

Computerpraktikum im GP II

Lineare Regression

Autor Daniel Brete (2004)
Anpassung Christoph Kohstall

22. September 2021

Dieses Skript gibt eine Einführung in die linearen Regression auf Grundlage der statistischen Maximum-Likelihood-Methode. Beweise und Ergänzungen, die für ein Verständnis der prinzipiellen Überlegungen der Maximum-Likelihood-Methode nicht unbedingt erforderlich sind, sind mit * gekennzeichnet.

Dank

Die Idee und ersten Versuche zu dieser Übung sind gemeinsam mit Jens Koesling entstanden, ohne den ich dieses Projekt nie angefangen hätte. Großer Dank gebührt Michael Karcher, der uns viele mathematische Fragen erklärt hat und uns bei Schwierigkeiten mit Mathematica immer wieder aus der Patsche geholfen hat. Tobias Burnus hat sich stets Zeit genommen, um uns bei unzähligen Computerproblemen zu helfen. Tristan Faber schließlich hat mich, nach dem er diese Übung betreut hat, davon überzeugt, das Skript völlig neu zu gliedern und zu überarbeiten.

Inhaltsverzeichnis

1	Warum Sie diesen Versuch machen	3
2	Grundlagen	3
2.1	Der Messprozess als Zufallsexperiment	3
2.2	Kontinuierliche Verteilung (Wahrscheinlichkeitsdichtefunktion)	3
2.3	Gaußverteilung	4

2.4	Erwartungswert	5
2.5	Varianz, Standardabweichung, physikalische Fehler	5
2.6	Gaußsches Fehlerfortpflanzungsgesetz	6
3	Parameterschätzung	7
3.1	Schätzfunktion	7
3.2	Maximum-Likelihood-Methode	8
3.2.1	am Beispiel einfacher Mittelwert	8
3.2.2	Der gewichtete Mittelwert	9
3.2.3	Schätzung des Fehlers einer Messung aus der Streuung .	10
3.3	Das Problem der Erwartungstreue, Bessels Korrektur	12
4	Der Fehler der geschätzten Parameter	14
4.1	Bekannte Messfehler	14
4.1.1	Fehler gleich groß	14
4.1.2	Fehler unterschiedlich groß	15
4.2	Unbekannte Fehler	15
4.3	*Bekannte Fehler / aus der Streuung geschätzt Fehler	16
5	Lineare Regression	17
5.1	Schätzung der Parameter A und B der Geraden	17
5.2	Berechnung der Fehler der Parameter	21
5.2.1	a priori-Fehler	21
5.2.2	Fehler aus der Streuung	22
6	Mehr zur linearen Regression	25
6.1	Die Bedeutung von χ^2 bei <i>bekannten</i> Fehlern	25
6.2	Falsches Modell	25
6.3	Das Bestimmtheitsmaß R^2	26
6.4	*Fehler in den x -Werten	27
6.5	* x - und y -Werte fehlerbehaftet	27
6.6	*Interpolation und Kalibrierkurven	28
6.7	*Ausblick: Matrixdarstellung und nichtlineare Modelle	28
	Literatur	28

1 Warum Sie diesen Versuch machen

Das Ziel dieser Praktikumsaufgabe ist es, Sie soweit mit den mathematischen Grundlagen der linearen Regression und der Software *Python* vertraut zu machen, dass Sie diese Technik anstelle grafischer Auswertungen bei den folgenden Praktikumsversuchen sinnvoll einsetzen können.

Das Skript zu dieser Aufgabe ist wesentlich umfangreicher, als Sie es von anderen Versuchen gewohnt sind. Das liegt zum einen daran, dass wir uns bemüht haben die theoretischen Grundlagen ausführlich darzustellen und nicht nur summarisch zu wiederholen, zum anderen erhalten Sie eine Anleitung zu *Python*, die die Lösung der ersten Teilaufgabe Schritt für Schritt vorführt. Mehr Text bedeutet für Sie in diesem Fall also hoffentlich nicht mehr sondern weniger Arbeit.

2 Grundlagen

In diesem Abschnitt werden einige Begriffe aus der Einführung in die Fehlerrechnung und Statistik im Grundpraktikum I wiederholt und vertieft.

2.1 Der Messprozess als Zufallsexperiment

Messen wir dieselbe Größe mehrfach, erhalten wir im allgemeinen verschiedene Messwerte (x_1, x_2, \dots) . Mathematisch betrachtet man die Messung daher als Zufallsexperiment. Das Ergebnis X der Messung heißt *Zufallsgröße*. Jedes Messergebnis x_i ist eine Realisation der Zufallsgröße. Geht man von einer diskreten Zufallsgröße aus, wird bei jeder Messung genau einer der Werte $x_i \in \{x_1, \dots, x_n\}$ angenommen. (z.B. Würfelexperiment) Bei einer Messung beobachtet man dann mit der Wahrscheinlichkeit p_i den Wert x_i .

2.2 kontinuierliche Verteilung (Wahrscheinlichkeitsdichtefunktion)

In der Praxis kommen meist kontinuierliche Größen vor. Hier kann man keine Wahrscheinlichkeit für einen diskreten Wert mehr angeben. Man verwendet stattdessen eine *Wahrscheinlichkeitsdichtefunktion* $P(x)$. Die Wahrscheinlichkeit einen Wert im Intervall $[x_a, x_b]$ zu messen, ist dann das bestimmte Integral über die Wahrscheinlichkeitsdichtefunktion. Ähnlich wie man bei der Integration über die Dichte die Masse erhält, erhält man bei der Integration über die

W-Dichte eine Wahrscheinlichkeit. [Bar, chap. 3.1.4]

$$p(x \in [x_a, x_b]) = \int_{x_a}^{x_b} P(x) dx \quad (1)$$

Genau wie bei diskreten Verteilungen die Summe über die Wahrscheinlichkeiten aller möglichen Ereignisse 1 ist, ist das Integral über den gesamten Wertebereich der Zufallsgröße 1. Man sagt die Verteilung ist *normiert*.

$$\int_{-\infty}^{+\infty} P(x) dx = 1 \quad (2)$$

2.3 Gaußverteilung

Die für die Fehlerrechnung wichtigste Verteilung ist die Gaußverteilung. Sie wird durch die Standardabweichung σ , die die Breite der Verteilung beschreibt, und den Erwartungswert μ , den Schwerpunkt der Verteilung, beschrieben.

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

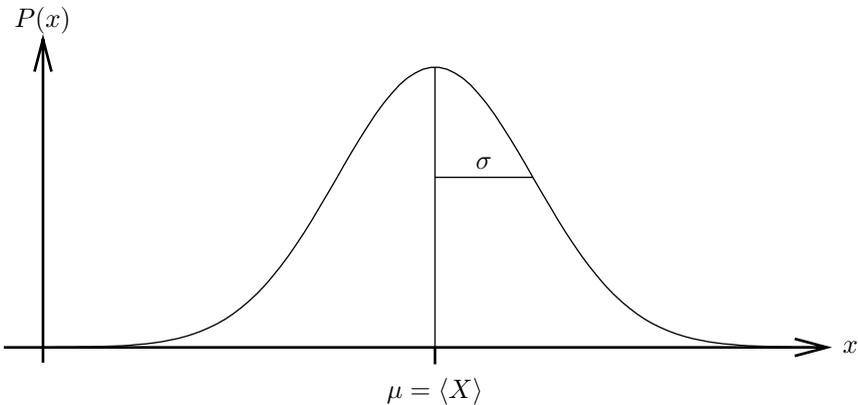


Abbildung 1: Gaußverteilung

Im folgenden werden wir stets davon ausgehen, dass unsere Messgrößen gaußverteilt sind. Warum das meist so ist, steht z.B. in [Bar, S. 49 ff].

2.4 Erwartungswert

Naiv möchten wir bei einer Messung den exakten “wahren” Wert x_w der Größe X erfahren, was auch abgesehen von der Frage ob es so etwas überhaupt gibt, auf Grund von unvermeidlichen Ungenauigkeiten des Messprozesses prinzipiell unmöglich ist. Als Ziel der Messung betrachten wir deshalb den Erwartungswert $\mu = \langle X \rangle$ der Messgröße. Im Fall einer diskreten Verteilung entspricht das Integral der mit den Eintrittswahrscheinlichkeiten gewichteten Summe über alle möglichen Werte. Wir werden der Anschaulichkeit halber in diesem Skript dennoch häufiger den Begriff des “wahren” Wertes x_w verwenden.

$$\mu = \langle X \rangle = \int_{-\infty}^{\infty} xP(x)dx \quad (4)$$

Der Erwartungswert kann auch für eine Funktion f der Zufallsgröße X definiert werden, die dann selbst als Zufallsgröße F aufgefasst werden kann. Für $\langle f(X) \rangle$ wird häufig kurz $\langle f \rangle$ geschrieben.

$$\langle F \rangle = \langle f(X) \rangle = \langle f \rangle = \int_{-\infty}^{\infty} f(x)xP(x)dx \quad (5)$$

Linearität des Erwartungswerts: Sind f und g Funktionen derselben Zufallsgröße X , dann gilt für Ihre Erwartungswerte:

$$\begin{aligned} \langle f + g \rangle &= \int_{-\infty}^{\infty} (f(x) + g(x))P(x)dx \\ &= \int_{-\infty}^{\infty} f(x)P(x)dx + \int_{-\infty}^{\infty} g(x)P(x)dx = \langle f \rangle + \langle g \rangle \end{aligned} \quad (6)$$

[Bar, S. 22]

2.5 Varianz, Standardabweichung, physikalische Fehler

In der Statistik ist die Varianz σ^2 das Maß für die Güte einer Messung der Größe X . Sie ist die mittlere quadratische Abweichung vom als bekannt vorausgesetzten Erwartungswert μ oder “wahren Wert” der Messung. Für kontinuierliche Verteilungen ist sie wie folgt definiert:

$$\sigma^2 := \langle (X - \mu)^2 \rangle = \langle (X - \langle X \rangle)^2 \rangle \stackrel{(5)}{=} \int_{-\infty}^{\infty} (x - \mu)^2 P(x)dx \quad (7)$$

$$\begin{aligned}
\sigma^2 &\stackrel{(7)}{=} \langle (X - \mu)^2 \rangle &= \int_{-\infty}^{\infty} (x - \mu)^2 P(x) dx \\
&= \langle X^2 - 2\mu X + \mu^2 \rangle &= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) P(x) dx \\
&= \langle X^2 \rangle - \langle 2\mu X \rangle + \langle \mu^2 \rangle &= \int_{-\infty}^{\infty} x^2 P(x) dx - \int_{-\infty}^{\infty} 2\mu x P(x) dx + \int_{-\infty}^{\infty} \mu^2 P(x) dx \\
&= \langle X^2 \rangle - 2\mu \langle X \rangle + \mu^2 &= \int_{-\infty}^{\infty} x^2 P(x) dx - 2\mu \underbrace{\int_{-\infty}^{\infty} x P(x) dx}_{\stackrel{(4)}{=} \mu} + \mu^2 \underbrace{\int_{-\infty}^{\infty} P(x) dx}_{\stackrel{(2)}{=} 1} \\
&= \langle X^2 \rangle - 2\mu^2 + \mu^2 &= \int_{-\infty}^{\infty} x^2 P(x) dx - 2\mu^2 + \mu^2 \\
&= \langle X^2 \rangle - \langle X \rangle^2 &= \int_{-\infty}^{\infty} x^2 P(x) dx - \mu^2
\end{aligned}$$

Abbildung 2: *Beweis der Varianzformel (8)

Für die Varianz gilt folgende vielverwendete Beziehung:

$$\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2 \quad (8)$$

In Worten: Die Varianz ist die Differenz aus dem Erwartungswert der Quadrate und dem Quadrat des Erwartungswerts.

Im Beweis (Abb. 2) wird die Linearität des Erwartungswerts verwendet.

Die *Standardabweichung* $\sigma := \sqrt{\sigma^2}$ hat dieselbe Dimension wie der Erwartungswert bzw. die Messgröße. Aus diesem Grund wird in der Physik als *Fehler* Δx meist die Standardabweichung oder ein Schätzwert dafür angegeben. Warum man in der Statistik lieber die Varianz verwendet, werden wir in Abschnitt 3.3 erfahren.

2.6 Gaußsches Fehlerfortpflanzungsgesetz

Ist die gesuchte Größe $z = z(x_1, \dots, x_n)$ eine Funktion mehrerer Messwerte (x_1, \dots, x_n) mit den Fehlern $(\sigma_1, \dots, \sigma_n)$ und sind die Messwerte gaußverteilt

und statistisch unabhängig gilt für den Fehler von z in erster Ordnung:

$$\sigma_z = \sqrt{\left(\frac{\partial z}{\partial x_1}\right)^2 \sigma_1^2 + \cdots + \left(\frac{\partial z}{\partial x_n}\right)^2 \sigma_n^2} = \sqrt{\sum_{i=1}^n \left(\frac{\partial z}{\partial x_i}\right)^2 \sigma_i^2} \quad (9)$$

[Bar, S. 55ff], [Bev, S. 41f]

3 Parameterschätzung

Wir werden am einfachen Beispiel des Mittelwerts neue Konzepte einführen und diese dann in den folgenden Kapiteln auf die Ausgleichsgerade übertragen.

3.1 Schätzfunktion

Stellen wir uns vor, wir haben dieselbe Größe X n -mal gemessen, dann können wir unsere Messreihe, die aus den Elementen x_i besteht, als Messwertvektor \vec{x} darstellen:

$$\vec{x} = (x_1, \dots, x_n) \quad (10)$$

Die Aufgabe besteht nun darin, aus den Messwerten $\vec{x} = (x_1, \dots, x_n)$ einen *Schätzwert* \hat{x} für den Erwartungswert $\langle X \rangle$ der Zufallsvariable X zu finden. Dieser Schätzwert muss offensichtlich aus den Messwerten $\vec{x} = (x_1, \dots, x_n)$ berechnet werden. Er ist also eine Funktion der Messwerte. Diese Funktion heißt *Schätzfunktion*.

$$\hat{x} = \hat{x}(\vec{x}) = \hat{x}(x_1, \dots, x_n) \quad (11)$$

Die Wahl dieser Funktion ist zunächst einmal willkürlich. Natürlich drängt sich in unserem Beispiel der gewöhnliche Mittelwert auf. In komplizierteren Fällen ist jedoch nicht immer klar, wie eine geeignete Schätzfunktion für das Problem aussieht.

Es gibt also für eine Aufgabenstellung prinzipiell beliebig viele verschiedene Schätzfunktionen.

Welche Eigenschaften charakterisieren nun eine *gute* Schätzfunktion? Wichtig sind die folgenden drei Eigenschaften, die jedoch auch bei häufig verwendeten Methoden nicht immer erfüllt werden können.

Konsistenz (asymptotische Erwartungstreue) Asymptotisch (für unendlich lange Messreihen) sollte der Schätzwert $\hat{x}(\vec{x})$ mit dem Erwartungswert des zu schätzenden Parameters übereinstimmen:

$$\lim_{n \rightarrow \infty} \left(\hat{x}(x_1, \dots, x_n) \right) = \langle X \rangle \quad (12)$$

Erwartungstreue Die Konsistenz macht nur eine Aussage für große Stichproben. Wir wünschen uns aber auch, dass im Mittel über viele kleine Stichproben der Mittelwert der Schätzwerte mit dem “wahren” Wert des Parameters übereinstimmt. Das heißt, der Erwartungswert der Schätzfunktion $\langle \hat{x} \rangle$ soll gleich dem Erwartungswert $\langle X \rangle$ des zu schätzenden Parameters sein.

$$\langle \hat{x} \rangle = \langle X \rangle \quad (13)$$

Abweichungen des *Erwartungswerts der Schätzfunktion* vom Erwartungswert des Parameters der Ausgangsverteilung nennt man *Bias*.

Wirksamkeit (efficiency) Eine Schätzfunktion heißt wirksamer als eine andere, wenn ihre Varianz kleiner als die der anderen ist. Auf Fragen der Wirksamkeit können wir in diesem Skript nicht weiter eingehen.

3.2 Maximum-Likelihood-Methode

3.2.1 am Beispiel einfacher Mittelwert

Eine systematische Methode eine Schätzfunktion zu finden, ist die *Maximum-Likelihood-Methode* (ML-Methode). Bei der Herleitung dieses Prinzips geht man zunächst davon aus, dass die Wahrscheinlichkeitsdichtefunktion $P(x)$, und damit auch der *unbekannte zu schätzende Erwartungswert* $\langle X \rangle = \mu$ *bekannt sei* und fragt nach der Wahrscheinlichkeit, mit der diese Ursprungsverteilung zu den beobachteten Messwerten $\vec{x} = (x_1, \dots, x_n)$ führt.

Die Wahrscheinlichkeit erst den Wert x_1 , dann x_2 usw. in dieser Reihenfolge zu messen, ist das Produkt der Einzelwahrscheinlichkeiten $P(x)$. Auch für kontinuierliche Wahrscheinlichkeitsverteilungen gilt diese Regel. Als Ergebnis erhält man eine Wahrscheinlichkeitsdichtefunktion $L(\vec{x})$, die die Wahrscheinlichkeit, den beobachteten Satz von Messwerten zu erhalten, beschreibt; die *Likelihood-Funktion*. Gehen wir davon aus, dass unsere Messwerte einer Gaußverteilung $P(x)$ folgen, erhält man:

$$\begin{aligned}
L(\vec{x}) &= L(x_1, \dots, x_n) = P(x_1)P(x_2) \cdots P(x_n) \\
&= \prod_i P(x_i) \\
&= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}
\end{aligned} \tag{14}$$

Leider kennen wir den Erwartungswert $\langle X \rangle = \mu$ nicht. (Sonst bräuchten wir X nicht zu messen!) Wir wollen deshalb jetzt μ so wählen, dass die Wahrscheinlichkeit die beobachteten Messwerte zu erhalten maximal wird. (Maximum-Likelihood) *Dieser Wert von μ ist unser Schätzwert \hat{x} für X !*

Suchen wir also das Maximum von L . Notwendige Bedingung für ein Maximum ist eine Nullstelle der ersten Ableitung. Wenn L ein Maximum aufweist, wird auch $\ln(L)$ maximal. Dieses Maximum ist leichter zu bestimmen, da das Produkt von Exponentialfunktionen so zu einer Summe wird, die sich leichter ableiten lässt.

$$\begin{aligned}
\ln(L(\vec{x})) &= -n \ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2 \\
0 = \frac{\partial \ln L}{\partial \mu} \Big|_{\mu=\hat{x}} &= \frac{1}{\sigma^2} \sum_i (x_i - \hat{x}) = \frac{1}{\sigma^2} \left(\left(\sum_i x_i \right) - n\hat{x} \right) \\
\hat{x} &= \frac{1}{n} \sum_{i=1}^n x_i (=:\bar{x})
\end{aligned} \tag{15}$$

Es ist nicht überraschend, dass die ML-Schätzfunktion \hat{x} in unserem Beispiel das arithmetische Mittel \bar{x} ist. Man müsste jetzt eigentlich zeigen, dass diese Schätzfunktion konsistent, erwartungstreu und wirksam ist. Darauf können wir jedoch an dieser Stelle leider nicht weiter eingehen.

Kommen wir nun zu einer Fragestellung mit etwas weniger offensichtlichem Ergebnis:

3.2.2 Der gewichtete Mittelwert

Angenommen eine Größe X wird mit *verschiedenen Methoden, mit unterschiedlichen Fehlern* σ_i gemessen. Wir werden naiv wieder eine Art Mittelwert bilden wollen. Aber wie sind die einzelnen Werte zu gewichten? Wir setzen wieder gaußverteilte Messwerte voraus. In der Sprache der Statistik entstammt

jetzt jeder Messwert x_i einer anderen Verteilung P_i – jede mit eigenem σ_i aber gemeinsamem μ . Wenden wir nun die oben eingeführte ML-Methode an, um einen Schätzwert für den Erwartungswert von X zu erhalten:

$$\begin{aligned}
 L(\vec{x}) &= L(x_1, \dots, x_n) = \prod_i P_i(x_i) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma_i^2}} \\
 \ln(L(x_1, \dots, x_n)) &= \sum_i -\ln\left(\sqrt{2\pi\sigma_i^2}\right) - \sum_i \frac{(x_i - \mu)^2}{2\sigma_i^2} \\
 0 = \frac{\partial \ln L}{\partial \mu} \Big|_{\mu=\hat{x}} &= \sum_i \left(\frac{x_i - \hat{x}}{\sigma_i^2} \right) = \sum_i \frac{x_i}{\sigma_i^2} - \hat{x} \sum_i \frac{1}{\sigma_i^2} \\
 \hat{x} &= \frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} (=:\bar{x})
 \end{aligned} \tag{16}$$

Bei der Mittelwertbildung sind also statistisch unabhängige, gaußverteilte Messwerte mit $w_i = \frac{1}{\sigma_i^2}$ zu gewichten. Es lässt sich zeigen, dass diese ML-Schätzfunktion die wirksamste Schätzfunktion für diesen Fall ist.

3.2.3 Schätzung des Fehlers einer Messung aus der Streuung

Gehen wir noch einen Schritt weiter, nehmen wir an, wir haben eine Messung mehrfach mit *derselben Methode* wiederholt und kennen den Fehler der Messwerte nicht, dann sollte es möglich sein, aus der Abweichung der Messwerte vom Mittelwert *den Fehler einer einzelnen Messung* σ zu schätzen.

Wird eine Messung mehrfach mit derselben Methode wiederholt, dann haben alle Messwerte x_i denselben Fehler σ , denn sie stammen aus derselben Ausgangsverteilung $P(x)$. Diese soll auch in diesem Abschnitt gaußförmig sein.

Die Aufgabe besteht nun darin, die beiden Parameter σ^2 und μ gleichzeitig zu schätzen. Für den Schätzwert von μ bei konstantem σ^2 haben wir bereits in (15) das arithmetische Mittel gefunden:

$$\hat{x} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \tag{17}$$

Der Schätzwert für μ ist also von den Fehlern unabhängig, solange sie nur für alle Messwerte gleich groß sind.

Um auch eine Schätzung für σ zu erhalten, wenden wir wieder das ML-Verfahren an.

$$P(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \quad (18)$$

$$L(\vec{x}) = \prod_i \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \quad (19)$$

$$\ln L = -n \ln \left(\sqrt{2\pi\sigma^2} \right) - \frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2 \quad (20)$$

Es sind zwei Parameter zu bestimmen; wir erhalten also ein System aus zwei Gleichungen. Diese Gleichungen heißen *Normalgleichungen*.

$$0 = \frac{\partial \ln L}{\partial \mu} \Bigg|_{\substack{\mu = \hat{x} \\ \sigma^2 = \widehat{\sigma^2}}} \Leftrightarrow 0 = \sum_i (x_i - \hat{x}) \Leftrightarrow \hat{x} = \frac{1}{n} \sum_i x_i =: \bar{x} \quad (21a)$$

$$0 = \frac{\partial \ln L}{\partial (\sigma^2)} \Bigg|_{\substack{\mu = \hat{x} \\ \sigma^2 = \widehat{\sigma^2}}} = -\frac{n}{2\widehat{\sigma^2}} + \frac{1}{2(\widehat{\sigma^2})^2} \sum_i (x_i - \hat{x})^2 \quad (21b)$$

Die erste Normalgleichung (21a) ist unabhängig von σ^2 und liefert das bekannte Ergebnis. Die zweite Normalgleichung (21b) liefert nach einsetzen von $\hat{x} = \bar{x}$ aus (21a) eine Schätzfunktion für σ^2 :

$$\widehat{\sigma^2} = \frac{1}{n} \sum_i (x_i - \bar{x})^2 \quad (22)$$

Über ihre Konsistenz und Erwartungstreue haben wir keine Aussage gemacht!

Die Wahl von ML zur Gewinnung einer Schätzfunktion ist willkürlich. Wir hätten genausogut direkt die naheliegende Formel (22) wählen können. Prinzipiell liefern verschiedene Methoden jedoch unterschiedliche Schätzfunktionen. Welche davon gut sind, muss im Einzelfall untersucht werden. *22 ist nicht gut!* Wir werden diese Funktion deshalb im nächsten Abschnitt untersuchen und eine bessere Schätzfunktion (25) für unser Problem finden.

3.3 Das Problem der Erwartungstreue Bessels Korrektur am Beispiel von 3.2.3

Die in (22) hergeleitete Schätzfunktion für die Varianz lässt sich analog zu Abb. 2 umformen:

$$\begin{aligned} \widehat{(\sigma^2)} &= \frac{1}{n} \sum_i (x_i - \bar{x}(x_1, \dots, x_n))^2 = \frac{1}{n} \sum_i (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \frac{1}{n} \left(\sum_i x_i^2 \right) - 2\bar{x} \underbrace{\frac{1}{n} \sum_i x_i}_{=\bar{x}} + \frac{1}{n} \sum_i \bar{x}^2 = \frac{1}{n} \sum_i (x_i^2 - \bar{x}^2) \end{aligned} \quad (23)$$

In dieser Form lässt sich zeigen, dass diese Schätzfunktion nicht erwartungstreu ist! Wir betrachten den Erwartungswert von $\widehat{(\sigma^2)}$. Dazu benötigen wir den Erwartungswert des Mittelwerts $\langle \bar{x} \rangle$.

Wir fassen den Mittelwert $\bar{x}(x_1, \dots, x_n)$ als Funktion der Zufallsgrößen X_i auf.

Da jeder Messwert x_i eine Realisation derselben Zufallsgröße X ist, hängt der Mittelwert $\bar{x}(X_1, \dots, X_n)$, nur von der einen Zufallsgröße $X = X_1 = \dots = X_n$ ab: $\bar{x}(X, \dots, X)$.

\bar{x} ist eine erwartungstreu Schätzfunktion für X , deshalb gilt: $\langle X \rangle = \langle \bar{x} \rangle$; jedoch gilt dies nicht für die Erwartungswerte der Quadrate: $\langle X^2 \rangle \neq \langle \bar{x}^2 \rangle$!

$$\begin{aligned} \widehat{(\sigma^2)} &= \left\langle \frac{1}{n} \left(\sum_i X^2 \right) - \bar{X}^2 \right\rangle = \langle X^2 \rangle - \langle \bar{X}^2 \rangle \\ &= \langle X^2 \rangle - \langle \bar{X}^2 \rangle - \underbrace{\left(\langle X \rangle^2 - \langle \bar{X} \rangle^2 \right)}_{=0} = \left(\langle X^2 \rangle - \langle X \rangle^2 \right) - \left(\langle \bar{X}^2 \rangle - \langle \bar{X} \rangle^2 \right) \\ &\stackrel{(8)}{=} \sigma_X^2 - \sigma_{\bar{X}}^2 \stackrel{(26)}{=} \sigma_X^2 - \frac{1}{n} \sigma_X^2 \\ &= \left(\frac{n-1}{n} \right) \sigma_X^2 \neq \sigma_X^2 \end{aligned} \quad (24)$$

Bessels Korrektur Offensichtlich ist die Schätzfunktion tatsächlich nicht erwartungstreu, allerdings konsistent, denn für große n geht der Faktor $\frac{n-1}{n}$ gegen 1. Ebenso offensichtlich ist, wie man aus dieser Funktion eine erwartungstreu Schätzfunktion erhält: Man multipliziert einfach mit dem Keh-

wert des störenden Faktors. Diese Korrektur heißt *Bessels Korrektur*, und die korrigierte Schätzfunktion wird üblicherweise s^2 genannt¹:

$$s^2 = \frac{n}{n-1} \widehat{\sigma^2} = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 \quad (25)$$

Dies ist also die Antwort auf die in (3.2.3) gestellte Frage, wie der Fehler einer Messung aus der Streuung geschätzt werden kann!

Freiheitsgrade Wie kann man dieses Ergebnis anschaulich deuten? Um den Fehler eines Messwertes aus der Streuung zu bestimmen, benötigt man zusätzliche Messungen. Man kann sinnvoll aus einer Messung einen Mittelwert (als Schätzwert für den Erwartungswert der Messung) bilden. Aus einem Punkt kann man jedoch keinen Schätzwert für die Streuung bestimmen. (22) liefert in diesem Fall stets den sinnlosen Wert 0. (25) hingegen ist sinnvoller Weise für $n = 1$ nicht definiert.

Man kann diese Überlegung verallgemeinern: Eine Gerade geht immer durch zwei Punkte und wird durch zwei Parameter (Steigung b und Achsenabschnitt a) vollständig bestimmt. Wurden nur zwei Punkte gemessen, ist eine Schätzung der Streuung aus der Abweichung von der Geraden nicht möglich.

Allgemein wird durch die Bestimmung eines Parameters ein Messwert verbraucht. Werden also m Parameter geschätzt, bleiben für die Schätzung der Fehler noch $(n - m)$ Werte übrig. Dies ist die Zahl der *Freiheitsgrade*.

Es gibt eine weitere Überlegung, die verständlich macht, warum Bessels Korrekturfaktor > 1 ist. s^2 ist ein Schätzwert für die Varianz σ^2 , den mittleren quadratischen Abstand eines Messwerts vom Erwartungswert. Im allgemeinen werden die Messwerte (x_1, \dots, x_n) näher an *ihrem* Mittelwert $\bar{x}(x_1, \dots, x_n)$, als am Erwartungswert der unterliegenden Verteilung liegen. (22) liefert also einen zu kleinen Wert.

Standardabweichung Für die Standardabweichung σ gibt es keine solche erwartungstreue Schätzfunktion! Glücklicherweise tritt in praktisch allen Rechnungen nur das Quadrat σ^2 auf, für das die erwartungstreue Schätzfunktion s^2 verwendet werden kann. Lediglich wenn $\Delta x = \sqrt{s^2}$ als Schätzung für den Fehler angegeben wird, tritt das Problem auf. Wenn man mit diesem Wert weiter rechnet, wird er aber wieder quadriert und die Welt ist in Ordnung. Man gibt nicht die Varianz direkt an, damit Wert und Fehler dieselbe Dimension haben.

¹Auf Taschenrechnern wird $\sqrt{s^2}$ häufig mit σ_{n-1} bezeichnet.

4 Der Fehler der geschätzten Parameter am Beispiel Mittelwert

Bisher haben wir nur Messwerte und ihre Fehler betrachtet und daraus möglichst gute Schätzwerte für den “wahren Wert” unserer Messgröße ermittelt. In diesem Abschnitt untersuchen wir die Frage, wie man aus den Fehlern der einzelnen Messwerte einen Fehler für diesen *Schätzwert* erhalten kann.

4.1 Bekannte Messfehler

4.1.1 Fehler gleich groß

Beispiel: Ein Praktikumsbetreuer klatscht zweimal vor der von ihm betreuten Gruppe von 6 Studenten, ausgestattet mit je einer Stoppuhr mit $1/100$ s Auflösung, für diese nicht sichtbar mit einer Knallpatsche. Die Gruppe hat die Aufgabe, die Zeit zwischen den Ereignissen möglichst genau zu bestimmen. Der Betreuer gibt an, die Reaktionszeitdifferenz zwischen Start und Stopp solle mit $\Delta t = 0.1$ s als Fehler angenommen werden.

Wurde dieselbe Größe, wie im Beispiel mehrfach unabhängig mit derselben Methode gemessen und ist die Varianz für alle Werte gleich und bekannt, geben wir als Ergebnis der Messung den ungewichteten Mittelwert nach (15) an. Wir erwarten, dass der Fehler dieses Mittelwertes kleiner ist, als der eines einzelnen Wertes. Wie groß der Fehler ist, erfahren wir, in dem wir das Gaußsche Fehlerfortpflanzungsgesetz (9) auf die Mittelwertbildung (15) anwenden:

$$\bar{x} = \frac{1}{n} \sum_i x_i \Rightarrow \sigma_{\bar{x}} = \frac{1}{\sqrt{n}} \sigma \quad (26)$$

Das Ergebnis ist auch als \sqrt{n} -Regel bekannt. Wie sind diese Fehler nun zu interpretieren? Sicher handelt es sich nicht um die Standardabweichung σ der Ursprungsverteilung. Aber in der Theorie der Fehlerrechnung interpretieren wir die Fehler als Standardabweichungen von Zufallsgrößen. Auch der oben berechnete Fehler des Mittelwertes $\sigma_{\bar{x}}$ ist eine Standardabweichung, und zwar die einer Verteilungsfunktion der Mittelwerte.

Wiederholen wir das Experiment, erhalten wir viele solche Mittelwerte die aus einer Gaußverteilung mit dem Erwartungswert $\mu_x = \mu_{\bar{x}}$ aber mit der Standardabweichung $\sigma_{\bar{x}}$ stammen. (Abb. 3)

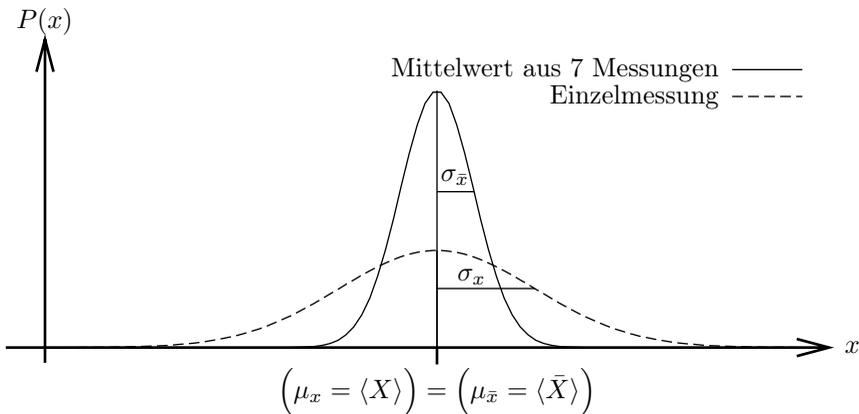


Abbildung 3: Wahrscheinlichkeitsdichtefunktionen für Einzelmessung und Mittelwert

4.1.2 Fehler unterschiedlich groß

Beispiel: Die Studenten 1,2,3 und 4 sind ausgeschlafen, 5 und 6 sind müde. Der Betreuer gibt an, für unausgeschlafene Praktikanten sei ein Fehler von $\Delta t = 0.2\text{s}$ zu verwenden. Wir setzen also $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.1\text{s}$ und $\sigma_5 = \sigma_6 = 0.2\text{s}$.

Wurde dieselbe Größe mit unterschiedlichen Methoden i mit unterschiedlichen bekannten Fehlern σ_i gemessen, verwenden wir den gewichteten Mittelwert nach (16) und erhalten durch Anwendung des Fehlerfortpflanzungsgesetzes:

$$\bar{x} = \frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \quad \Rightarrow \quad \sigma_{\bar{x}} = \frac{1}{\sqrt{\sum_i \frac{1}{\sigma_i^2}}} \quad (27)$$

4.2 Unbekannte Fehler - alle Fehler gleich groß

Bisher haben wir stets Situationen betrachtet, in denen der Fehler eines Messwerts unabhängig von der Messung bekannt war. (z.B.: vom Betreuer, aus einer Bedienungsanleitung, durch eine Schätzung des Ablesefehlers, ...). Der Fehler der Messung war also prinzipiell *a priori*, also bevor die Messung durchgeführt wurde, bekannt.

Nun gehen wir davon aus, dass wir aus *derselben* Messreihe sowohl unser Messergebnis (Einen Schätzwert für den Erwartungswert) bestimmen wollen, als auch eine Schätzwert für den Fehler dieses Ergebnisses.

Gehen wir wieder davon aus, dass alle Messwerte denselben Fehler haben, d.h., dass sie aus derselben Wahrscheinlichkeitsverteilung mit der *unbekannten* Varianz σ^2 stammen, dann ist s^2 (25) ein Schätzwert für die Varianz eines *einzelnen* Wertes x_i . Wir können diesen Schätzwert nun in Gleichung (26) für den Fehler des Mittelwertes einsetzen und erhalten so:

$$\sigma_{\bar{X}} \approx \sqrt{\frac{s^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}} \quad (28)$$

4.3 *Bekannte Fehler / aus der Streuung geschätzt Fehler

Betrachtet man eine gaußverteilte Messgröße x mit bekannter Standardabweichung σ . x_w sei ihrer wahrer Wert, x_i der Messwert. Man kann diese Angabe auf zwei Arten interpretieren.

Zum einen ist diese Standardabweichung definitionsgemäß die Wurzel aus dem mittleren quadratischen Fehler.

Zum anderen kann man die Standardabweichung als Konfidenzintervall auffassen. Denn auf Grund der zugrundeliegenden Gaußverteilung, liegt der Messwert x_i mit einer Wahrscheinlichkeit von 68% (1σ -Wahrscheinlichkeit) in dem Intervall $]x_w - \sigma; x_w + \sigma[$. Dann gilt natürlich auch: „Der wahre Wert x_w liegt mit der selben Wahrscheinlichkeit im Intervall $]x_i - \sigma; x_i + \sigma[$ um den Messwert x_i . Genau so sind Konfidenzintervalle definiert. Dabei kann die Wahrscheinlichkeit, die jetzt statistische Sicherheit oder Vertrauensniveau heißt, beliebig vorgegeben werden. Die Intervallgröße ändert sich dann entsprechend.

Wenn die Standardabweichung nun nicht a priori bekannt ist, sondern geschätzt wird, erhält man zwar einen Schätzwert für die Wurzel aus dem mittleren quadratischen Fehler, die zweite Interpretationsmöglichkeit als Konfidenzintervall geht allerdings verloren, weil die Grenzen des Intervalls selbst mit einer Unsicherheit behaftet sind.

Dennoch ist es auch in diesem Fall möglich, wieder Konfidenzintervalle anzugeben. Diese sind allerdings für kleine n wesentlich größer, als durch die Schätzwerte der Standardabweichung suggeriert wird. Lediglich für große n stimmen diese Konfidenzintervalle mit den geschätzten Standardabweichungen überein.

Wir können hier nicht auf die Berechnung der Konfidenzintervalle aus den Schätzwerten der Standardabweichung eingehen. Mehr dazu findet man in [Mand, S. 114]. [Bar, S. 134ff] Stichwort: Students t Verteilung.

5 Lineare Regression

In diesem Abschnitt kommen wir endlich zu der Frage, wie man die Lage einer Ausgleichsgerade durch die Messpunkte bestimmen kann. Um den Leser nicht zu ermüden, betrachten wir von Anfang an den Fall, dass die einzelnen Messwerte unterschiedliche Fehler haben.

Wir haben Paare von Punkten (x_i, y_i) gemessen. Wir nehmen an, dass der Zusammenhang zwischen x und y durch eine Gerade, d.h., durch eine Funktion vom Typ

$$Y(X) = A + BX \quad (29)$$

beschrieben werden kann. Diese Funktion heißt *Modellfunktion*. Wir können unsere Messdaten durch zwei Vektoren $\vec{x} = (x_1, \dots, x_n)$ und $\vec{y} = (y_1, \dots, y_n)$ darstellen. A und B sind Zufallsgrößen und unsere Aufgabe besteht darin, möglichst gute Schätzwerte \hat{a} und \hat{b} für Ihre Erwartungswerte $\langle A \rangle$ und $\langle B \rangle$ (anschaulich ihre wahren Werte a_w, b_w) zu bestimmen. Wir machen folgende Voraussetzungen:

- die x_i seien vom Experimentator frei wählbar und ihre Fehler seien gegenüber denen der y_i vernachlässigbar klein. In diesem Fall besteht also kein Unterschied zwischen den Messwerten x_i und ihren Erwartungswerten $\langle X_i \rangle$ bzw. ihren wahren Werten x_{wi} .
- Jeder einzelne Messwert y_i stammt aus einer *eigenen* Gaußverteilung mit dem Erwartungswert $\mu_i = \langle Y_i \rangle$ und der Standardabweichung σ_i . Dies wird in Abb. 4 verdeutlicht.

In Abb. 5 ist der Unterschied zwischen Mess-, Schätz-, und Erwartungswert dargestellt.

5.1 Schätzung der Parameter A und B der Geraden

Wenden wir nun die Maximum-Likelihoodfunktion auf dieses Problem an: Für jede Stelle x_i gibt es nun einen "wahren Wert" $y_{iw} = y_w(x_i) = \langle Y(x_i) \rangle = \langle Y_i \rangle$. Die Wahrscheinlichkeit an der Stelle x_i die Messung y_i zu machen, wird durch die folgende Gaußfunktion beschrieben:

$$P_i(y_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i - \langle Y_i \rangle)^2}{2\sigma_i^2}\right) \quad (30)$$

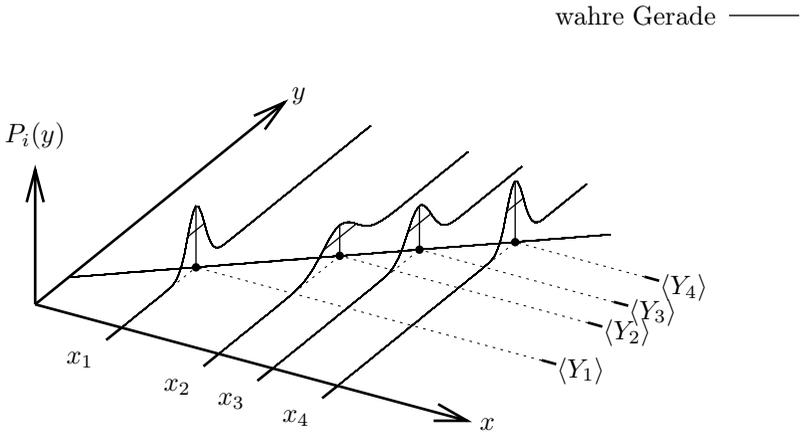


Abbildung 4: Wahrscheinlichkeitsdichte bei der linearen Regression

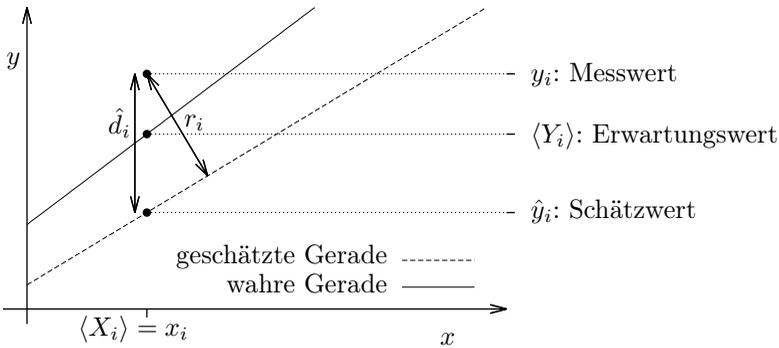


Abbildung 5: Unterschied zwischen Mess-, Schätz-, und Erwartungswert bei der linearen Regression

Die Likelihoodfunktion, also die Wahrscheinlichkeitsdichte, wenn man an den Stellen \vec{x} misst die Werte \vec{y} zu beobachten, ergibt sich wieder als Produkt der Einzelwahrscheinlichkeiten:

$$L(\vec{y}) = L(y_1, \dots, y_n) = \prod_i P_i = \prod_i \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \right) \exp \left(-\frac{(y_i - \langle Y_i \rangle)^2}{2\sigma_i^2} \right) \quad (31)$$

Die $\langle Y_i \rangle$ können wir nun mit Hilfe des Modells (29) durch die Erwartungswerte für die Parameter $\langle A \rangle$ und $\langle B \rangle$ ausdrücken:

$$\langle Y_i \rangle = \langle A \rangle + \langle B \rangle x_i \quad (32)$$

Setzen wir dies in (31) ein, erhalten wir:

$$\begin{aligned} L(\vec{y}) &= \prod_i P_i = \prod_i \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \right) \exp \left(-\frac{(y_i - \langle A \rangle - \langle B \rangle x_i)^2}{2\sigma_i^2} \right) \\ &= \left(\prod_i \left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \right) \right) \exp \left(\sum_i \left(-\frac{(y_i - \langle A \rangle - \langle B \rangle x_i)^2}{2\sigma_i^2} \right) \right) \end{aligned} \quad (33)$$

Um die Werte für A und B zu finden, die unsere Messwerte am wahrscheinlichsten machen, müssen wir nun die Likelihoodfunktion maximieren. Mit denselben Überlegungen wie in 3.2 bestimmen wir deshalb die Nullstellen der Ableitung von $\ln L$.

$$0 = \left. \frac{\partial \ln L}{\partial \langle A \rangle} \right|_{\substack{\langle A \rangle = \hat{a} \\ \langle B \rangle = \hat{b}}} \quad \wedge \quad 0 = \left. \frac{\partial \ln L}{\partial \langle B \rangle} \right|_{\substack{\langle A \rangle = \hat{a} \\ \langle B \rangle = \hat{b}}} \quad (34)$$

Da (33) verhältnismäßig lang ist, ist es sicherlich gut, sich auf den Term zu beschränken, der für die Ableitung von Bedeutung ist. Das ist allein die Summe im Exponenten. Diese Summe nennen wir nun χ^2 (chiquadrat).² Um L zu maximieren, müssen wir diese Summe minimieren.

$$\chi^2 := \sum_i \left(\frac{y_i - \hat{y}(x_i)}{\sigma_i} \right)^2 = \sum_i \left(\frac{1}{\sigma_i} (y_i - \hat{a} - \hat{b}x_i) \right)^2 \quad (35)$$

Wir wollen also die beste Ausgleichsgerade dadurch finden, dass wir Werte für die Parameter \hat{a} und \hat{b} finden, die die gewichtete Summe der Quadrate der

²Diese Größe wird häufig auch *WSSR* („**W**eighted **S**um of **S**quared **R**esiduals“) genannt. Eine analoge Größe findet sich auch schon bei der Herleitung des Mittelwertes in (14).

Abweichungen χ^2 minimieren. Wir wollen also die Ausgleichsgerade finden, die die kleinste Summe der Quadrate hervorruft. Daher der Name *Methode der kleinsten Quadrate* oder auf Englisch *least-squares fit* für das Verfahren.

Minimierung von χ^2 : Um die Werte der Parameter \hat{a} und \hat{b} , die χ^2 minimieren, zu finden, bilden wir die partiellen Ableitungen von χ^2 nach den Parametern und setzen diese gleich Null. Die so erhaltenen Gleichungen heißen *Normalgleichungen*.

$$\frac{\partial}{\partial \hat{a}} \chi^2 = \frac{\partial}{\partial \hat{a}} \sum_i \left(\frac{1}{\sigma_i^2} (y_i - \hat{a} - \hat{b}x_i)^2 \right) = -2 \sum_i \left(\frac{1}{\sigma_i^2} (y_i - \hat{a} - \hat{b}x_i) \right) = 0 \quad (36a)$$

$$\frac{\partial}{\partial \hat{b}} \chi^2 = \frac{\partial}{\partial \hat{b}} \sum_i \left(\frac{1}{\sigma_i^2} (y_i - \hat{a} - \hat{b}x_i)^2 \right) = -2 \sum_i \left(\frac{x_i}{\sigma_i^2} (y_i - \hat{a} - \hat{b}x_i) \right) = 0 \quad (36b)$$

Die Normalgleichungen bilden ein lineares Gleichungssystem in zwei Variablen \hat{a} und \hat{b} , das leicht gelöst werden kann:

$$\left(\sum_i \frac{1}{\sigma_i^2} \right) \hat{a} + \left(\sum_i \frac{x_i}{\sigma_i^2} \right) \hat{b} = \left(\sum_i \frac{y_i}{\sigma_i^2} \right) \quad (37a)$$

$$\left(\sum_i \frac{x_i}{\sigma_i^2} \right) \hat{a} + \left(\sum_i \frac{x_i^2}{\sigma_i^2} \right) \hat{b} = \left(\sum_i \frac{x_i y_i}{\sigma_i^2} \right) \quad (37b)$$

Um die Lösung übersichtlicher darstellen zu können, führen wir als Abkürzung die Determinante S der Koeffizientenmatrix ein.

$$S = \sum_i \frac{1}{\sigma_i^2} \sum_i \frac{x_i^2}{\sigma_i^2} - \left(\sum_i \frac{x_i}{\sigma_i^2} \right)^2$$

$$\hat{a} = \frac{1}{S} \left(\sum_i \frac{x_i^2}{\sigma_i^2} \sum_i \frac{y_i}{\sigma_i^2} - \sum_i \frac{x_i}{\sigma_i^2} \sum_i \frac{x_i y_i}{\sigma_i^2} \right) \quad (38a)$$

$$\hat{b} = \frac{1}{S} \left(\sum_i \frac{1}{\sigma_i^2} \sum_i \frac{x_i y_i}{\sigma_i^2} - \sum_i \frac{x_i}{\sigma_i^2} \sum_i \frac{y_i}{\sigma_i^2} \right) \quad (38b)$$

Dies sind also die ML-Schätzfunktionen $\hat{a}(\vec{x}, \vec{y})$ und $\hat{b}(\vec{x}, \vec{y})$.

5.2 Berechnung der Fehler der Parameter

5.2.1 a priori-Fehler

Gehen wir wie in Abschnitt 4.1.1 davon aus, dass die Fehler σ_i der Messwerte y_i unabhängig von der Messung bekannt sind, (aus der Bedienungsanleitung etc.) dann erhalten wir mit dem Gaußschen Fehlerfortpflanzungsgesetz die Fehler der Parameter \hat{a} und \hat{b} .

$$\begin{aligned}
 \sigma_a^2 &= \sum \left(\frac{\partial \hat{a}}{\partial y_i} \right)^2 \sigma_i^2 \\
 &= \sum_{i=1}^N \frac{\sigma_i^2}{S^2} \left[\frac{1}{\sigma_i^4} \left(\sum_j \frac{x_j^2}{\sigma_j^2} \right)^2 - \frac{2x_i}{\sigma_i^4} \sum_j \frac{x_j^2}{\sigma_j^2} \sum_j \frac{x_j}{\sigma_j^2} + \frac{x_i^2}{\sigma_i^4} \left(\sum_j \frac{x_j}{\sigma_j^2} \right)^2 \right] \\
 &= \frac{1}{S^2} \left(\sum_j \frac{x_j^2}{\sigma_j^2} \right) \left[\sum \frac{1}{\sigma_i^2} \sum \frac{x_j^2}{\sigma_j^2} - \left(\sum \frac{x_j}{\sigma_j^2} \right)^2 \right] \\
 &= \frac{1}{S} \sum_j \frac{x_j^2}{\sigma_j^2} \tag{39a}
 \end{aligned}$$

$$\begin{aligned}
 \sigma_b^2 &= \sum \left(\frac{\partial \hat{b}}{\partial y_i} \right)^2 \sigma_i^2 \\
 &= \sum_{i=1}^N \frac{\sigma_i^2}{S^2} \left[\frac{x_i^2}{\sigma_i^4} \left(\sum_j \frac{1}{\sigma_j^2} \right)^2 - \frac{2x_i}{\sigma_i^4} \sum_j \frac{1}{\sigma_j^2} \sum_j \frac{x_j}{\sigma_j^2} + \frac{1}{\sigma_i^4} \left(\sum_j \frac{x_j}{\sigma_j^2} \right)^2 \right] \\
 &= \frac{1}{S^2} \left(\sum_j \frac{x_j^2}{\sigma_j^2} \right) \left[\sum \frac{1}{\sigma_i^2} \sum \frac{1}{\sigma_j^2} - \left(\sum \frac{x_j}{\sigma_j^2} \right)^2 \right] \\
 &= \frac{1}{S} \sum_j \frac{1}{\sigma_j^2} \tag{39b}
 \end{aligned}$$

Für die Fehler $\Delta \hat{a}$ und $\Delta \hat{b}$ der Parameter \hat{a} und \hat{b} gilt nun

$$\Delta \hat{a} = \sqrt{\sigma_a^2} \quad \text{und} \quad \Delta \hat{b} = \sqrt{\sigma_b^2} \tag{40}$$

Im Spezialfall gleicher Fehler ($\sigma_i = \sigma$) in den gemessenen Werten y_i , vereinfachen sich die Ausdrücke (38) und (39) zu

$$S' = S\sigma^4 = n \sum x_i^2 - \left(\sum x_i\right)^2$$

$$\hat{a} = \frac{1}{S'} \left(\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i \right) \quad \sigma_a^2 = \frac{\sigma^2}{S'} \sum x_i^2 \quad (41a)$$

$$\hat{b} = \frac{1}{S'} \left(\sum_i x_i y_i - \sum_i y_i \right) \quad \sigma_b^2 = N \frac{\sigma^2}{S'} \quad (41b)$$

5.2.2 Fehler aus der Streuung

In diesem Abschnitt wollen wir analog zu 4.2 aus *derselben Messreihe* sowohl die Parameter A und B , als auch ihre Fehler schätzen.

Absolute Fehler gleich

Die y -Werte sollen alle denselben unbekanntem Fehler $\sigma_i = \sigma$ aufweisen. Wir möchten nun diesen gemeinsamen Fehler schätzen, um daraus Fehler für Steigung und Achsabschnitt berechnen zu können. In (41) haben wir gezeigt, dass für konstante Fehler die Parameter A und B unabhängig vom Fehler geschätzt werden können. Dies ermöglicht es, an jeder Stelle x_i einen zugehörigen Funktionswert \hat{y}_i zu schätzen. Die Abstände von Schätzwert \hat{y} und Messwert y_i heißen *Residuen* \hat{d}_i . Das *d* rührt daher, dass es sich nicht um die Abweichung von den wahren Werten y_{iw} , sondern um Schätzwerte für diesen Abstand handelt. (Abb. 5)

$$\hat{d}_i = y_i - \hat{y}_i = y_i - \hat{a} - \hat{b}x \quad (42)$$

Als Schätzwert für die Varianz σ^2 der y -Werte wählen wir nun in naheliegender Weise:

$$\widehat{\sigma^2} = \frac{1}{n} \sum \hat{d}_i^2 \quad (43)$$

Die Anwendung des ML-Prinzips hätte auch in diesem Fall wieder auf die Methode der kleinsten Quadrate und damit auf das obige Ergebnis geführt.

Der so gewonnene Schätzwert ist nur konsistent, aber nicht erwartungstreu, wir wenden daher BESSELS Korrektur analog zu 3.3 an und definiert die erwartungstreu Schätzfunktion s^2 für die Varianz eines einzelnen Messpunktes y_i . Da 2 Parameter a und b geschätzt wurden, ist die Zahl der Freiheitsgrade

hier $(n - 2)$.

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x)^2 \quad (44)$$

Aus der so geschätzten Varianz der Messwerte kann man in der Näherung $\sigma^2 \approx s^2$ die Fehler von \hat{a} und \hat{b} nach (39) schätzen. Bei dieser Schätzung tritt wieder das Problem auf, dass der Fehler mit dem s behaftet ist in diese Überlegung nicht eingeht. Deshalb gilt für die so geschätzten $\Delta\hat{a}$ und $\Delta\hat{b}$ das in 4.3 gesagte. Bei wiederholter Durchführung der Messreihe fluktuieren die Varianzen s^2 bei kleinem n (Zahl der Punkte einer Messreihe) sehr stark.

Absolute Fehler verschieden – Verhältnis bekannt (gewichtete Streuung)

Beispiel: Exponentialfunktion Wir untersuchen eine Messreihe bei der der Zusammenhang zwischen x und z durch eine Exponentialfunktion $Z(x) = \exp(A + Bx)$ beschrieben wird. Wir wissen, dass alle Punkte z_i denselben unbekanntem Fehler σ aufweisen. Wir möchten diesen Fehler und damit auch die Fehler der Parameter a und b aus derselben Messreihe schätzen.

Um die Daten auswerten zu können, müssen wir die z_i logarithmieren und den unbekanntem Fehler der z_i nach dem Fehlerfortpflanzungsgesetz umrechnen. Wir erhalten So einen neuen Datensatz mit $y_i = \ln(z_i)$ und $\sigma_i \stackrel{(9)}{=} \frac{1}{y_i} \sigma$. Die Fehler der y_i können also in der Form $\sigma_i = f_i \sigma$ dargestellt werden. In diesem Beispiel ist $f_i = \frac{1}{y_i}$.

Im Beispiel wurde eine Situation dargestellt, in der die Fehler der einzelnen Punkte zwar nicht bekannt sind, wir jedoch ihre relative Größe untereinander kennen. Im Praktikum ist zwar vermutlich die Exponentialfunktion der einzige (und nicht seltene) Fall in dem dies vorkommt, da andere linearisierbare Zusammenhänge auf andere f_i führen, wollen in diesem Abschnitt allgemein den Fall untersuchen, in dem sich die unbekanntem Fehler σ_i als Produkt aus einem bekannten individuellen Faktor f_i und dem unbekanntem gemeinsamen Anteil σ darstellen lässt.

$$\sigma_i = f_i \sigma \quad (45)$$

Die Parameter \hat{a} und \hat{b} können wir weiterhin problemlos nach (38) schätzen. Dazu setzen wir (45) in (38) ein. Dabei kürzt sich der unbekanntem gemeinsame Faktor σ heraus! In (38) treten lediglich die f_i an die Stelle der σ_i . Die Gewichte der Messwerte sind also $w_i = \frac{1}{f_i^2}$:

$$S' = S\sigma^4 = \sum_i \frac{1}{f_i^2} \sum_i \frac{x_i^2}{f_i^2} - \left(\sum_i \frac{x_i}{f_i^2} \right)^2$$

$$\hat{a} = \frac{1}{S'} \left(\sum_i \frac{x_i^2}{f_i^2} \sum_i \frac{y_i}{f_i^2} - \sum_i \frac{x_i}{f_i^2} \sum_i \frac{x_i y_i}{f_i^2} \right) \quad (46a)$$

$$\hat{b} = \frac{1}{S'} \left(\sum_i \frac{1}{f_i^2} \sum_i \frac{x_i y_i}{f_i^2} - \sum_i \frac{x_i}{f_i^2} \sum_i \frac{y_i}{f_i^2} \right) \quad (46b)$$

Wir schätzen nun analog zu (44) das gemeinsame σ^2 . Dabei müssen wir beachten, dass es nicht mehr nur einen Erwartungswert $\langle Y \rangle$, sondern zu jedem Messwert eine Erwartungswert $\langle Y_i \rangle = \langle y(x_i) \rangle$ gibt. So erhalten wir die folgende erwartungstreue Schätzung:

$$s^2 = \frac{\chi^2}{n-2} = \frac{1}{n-2} \sum_{i=1}^n \frac{1}{f_i^2} (y_i - \hat{a} - \hat{b}x_i)^2 \quad (47)$$

Nun können wir durch Einsetzen von $\sigma_i^2 \simeq f_i^2 s^2$ in (39a) und (39b) Schätzungen für die Fehler der Parameter \hat{a} und \hat{b} berechnen.

$$\Delta \hat{a} = \sqrt{\frac{1}{S'} \sum_{i=1}^n \frac{x_i^2}{f_i^2} s^2} \quad \text{und} \quad \Delta \hat{b} = \sqrt{\frac{1}{S'} \sum_{i=1}^n \frac{1}{f_i^2} s^2} \quad (48)$$

Geschafft! Sie sind fast am Ziel!

Noch ein paar Hinweise zu den so aus der Streuung berechneten Fehlern:

- $\Delta \hat{a}$ und $\Delta \hat{b}$ sind für kleine n nicht verlässlich. Hier müssten eigentlich Konfidenzintervalle betrachtet werden. (Siehe **4.3**). Ein Beispiel, in dem Konfidenzintervalle mit der geschätzten Standardabweichung verglichen werden, findet man in [Mand, S. 141-147]
- Die f_i sind nur bis auf einen gemeinsamen Faktor bestimmt. $\frac{\chi^2}{n-2}$ ist von diesem Faktor ebenfalls abhängig. Die daraus berechneten $\Delta \hat{a}$ und $\Delta \hat{b}$ sind davon unabhängig.
- Der Schätzwert für die Varianz eines Messpunktes ist $s_i^2 = f_i^2 s^2 = f_i^2 \frac{\chi^2}{n-2}$.

6 Mehr zur linearen Regression

6.1 Die Bedeutung von χ^2 bei *bekannt* Fehlern

Die in Abschnitt 5.2.2 hergeleiteten Relationen geben uns jetzt ein Mittel in die Hand, die Güte unserer linearen Approximation zu beurteilen, *falls* wir die Fehler der Messwerte *doch* kennen.

Wählt man $f_i = \sigma_i$, dann ist σ definitionsgemäß gleich 1. Der Erwartungswert für s^2 ist (da s^2 erwartungstreu ist) dann ebenfalls gleich 1. Weicht $s^2 = \frac{\chi^2}{n-m}$ ³ stark von 1 ab, können wir mit einer gewissen Wahrscheinlichkeit schließen, dass die Eingangsfehler σ_i schlecht geschätzt wurden. Um dieses Werkzeug systematisch einsetzen zu können, müsste man sich Gedanken machen, mit welcher Wahrscheinlichkeit $\frac{\chi^2}{n-m}$ nicht mehr als um einen bestimmten Betrag vom Erwartungswert 1 abweicht. Darauf können wir hier nicht weiter eingehen.

Im Praktikum werden Sie in der Regel $\frac{\chi^2}{n-m} \ll 1$ beobachten, weil die Eingangsfehler auch systematische Anteile enthalten und nach oben – also zu groß – geschätzt wurden. Erhalten Sie allerdings $\frac{\chi^2}{n-m} > 1$, sollten Sie sich Gedanken machen, ob nicht ungenauer gemessen wurde, als man hinterher angegeben hat oder Fehlerquellen übersehen wurden. Bei einer ausreichenden Anzahl von Messpunkten kann man auch eine Fehlerschätzung nach Abschnitt 5.2.2 in Erwägung ziehen. Dies entspricht etwa dem Vorgehen beim Zeichnen einer Grenzgerade, bei der man sich an der Lage der Punkte orientiert.

6.2 Falsches Modell

Wir haben bisher immer vorausgesetzt, dass der wahre Zusammenhang tatsächlich eine Gerade ist. Wenn dies nicht der Fall ist, sind die Schätzwerte für die Parameter *und* die Fehler bedeutungslos. Es kann leicht passieren, dass Achsenabschnitt und Steigung mit sehr kleinem Fehler geschätzt werden, aber eine graphische Darstellung offenbart, dass die Werte überhaupt keinen linearen Verlauf zeigen. Dass die Fehler trotzdem klein sind, liegt daran, dass in ihre Berechnung die tatsächliche Abweichung der Messwerte von der Schätzgeraden überhaupt nicht eingeht! Verfahren, die diese Abweichungen berücksichtigen, liefern in einem solchen Fall zwar größere Fehler (siehe später). Aber

³In diesem Kontext wird s^2 in der Literatur praktisch ausschließlich als „reduced chi squared“ bezeichnet. Um dieser Konvention zu entsprechen, schreiben wir an dieser Stelle $\frac{\chi^2}{n-m}$, dabei ist $(n - m)$ die Zahl der Freiheitsgrade.

die damit berechneten Werte sind genauso sinnlos. Deshalb ist die graphische Darstellung auch bei rechnerischer Auswertung so wichtig.

6.3 Das Bestimmtheitsmaß R^2

Die Residuen ($\hat{d}_i = y_i - \hat{y}_i$) beschreiben, wie gut jeder einzelne Punkt durch das Modell vorhergesagt wird. Wir möchten jedoch eine Größe finden, die beschreibt, wie gut das Modell insgesamt passt. Hierzu definieren wir zunächst die Streuung der Messwerte y_i als ihre quadratische Abweichung vom ihrem Mittelwert \bar{y} . Die Summe über diese Abweichungen nennt man *SQT* („Sum of Squares Total“):

$$SQT = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (49)$$

Um nun ein Modell zu beurteilen, können wir die Streuung, die dieses Modell erklärt, untersuchen. Die zugehörige Summe heißt *SQE* („Sum of Squares Explained“) und wird aus den Abständen zwischen den vom Modell vorhergesagten Werten \hat{y}_i und dem Mittelwert aller Messwerte \bar{y} gebildet:

$$SQE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (50)$$

Als Maß für die Güte des Modells definieren wir das Bestimmtheitsmaß R^2 als den Anteil der Streuung, den das Modell erklärt. Also bewegt sich R^2 zwischen 0 und 1.

$$R^2 = \frac{SQE}{SQT} \quad (51)$$

Für diese Größen gilt die folgende *Streuungszerlegung* (52), deren Gültigkeit nicht ohne weiteres einsichtig ist. Dabei stellt sich heraus, dass die Streuung der Messwerte einfach die Summe der durch das Modell erklärten Streuung und der Residualstreuung (die Summe der Quadrate der Residuen \hat{d}_i SQR („Sum of Squared Residuals“)) ist:

$$SQT = SQE + SQR$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (52)$$

Daraus folgt eine praktischere Formel für die Berechnung von R^2 :

$$R^2 = \frac{SQE}{SQT} = \frac{SQT - SQR}{SQT} = 1 - \frac{SQR}{SQT} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Liegen alle beobachteten Punkte exakt auf einer Geraden, so sind die Residuen alle gleich Null und ebenso die Residualstreuung. In diesem Fall ist also die Gesamtstreuung gleich der erklärten Streuung, d.h. die gesamte Variation von Y lässt sich

durch die Variation von X zusammen mit der postulierten linearen Beziehung erklären. In diesem Fall ist R^2 also 1. Je grösser nun die Residualstreuung ist, desto schlechter beschreibt das Modell die Daten, d.h. desto weniger wird die in den Daten vorhandene Streuung durch das Modell erklärt.

Haben die Messwerte unterschiedliche Fehler, treten an die Stelle der Mittelwerte gewichtete Mittelwerte. Auch die einzelnen Summanden müssen gewichtet werden.

$$R^2 = 1 - \frac{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2}}{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{\sigma_i^2}} = 1 - \frac{\sum_{i=1}^n \frac{(y_i - \hat{a} - \hat{b}x_i)^2}{\sigma_i^2}}{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{\sigma_i^2}} \quad \text{mit } \bar{y} = \frac{\sum_{i=1}^n \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \quad (53)$$

[Fahr, S. 158ff] Grenzen innerhalb derer ein Wert für R^2 akzeptabel ist, lassen sich nicht so ohne weiteres angeben. Man könnte jedoch eine Grenze angeben, so dass mit einer vorgegebenen Wahrscheinlichkeit R^2 kleiner als diese Grenze ist, falls es sich um einen linearen Zusammenhang handelt. Diese Grenze ist natürlich von der Zahl n der Messwerte abhängig. Das heisst *nicht*, dass wenn R^2 kleiner als diese Grenze ist, der wahre Zusammenhang mit der angegebenen Wahrscheinlichkeit linear ist!

6.4 *Fehler in den x -Werten

Sind nur die x -Werte fehlerbehaftet, bzw. die y -Fehler gegenüber den x -Fehlern vernachlässigbar, vertauschen Sie einfach abhängige und unabhängige Variable.

Die Umrechnung der so erhaltenen Parameter ist unproblematisch. Auch der Fehler der Steigung kann problemlos nach dem Fehlerfortpflanzungsgesetz berechnet werden. Bei der Umrechnung des Achsenabschnitts tritt das selbe Problem wie in 6.6 auf, da die Fehler von Achsenabschnitt und Steigung nicht unabhängig sind. Glücklicherweise wird im Praktikum häufig nur die Steigung benötigt.

6.5 * x - und y -Werte fehlerbehaftet

Diese Aufgabe lässt sich nicht ganz so einfach lösen. Das Prinzip der kleinsten Quadrate kann hier nicht so ohne weiteres angewendet werden, denn es verrät nicht, welche quadratischen Abstände minimiert werden sollen. Der kürzeste Abstand r_i vom Messpunkt zur Ausgleichsgeraden ist es im Allgemeinen jedenfalls nicht (Abb. 5), denn dann wäre die Lage der Geraden von der Achsenskalierung abhängig. Glücklicherweise hilft auch hier das ML-Prinzip weiter. [Bar, S. 109]

6.6 *Interpolation und Kalibrierkurven

Gelegentlich führt man zunächst eine Messreihe durch, aus der man eine Ausgleichsgerade bestimmt, um dann später die Grösse y zu messen indem man x misst und y mit Hilfe dieser *Kalibrierkurve* ausrechnet – oder umgekehrt. Das funktioniert indem man einfach den Messwert x in die Geradengleichung der Kalibrierkurve einsetzt. Leider ist die Berechnung des Fehlers des so geschätzten y -Wertes nicht ganz so unproblematisch, weil Steigung und Achsenabschnitt nicht statistisch unabhängig sind. Wenn man einfach das Gaußsche Fehlerfortpflanzungsgesetz anwendet, um den Fehler dieses Schätzwertes zu bestimmen, erhält man deshalb einen zu kleinen Wert. Mit dem Größtfehler liegt man zwar auf der sicheren Seite, aber er ist wie der Name schon sagt unangemessen groß. Auch hier lässt uns die Mathematik nicht im Stich. Durch eine einfache Koordinatentransformation ist es möglich, eine neue Geradengleichung mit statistisch unabhängigen Parametern zu erhalten. [Bar, S. 103]

6.7 *Ausblick: Matrixdarstellung und nichtlineare Modelle

Wir haben gesehen, dass die Normalgleichungen ein lineares Gleichungssystem darstellen. Man kann dieses Gleichungssystem in der Sprache der linearen Algebra als Matrixgleichung schreiben. Die übrigen Rechnungen werden dann viel kürzer. ([Bar] S. 103) Insbesondere wenn kompliziertere lineare Modelle als eine einfache Gerade verwendet werden, ist diese Darstellung vorzuziehen, auch weil sich solche Gleichungssysteme besonders gut programmieren lassen.

Auch auf nicht lineare Modelle kann das ML-Prinzip bzw. die Methode der kleinsten Quadrate angewendet werden. In diesem Fall sind die Normalgleichungen leider nicht linear und man ist auf numerische Lösungen angewiesen. Eine Anwendung eines solchen nichtlinearen Modells auf die simultane Bestimmung der Halbwertszeiten von ^{108}Ag und ^{110}Ag , die Teil des Versuchs *radioaktiver Zerfall* ist, findet sich in [Bev].

Literatur

- [Bar] BARLOW, ROGER J. *Statistics: a guide to the use of statistical methods in the physical sciences*. Wiley, 1989; Nachdr. 1999.
- [Bev] BEVINGTON, PHILIP R.; ROBINSON, D. KEITH. *Data reduction and error analysis for the physical sciences*. McGraw-Hill, 2. Aufl. 1992.

- [Fahr] FAHRMEIR, LUDWIG [u.a.] *Statistik – der Weg zur Datenanalyse*. Springer, 1997.
- [Mand] MANDEL, JOHN. *The statistical analysis of experimental Data*. New York: Interscience, 1964; korr. Nachdr. New York: Dover, 1984.
- [Squi] SQUIRES, G. L.. *Messergebnisse und ihre Auswertung: Eine Anleitung zum praktischen naturwissenschaftlichen Arbeiten*. Walter de Gruyter, 1971.

Als Einführung in die Fehlerrechnung und Datenanalyse für Physiker sind vor allem [Bar] und [Bev] gut geeignet. Am Ende einiger Abschnitte in diesem Skript finden Sie Hinweise, wo das jeweilige Thema ausführlicher dargestellt wird.