hat{\sigma}^2 = \frac{1}{n-p} \hat{\epsilon}^T \mathbf{P} \hat{\epsilon}$$

ersetzt. (Fahrmeir (2009) Seite 94, Wakefield (2013) Seite 44 ff.)

⁶ Im Praktikum wird meist mit (1σ) -Fehlerintervallen gearbeitet. Das zugehörige Signifikanzniveau ist 68.2689%, woraus $\alpha = 31.7311\%$ folgt.

⁷ Standard t-Verteilung $x \sim t_n$ (z.B. Fahrmeir (2009) Seite 461)

$$p(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

⁸ Zu bedingten Wahrscheinlichkeiten und zum Bayes-Theorem siehe z.B. K. Koch (2000) (Kapitel 2) oder Wakefield (2013) Kapitel 3

⁹Die in den Gleichungen 10 und 11 genutzte, frei wählbare Konstante c beeinflusst den Varianzfaktor σ^2 , der auch als Varianz der Gewichtseinheit bezeichnet wird.

¹⁰ Die Umformung erfolgt in mehreren Schritten. Dabei wird ausgenutzt, dass die Gewichtsmatrix \mathbf{P} entsprechend ihrer Definition symmetrisch ist. Daraus ergibt

sich , dass auch die Matrix $(\mathbf{A}^T \mathbf{P} \mathbf{A})$ symmetrisch ist.

entwickle:

$$\begin{aligned} & (\mathbf{y} - \mathbf{A} \theta)^T \mathbf{P} (\mathbf{y} - \mathbf{A} \theta) \\ &= \mathbf{y}^T \mathbf{P} \mathbf{y} - \theta^T \mathbf{A}^T \mathbf{P} \mathbf{y} - \mathbf{y}^T \mathbf{P} \mathbf{A} \theta + \theta^T \mathbf{A}^T \mathbf{P} \mathbf{A} \theta \end{aligned}$$

alle Terme sind Skalare $\rightarrow \theta^T \mathbf{A}^T \mathbf{P} \mathbf{y} = \mathbf{y}^T \mathbf{P} \mathbf{A} \theta$

$$= \mathbf{y}^T \mathbf{P} \mathbf{y} - 2\theta^T \mathbf{A}^T \mathbf{P} \mathbf{y} + \theta^T \mathbf{A}^T \mathbf{P} \mathbf{A} \theta$$

mit: $2\theta^T \mathbf{A}^T \mathbf{P} \mathbf{y} = 2\theta^T (\mathbf{A}^T \mathbf{P} \mathbf{A}) (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y}$

und: $\mu_0 = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y}$

$$= \mathbf{y}^T \mathbf{P} \mathbf{y} - 2\theta^T (\mathbf{A}^T \mathbf{P} \mathbf{A}) \mu_0 + \theta^T \mathbf{A}^T \mathbf{P} \mathbf{A} \theta$$

setze:

$$\begin{aligned} & \mathbf{y}^T \mathbf{P} \mathbf{y} \\ &= \mathbf{y}^T \mathbf{P} \mathbf{y} - \mathbf{y}^T \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} + \mathbf{y}^T \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} \\ &= \mathbf{y}^T (\mathbf{P} - \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}) \mathbf{y} + \mathbf{y}^T \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} \end{aligned}$$

aus

$$\begin{aligned}
& (\mathbf{y} - \mathbf{A} \mu_0)^T \mathbf{P} (\mathbf{y} - \mathbf{A} \mu_0) \\
&= (\mathbf{y} - \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y})^T \mathbf{P} (\mathbf{y} - \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y}) \\
&= \mathbf{y}^T (\mathbf{I} - \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T) \mathbf{P} (\mathbf{I} - \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}) \mathbf{y} \\
&= \mathbf{y}^T (\mathbf{P} - 2 \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \\
&\quad + \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}) \mathbf{y} \\
&= \mathbf{y}^T (\mathbf{P} - \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}) \mathbf{y}
\end{aligned}$$

und

$$\begin{aligned}
& \mathbf{y}^T \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} \\
&= \mathbf{y}^T \mathbf{P} \mathbf{A} (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{P} \mathbf{A}) (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{y} \\
&= \mu_0^T (\mathbf{A}^T \mathbf{P} \mathbf{A}) \mu_0
\end{aligned}$$

folgt mit: $(n - p) s^2 = (\mathbf{y} - \mathbf{A} \mu_0)^T \mathbf{P} (\mathbf{y} - \mathbf{A} \mu_0)$

$$\begin{aligned}
& \mathbf{y}^T \mathbf{P} \mathbf{y} \\
&= (n - p) s^2 + \mu_0^T (\mathbf{A}^T \mathbf{P} \mathbf{A}) \mu_0
\end{aligned}$$

damit ergibt sich:

$$\begin{aligned}
& (\mathbf{y} - \mathbf{A} \theta)^T \mathbf{P} (\mathbf{y} - \mathbf{A} \theta) \\
&= (n - p) s^2 + \mu_0^T (\mathbf{A}^T \mathbf{P} \mathbf{A}) \mu_0 - 2 \theta^T (\mathbf{A}^T \mathbf{P} \mathbf{A}) \mu_0 + \theta^T \mathbf{A}^T \mathbf{P} \mathbf{A} \theta \\
&= (n - p) s^2 + (\theta - \mu_0)^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\theta - \mu_0)
\end{aligned}$$

¹¹ Normal-inverse Gamma-Verteilung $\theta, \sigma^2 \sim \text{NIG}(\mathbf{m}, \mathbf{\Sigma}, a, b)$
(z.B. L. Fahrmeier u.a. (2013) Seite 652):

$$\begin{aligned}
p(\theta, \sigma^2) &= \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}|^{1/2}} \frac{b^a}{\Gamma(a)} \\
&\quad \frac{1}{(\sigma^2)^{a+1}} \exp\left(-\frac{b}{\sigma^2} - \frac{1}{2} (\theta - \mathbf{m})^T \mathbf{\Sigma}^{-1} (\theta - \mathbf{m})\right)
\end{aligned}$$

Für die praktische Anwendung wird oftmals $\mathbf{\Sigma} = \sigma^2 \mathbf{M}$ gesetzt und $\theta, \sigma^2 \sim \text{NIG}(\mathbf{m}, \mathbf{M}, a, b)$ verwendet (z.B. Fahrmeier (2013) Seite 227).

$$\begin{aligned}
p(\theta, \sigma^2) &= \frac{1}{(2\pi)^{p/2} |\mathbf{M}|^{1/2}} \frac{b^a}{\Gamma(a)} \\
&\quad \frac{1}{(\sigma^2)^{p/2+a+1}} \exp\left(-\frac{b}{\sigma^2} - \frac{1}{2\sigma^2} (\theta - \mathbf{m})^T \mathbf{M}^{-1} (\theta - \mathbf{m})\right)
\end{aligned}$$

¹² Inverse Gamma-Verteilung $x \sim \text{InvGa}(a, b)$

(z.B. Fahrmeir (2009) Seite 461):

$$\begin{aligned}
 p(x) &= \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp(-b/x), \quad x > 0 \\
 \text{mit:} \quad E[x] &= \frac{b}{a-1} \\
 \text{es gilt:} \quad &\rightarrow \int p(x) dx = 1 \\
 \text{daraus folgt:} \quad &\rightarrow \int x^{-(a+1)} \exp(-b/x) dx = \Gamma(a) b^{-a}
 \end{aligned}$$

¹³ p -dimensionale t-Verteilung $\mathbf{x} \sim T_p(\mu, \Sigma, d)$
(z.B. Fahrmeir (2009) Seite 467)

$$p(\mathbf{x}) = \frac{\Gamma[(d+p)/2]}{\Gamma[d/2](d\pi)^{p/2}} |\Sigma|^{-1/2} \left[1 + \frac{(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)}{d} \right]^{-(d+p)/2}$$

¹⁴ Zu der gleichen Verteilung $\theta \mid \mathbf{y}$ gelangt Koch (2000) (Seite 111 ff) mit einer anderen Priori-Verteilung. Er substituiert den Varianzfaktor σ^2 mit dem Gewichtsparameter $\tau = 1/\sigma^2$ und benutzt die Priori-Verteilungen

$$\begin{aligned}
 p(\theta) &\propto \text{const} \quad \text{für} \quad -\infty < \theta_i < \infty, \quad i = 1, \dots, n \\
 p(\tau) &\propto \tau^{-1} \quad \text{für} \quad 0 < \tau < \infty \\
 p(\theta, \tau) &= p(\theta) \times p(\tau) \propto \tau^{-1}
 \end{aligned}$$

Daraus ergibt sich die Posteriori-Verteilung zu:

$$p(\theta, \tau \mid \mathbf{y}) \propto \tau^{n/2-1} \exp\left(-\frac{\tau}{2} (\mathbf{y} - \mathbf{A}\theta)^T \mathbf{P} (\mathbf{y} - \mathbf{A}\theta)\right)$$

Der Exponent wird in der gleichen Weise umgeformt¹⁰. Damit lässt sich die Posteriori-Verteilung als Normal-Gammaverteilung darstellen. Aus dieser folgt dann die mehrdimensionale t-Verteilung:

$$\theta \mid \mathbf{y} \sim T_p(\mu_0, s^2(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1}, n-p)$$

als Posteriori-Randverteilung für θ .

¹⁵ Fahrmeir (2013) (Seite 227 ff.) verwendete als konjugierte Priori-Verteilung eine Normal-inverse Gammaverteilung $\theta, \sigma^2 \sim \text{NIG}(\mathbf{m}, \mathbf{M}, a, b)$. Aus dieser folgt als Posteriori-Verteilung wieder eine Normal-inverse Gammaverteilung. Mit den

von Fahrmeir (2013) auf Seite 231 angegebenen Werten $\mathbf{m} = \mathbf{0}$, $\mathbf{M}^{-1} = \mathbf{0}$, $a = -p$ und $b = 0$ folgt als Priori-Verteilung:

$$p(\theta, \sigma^2) \propto \frac{1}{(\sigma^2)^{-p/2+1}} = (\sigma^2)^{p/2-1}$$

die nicht mit der von Fahrmeir (2013)(Gleichung. 4.16) angestrebten, nichtinformativen Priori-Verteilung

$$p(\theta, \sigma^2) \propto \frac{1}{\sigma^2}$$

übereinstimmt. Diese ist mit der Vorgabe $a = -p/2$ zu erreichen. Daraus ergibt sich dann als Posteriori-Verteilung

$$p(\theta, \sigma | \mathbf{y}) \propto \frac{1}{(\sigma^2)^{n/2+1}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{A} \theta)^T \mathbf{P} (\mathbf{y} - \mathbf{A} \theta)\right)$$

die sich durch die Normal-inverse Gammaverteilung $\text{NIG}(\mathbf{m}, \mathbf{M}, a, b)$ mit

$$\begin{aligned} \text{mit } \mathbf{m} &\rightarrow \mu_0, & \mathbf{M} &\rightarrow (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \\ \text{und } a &\rightarrow \frac{n-p}{2}, & b &\rightarrow \frac{n-p}{2} s^2 \end{aligned}$$

beschreiben lässt.

¹⁶ Wakefield (2013)(Seite 222) kommt auf eine inverse Chi Quadrat Verteilung mit $n - p$ Freiheitsgraden

$$\sigma^2 | \mathbf{y} \sim (n - p) s^2 \times \chi_{n-p}^{-2}$$

wobei zu beachten ist, dass Wakefield (2013) den Parametervektor θ von $0 \dots k = p_W - 2$ indiziert und dort in p_W , der Anzahl der unbekannt Parameter, σ enthalten ist (Seite 209).

Der Erwartungswert der inversen Chi Quadrat Verteilung mit k Freiheitsgraden ist gegeben durch

$$\text{E}[\chi_k^{-2}] = \frac{1}{k - 2}$$

Daraus folgt:

$$\hat{\sigma}_B^2 = \frac{n - p}{n - p - 2} s^2$$

¹⁷ Bei der Herleitung wird ausgenutzt, dass die Matrizen \mathbf{P} und $\mathbf{A}^T \mathbf{P} \mathbf{A}$ sym-

metrisch sind.

$$\begin{aligned}\text{Cov}[\theta] &= \mathbf{H} \boldsymbol{\Sigma} \mathbf{H}^T \\ &= (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \boldsymbol{\Sigma} ((\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P})^T \\ &= (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \boldsymbol{\Sigma} \mathbf{P}^T \mathbf{A} ((\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1})^T \\ &\quad \text{mit : } \boldsymbol{\Sigma} = c \mathbf{P}^{-1} \\ &= c (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{P} \mathbf{A}) (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \\ &= c (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1}\end{aligned}$$

Literatur

Physikalisches Grundpraktikum: Einführung in die Messung, Auswertung und Darstellung experimenteller Ergebnisse in der Physik
Mathematisch-Naturwissenschaftliche Fakultät I der HUB
Institut für Physik **2007**
<http://gpr.physik.hu-berlin.de/Skripten/Einfuehrung/PDF-Datei/Einfuehrung.pdf> abgerufen am 4.1.2016 10:39 Uhr

Physikalisches Grundpraktikum: Einführungspraktikum
Mathematisch-Naturwissenschaftliche Fakultät I der HUB
Institut für Physik **2007**
<http://gpr.physik.hu-berlin.de/Skripten/Einfuehrungspraktikum/PDF-Dateien/Einfuehrungspraktikum.pdf>

R.M. Barnett et al.
Review of Particle Properties
28. STATISTICS page 159
Physical Review D Volume **54**, 1 **1996**

J. Beringer et al. (Particle Data Group)
REVIEW OF PARTICLE PHYSICS
36. STATISTICS page 390
Physical Review D Volume **86**, 010001 **2012**

P.R. Bevington and D.K. Robinson
Data Reduction and Error Analysis for the Physical Sciences
McGraw-Hill Book Co., New York, 3. edition **2003**

Frederick James
Statistical Methods in Experimental Physics
World Scientific Publishing, Singapore, 2nd edition **2006**

Ludwig Fahrmeir, Thomas Kneib, Stefan Lang
Regression Modelle, Methoden und Anwendungen
Springer-Verlag, Berlin Heidelberg 2. Auflage **2009**

Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, Brian Marx
Regression Models, Methods and Applications
Springer-Verlag, Berlin Heidelberg First edition **2013**

Michael Grabe
Grundriss der Generalisierten Gauß´chen Fehlerrechnung
Springer-Verlag, Heidelberg Dordrecht London **2011**

Karl-Rudolf Koch

Einführung in die Bayes-Statistik

Springer Verlag Berlin Heidelberg 1.Auflage **2000**

Brain R. Martin

Statistics for physical Sciences

Academic Press Elsevier, Amsterdam, First edition **2012**

Volker Nollau

Statistische Analysen

Mathematische Methoden der Planung und Auswertung von Versuchen

Fachbuchverlag Leipzig 1.Auflage **1975**

Wolfgang Schenk, Friedrich Kremer, Gunter Beddies, Thomas Franke, Petrik Galvosas, Peter Rieger

Physikalisches Praktikum 14., überarbeitete und erweiterte Auflage

Springer Fachmedien Wiesbaden **2014**

Jon Wakefield

Bayesian and Frequentist Regression Methods

Springer Verlag New York Heidelberg 1.Edition **2013**

Ulli Wolf, Peter Galler, Tomasz Korzec

Vorlesung Computational Physics I

Institut für Physik der HUB **2014**

<http://www.physik.hu-berlin.de/com/teachingandseminars/previousCPI>

abgerufen am 4.1.2016 11:31 Uhr

Wolfram Research, Inc.

Mathematica, Version 10.0, Champaign, IL **2014**

<http://www.wolframalpha.com>

Ion Vasilief, Scott Howard (Debian Maintainer)

QtiPlot Version 0.9.8.9-9+b3

Released: 2014/04/29

ubuntu package: qtiplot_0.9.8.9-9+b3_amd64.deb

<http://www.qtiplot.com>

Russell Standish, Georges Khaznadar (Debian Maintainer)

SciDAVis Version 1.D9-1

Released: 2015/12/10

Debian package: scidavis_1.D9-1_amd64.deb

<http://scidavis.sourceforge.net>

Anton Gladky (Debian Maintainer)
GnuPlot Version 5.0.1+dfsg1-3
Released: 2015/08/11
Debian package: gnuplot5_5.0.1+dfsg1-3_all.deb
<http://gnuplot.sourceforge.net>

GNU Scientific Library Reference Manual
Edition 1.15 for GSL Version 1.15, 29 April 2011
Abschnitt 37 page 403
<https://www.gnu.org/s/gsl/manual/gsl-ref.pdf>

6 Anhang

```

lineare Regression von DataLinReg.dat

LinearFit mit Option "Scale Errors with sqrt(Chi2/doF)"

Linear Regression using function: A*x+B
B (y-intercept) = 1.957232142857143e+02 +/- 2.252921990469467e-01
A (slope) = -1.001488095238095e+01 +/- 4.461450346762922e-02
-----
Chi^2/doF = 1.671981292517004e-01
R^2 = 0.999722241130897
Adjusted R^2 = 0.999679508997189
RMSE (Root Mean Squared Error) = 0.408898678466562
RSS (Residual Sum of Squares) = 2.34077380952381
-----

LinearFit ohne Option "Scale Errors with sqrt(Chi2/doF)"

Linear Regression using function: A*x+B
B (y-intercept) = 1.957232142857143e+02 +/- 2.252921990469467e-01
A (slope) = -1.001488095238095e+01 +/- 4.461450346762922e-02
-----
Chi^2/doF = 1.671981292517004e-01
R^2 = 0.999722241130897
Adjusted R^2 = 0.999679508997189
RMSE (Root Mean Squared Error) = 0.408898678466562
RSS (Residual Sum of Squares) = 2.34077380952381
#####

gewichtete lineare Regression von DataWeightedLinReg1.dat
(gewichtet mit systematischem Restfehler)

LinearFit mit Option "Scale Errors with sqrt(Chi2/doF)"

Linear Regression using function: A*x+B
Weighting Method: Instrumental, using error bars
B (y-intercept) = 1.957067396883333e+02 +/- 1.566352195152149e-01
A (slope) = -1.001137532586336e+01 +/- 3.001222094441872e-02
-----
Chi^2/doF = 2.225043810434601e+00
R^2 = 0.999722071005109
Adjusted R^2 = 0.999679312698202
RMSE (Root Mean Squared Error) = 1.49165807423639
RSS (Residual Sum of Squares) = 31.1506133460845
(ok)-----

LinearFit ohne Option "Scale Errors with sqrt(Chi2/doF)"

Linear Regression using function: A*x+B
Weighting Method: Instrumental, using error bars
B (y-intercept) = 1.957067396883333e+02 +/- 1.566352195152149e-01
A (slope) = -1.001137532586336e+01 +/- 3.001222094441872e-02
-----
Chi^2/doF = 2.225043810434601e+00
R^2 = 0.999722071005109
Adjusted R^2 = 0.999679312698202
RMSE (Root Mean Squared Error) = 1.49165807423639
RSS (Residual Sum of Squares) = 31.1506133460845
(ok)-----

```

Abbildung 1: Ergebnisse QtiPlot

F4 Federkonstante statisch

Alle Längen in mm, alle Massen in g

```
In[1]= uSpiegelscala[x_] := 0.2 + 5 * 10^-4 * x;
```

Anleitung Einführungspraktikum (2007) Seite 11

Lineare Regression

Einführung (2007): Seite 42 Gln. 49, 50 + Seite 42 Gln unter Gln 51

Nollau: Seiten 129...151

Fahrmeir Seiten 104...120

Schenk: Seiten 9 und 17...18

```
In[2]= linRegScript[data_] := Module[{x, y, a, b, sa, sb, n, xi, xxi, yi, xyi, D, sy},
  x = Transpose[data][[1]];
  y = Transpose[data][[2]];
  n = Length[data];
  xi = Sum[x[[i]], {i, n}];
  xxi = Sum[x[[i]]^2, {i, n}];
  yi = Sum[y[[i]], {i, n}];
  xyi = Sum[x[[i]] * y[[i]], {i, n}];
  D = n * xxi - xi^2;
  a = (n * xyi - xi * yi) / D;
  b = (xxi * yi - xi * xyi) / D;
  sy = Sqrt[Sum[(y[[i]] - a * x[[i]] - b)^2, {i, n}] / (n - 2)];
  sa = sy * Sqrt[n / D];
  sb = sy * Sqrt[xxi / D];
  {a, sa, b, sb}];
```

gewichtete lineare Regression

Einführung (2007) Seite 41 Gln. 43 + Seite 42 Gln. 45

```
In[3]= WeightedlinRegScript[data_, s_] :=
  Module[{x, y, a, b, sa, sb, n, ni, xi, xxi, yi, xyi, D},
  x = Transpose[data][[1]];
  y = Transpose[data][[2]];
  n = Length[data];
  ni = Sum[1 / s[[i]]^2, {i, n}];
  xi = Sum[x[[i]] / s[[i]]^2, {i, n}];
  xxi = Sum[x[[i]]^2 / s[[i]]^2, {i, n}];
  yi = Sum[y[[i]] / s[[i]]^2, {i, n}];
  xyi = Sum[x[[i]] * y[[i]] / s[[i]]^2, {i, n}];
  D = ni * xxi - xi^2;
  a = (ni * xyi - xi * yi) / D;
  b = (xxi * yi - xi * xyi) / D;
  sa = Sqrt[ni / D];
  sb = Sqrt[xxi / D];
  {a, sa, b, sb}];
```

Nollau : Seiten 152 ... 156

Fahrmeir Seiten 178 ... 182

Abbildung 2: Mathematica Notebook Seite 1

```
In[4]= WeightedlinRegNollau[data_, s_] :=
Module[{x, y, a, b, sa, sb, n, ni, xi, xxi, yi, xyi, D, sy},
  x = Transpose[data][[1]];
  y = Transpose[data][[2]];
  n = Length[data];
  ni = Sum[1 / s[[i]]^2, {i, n}];
  xi = Sum[x[[i]] / s[[i]]^2, {i, n}];
  xxi = Sum[x[[i]]^2 / s[[i]]^2, {i, n}];
  yi = Sum[y[[i]] / s[[i]]^2, {i, n}];
  xyi = Sum[x[[i]] * y[[i]] / s[[i]]^2, {i, n}];
  D = ni * xxi - xi^2;
  a = (ni * xyi - xi * yi) / D;
  b = (xxi * yi - xi * xyi) / D;
  sy = Sqrt[Sum[(y[[i]] - a * x[[i]] - b) / s[[i]]^2, {i, n}] / (n - 2)];
  sa = sy * Sqrt[ni / D];
  sb = sy * Sqrt[xxi / D];
  {a, sa, b, sb}];
```

Mittelwert, Standardabweichung der verwendeten insgesamt 32 Massestücke

```
In[5]= massePlatz1 = {50.169, 50.954, 50.322, 50.719, 50.207, 50.034, 49.996, 50.612};
massePlatz2 = {50.651, 50.598, 49.840, 50.502, 50.333, 50.016, 49.948, 49.806};
massePlatz3 = {50.871, 50.676, 50.251, 49.966, 50.047, 49.970, 49.945, 49.978};
massePlatz4 = {49.922, 49.958, 49.966, 49.562, 49.988, 50.813, 49.968, 50.400};
masse = Join[massePlatz1, massePlatz2, massePlatz3, massePlatz4];
{Mean[masse], StandardDeviation[masse]}

Out[10]= {50.2184, 0.35794}
```

Messdatenreihe vom 7.1.2015, Meßplatz 2

Anleitung Einführungspraktikum (2007) Seite 11

Die Feder wird schrittweise bis 400 g mit Massestücken von 50 g belastet

und mit Hilfe einer Spiegelskala an der Marke M die Auslenkung x gemessen (erste Messreihe).

Nach Erreichen der Höchstbelastung von 400 g ist dann die Feder wieder schrittweise zu entlasten (zweite Messreihe).

```
In[11]= messdaten = {{186.0, 175.5, 165.0, 155.0, 145.5, 135.5, 126.0, 116.0},
  {115.5, 125.0, 136.0, 146.0, 156.0, 165.5, 176.0, 186.0}};

In[12]= n1 = Length[messdaten[[1]];
n2 = Length[messdaten[[2]];
xlist = Join[Table[i, {i, n1}], Table[i, {i, n2, 1, -1}]];
ylist = Flatten[messdaten];
fitlist = Transpose[{xlist, ylist}]

Out[16]= {{1, 186.}, {2, 175.5}, {3, 165.}, {4, 155.}, {5, 145.5},
  {6, 135.5}, {7, 126.}, {8, 116.}, {8, 115.5}, {7, 125.},
  {6, 136.}, {5, 146.}, {4, 156.}, {3, 165.5}, {2, 176.}, {1, 186.}}
```

Abschätzung des y-Fehlers aus dem systematischen Restfehler

```
In[17]= usysResty = uSpiegelskala[ylist]

Out[17]= {0.293, 0.28775, 0.2825, 0.2775, 0.27275, 0.26775, 0.263,
  0.258, 0.25775, 0.2625, 0.268, 0.273, 0.278, 0.28275, 0.288, 0.293}
```

Abbildung 3: Mathematica Notebook Seite 2

Lineare Regression

```
In[18]= fitresult = LinearModelFit[fitlist, x, x];
Print[fitresult["ParameterTable"]];
```

	Estimate	Standard Error	t-Statistic	P-Value
1	195.723	0.225292	868.753	1.58279×10^{-34}
x	-10.0149	0.0446145	-224.476	2.67212×10^{-26}

```
In[20]= {a, ua, b, ub} = linRegScript[fitlist];
Print["linRegScript Gln.49 50 und 51a : y = a ' x + b "];
Print[TableForm[{b, ub}, {a, ua}],
TableHeadings -> {"b", "a"}, {"Estimate", "StandardError"}]];
linRegScript Gln.49 50 und 51a : y = a ' x + b
```

	Estimate	StandardError
b	195.723	0.225292
a	-10.0149	0.0446145

gewichtete lineare Regression (mit systematischem Restfehler)

```
In[23]= weightedfitresult = LinearModelFit[fitlist, x, x, Weights -> 1 / (usysResty) ^ 2];
Print[weightedfitresult["ParameterTable"]];
```

	Estimate	Standard Error	t-Statistic	P-Value
1	195.707	0.233646	837.62	2.6382×10^{-34}
x	-10.0114	0.044768	-223.628	2.81746×10^{-26}

```
In[25]= {aS, uaS, bS, ubS} = WeightedlinRegScript[fitlist, usysResty];
Print["WeightedlinRegScript Gln.43 ... 45 : y = a ' x + b "];
Print[TableForm[{bS, ubS}, {aS, uaS}],
TableHeadings -> {"b", "a"}, {"Estimate", "StandardError"}]];
WeightedlinRegScript Gln.43 ... 45 : y = a ' x + b
```

	Estimate	StandardError
b	195.707	0.156635
a	-10.0114	0.0300122

```
In[28]= {aN, uaN, bN, ubN} = WeightedlinRegNollau[fitlist, usysResty];
Print["WeightedlinRegNollau pp. 152 - 155 : y = a ' x + b "];
Print[TableForm[{bN, ubN}, {aN, uaN}],
TableHeadings -> {"b", "a"}, {"Estimate", "StandardError"}]];
WeightedlinRegNollau pp. 152 - 155 : y = a ' x + b
```

	Estimate	StandardError
b	195.707	0.233646
a	-10.0114	0.044768

gewichtete lineare Regression (mit systematischem Restfehler) VarianceEstimatorFunction-> (1&)

```
In[31]= weightedfitresult = LinearModelFit[fitlist, x, x,
Weights -> 1 / (usysResty) ^ 2, VarianceEstimatorFunction -> (1 &)];
Print[weightedfitresult["ParameterTable"]];
```

	Estimate	Standard Error	t-Statistic	P-Value
1	195.707	0.156635	1249.44	9.77167×10^{-37}
x	-10.0114	0.0300122	-333.577	1.04454×10^{-28}

Abbildung 4: Mathematica Notebook Seite 3

qtiplot/src/analysis

Class Fit

enthält protected Variable

bool d_scale_errors; Fit.h: 251

und public function Fit.h: 133

```
void scaleErrors (bool yes=true)
    {d_scale_errors = yes;};
```

Variable d_scale_errors wird in

```
void Fit::init ()
```

```
    d_scale_errors = false; Fit.cpp: 94
```

gesch.

* init wird in Constructor Fit::Fit aufgerufen

↳ Defaultmäßig ist d_scale_errors = false

scale_Errors wird aufgerufen in

```
ApplicationWindow.cpp: 13381 * (1)
```

```
fitter -> scaleErrors (Fit_scale_errors); ==>
```

```
↑
class Fit
```

```
public: bool Fit_scale_errors; ApplicationWindow.cpp: 1310
```

```
    Fit_scale_errors = true; ApplicationWindow.cpp: 731
```

QtiPlot setzt Vorgabewert auf true

dieser muss vom gewählten Fitter übernommen werden !

Abbildung 5: Quellcodeanalyse Seite 1

*1

src/core/ApplicationWindow.cpp : 13549

```
if (operation != FitLinear && operation != FitSlope)
{
    Fitter -> guessInitialValues();
    Fitter -> scaleErrors (fit_scale_errors);
    Fitter -> generateFunction (generateUniformFitPoints,
                              fitPoints);
} else if (d_2_Linear_fit_points)
```

13381:

```
    Fitter -> generateFunction (generateUniformFitPoints,
```

LinearFit und LinearSlopeFit übernehmen nicht den Wert aus dem ApplicationWindow.
Der default Wert false aus Class Fit wird nicht überschrieben.

13328: void ApplicationWindow::analyzeCurve

(Graph *g, QwtPlotCurve *C, Analysis operation)

```
switch (operation) {
```

```
    case FitLinear:
```

13353: Fitter = new LinearFit(this, g); #

```
    break;
```

```
    case FitSlope:
```

14492: Fitter = new LinearSlopeFit(this, g); #

```
    break;
```

```
# defined in src/analysis/PolynomialFit.h
```

Abbildung 6: Quellcodeanalyse Seite 2

SVC/analysis/PolynomialFit.h

```
65: class LinearFit : public Fit
    {
        void fit();
    private:
        void init()
    }

```

```
85: class LinearSlopeFit : public Fit
    {
        void fit();
    private:
        void init();
    }

```

PolynomialFit.cpp

```
242: void LinearFit::init() {
```

```
244:     d_scale_errors = false
```

```
378 void LinearSlopeFit::init() {
```

```
380:     d_scale_errors = false
```



Abbildung 7: Quellcodeanalyse Seite 3

svc/analysis/PolynomialFit.epp
 293: void LinearFit::Fit() {
 305 ~~if~~ if (d_weighting == NoWeighting)
 gsl_fit_linear(...) GSL-Reference p 403
 else
 gsl_fit_wlinear(...) GSL-Reference p 403
 weighted data =>
 ~~Variance~~ Covariance Matrix

$$C_{ab} = \sum_i \frac{1}{w_i} \frac{\partial c_a}{\partial y_i} \frac{\partial c_b}{\partial y_i} \quad w_i = \frac{1}{\sigma_i^2}$$

 unweighted data => Variance-covariance Matrix

$$w_i = \frac{1}{\sigma^2} \text{ mit}$$

$$\sigma^2 = \frac{\sum (y_i - Y(c, x_i))^2}{(n-p)}$$

$$C_{ab} = \sigma^2 \sum_i \frac{\partial c_a}{\partial y_i} \frac{\partial c_b}{\partial y_i}$$

 Covariance Matrix

$$\sigma_{c_a} = \sqrt{C_{aa}}$$

=> weighted Data : entspricht Scale Errors = No
 unweighted Data : entspricht Scale Errors = Yes
 Option „Scale Errors“ ist wirkungslos

Notwendig, da Festlegung durch die Wahl der GSL-Funktion vorgegeben ist. Eine Anwendung von Scale Errors=true auf unweighted Data führt zu falschen Ergebnissen.

Abbildung 8: Quellcodeanalyse Seite 4