

VORLESUNG

# **ANALYTISCHE CHEMIE III**

## **Chemometrie Teil 1: Quantitative Auswertung analytisch-chemischer Messresultate**

Dr. M. Badertscher

Das vorliegende Skript entstand unter Mitarbeit von

Prof. Dr. E. Pretsch  
Dr. D. Wegmann †



# Inhaltsverzeichnis

1. Einleitung .....	2
2. Ausgewählte Literatur .....	3
3. Zusammenhänge zwischen Variablen .....	4
3.1. Beschreibung von Zusammenhängen .....	5
4. Verhalten der Messfehler .....	6
4.1. Qualitative Beschreibung von Messfehlern .....	6
4.2. Quantitative Erfassung von Messfehlern .....	7
4.3. Modellverteilungen .....	10
4.3.1. Normalverteilung .....	10
4.3.2. Alternativen zur Normalverteilung .....	11
4.3.3. Lognormalverteilung .....	13
4.4. Zentraler Grenzwertsatz .....	13
4.5. Berechnung von Fehlerschranken aus geschätzten Verteilungen .....	14
4.6. Fehlerfortpflanzung .....	17
4.6.1. Fortpflanzung systematischer Fehler .....	18
4.6.2. Fortpflanzung zufälliger Fehler .....	20
5. Grundlagen statistischer Tests .....	23
5.1. Risiko falscher Entscheide .....	23
5.2. Beispiel eines zweiseitigen statistischen Tests .....	24
5.3. Beispiel eines einseitigen statistischen Tests .....	26
5.4. Entscheidungsstrategie .....	29
6. Schrittweises Vorgehen beim Entwurf eines Modells .....	31
6.1. Charakterisierung der Variablen und ihrer Zusammenhänge .....	31
6.2. Charakterisierung der Messfehler .....	32
6.3. Mathematische Behandlung des Modells .....	35
6.4. Test des Modells und des Datensatzes .....	41
6.4.1. Residuenanalyse .....	41
6.4.2. Systematische Fehler .....	41
6.4.3. Ausreisser .....	42
6.4.4. Hebelpunkte .....	42
6.4.5. Übermäßige Flexibilität des Modells .....	44
6.5. Entscheidung aufgrund des Schlussresultates .....	47
7. Herleitung der Linearen Regression .....	48
7.1. Formulierung der Modellfunktion .....	48
7.2. Beschreibung der experimentellen Messwerte .....	48
7.3. Schätzungen der Parameter .....	48
7.4. Schätzung der Fehlervarianz .....	51
7.5. Vertrauensbereich der Parameter .....	51
7.6. VB für den Erwartungswert eines Messwertes bei ausgewählten x-Werten ....	53
7.7. Vertrauensbereich für einen Wert x aus Messungen von y .....	55
8. Sammlung statistischer Modelle .....	59
8.1. Proportionaler Zusammenhang mit Lognormalverteilung .....	59
8.2. Proportionaler Zusammenhang mit Normalverteilung .....	66
8.3. Linearer Zusammenhang mit Normalverteilung .....	71
8.4. Standardaddition .....	76
9. Sammlung statistischer Tests .....	80
9.1. Vergleich zweier Varianzen, F-Test .....	80
9.2. Vergleich zweier Mittelwerte, t-Test .....	81
9.3. Vergleich zusammengehörender Datenpaare .....	82

# 1. Einleitung

In der analytischen Chemie werden typischerweise Messungen durchgeführt, die zu Entscheidungen führen sollen. So wird etwa in der Qualitätskontrolle eines pharmazeutischen Betriebs die Konzentration von Verunreinigungen in einem Medikament bestimmt, um zu gewährleisten, dass die gesetzlichen Richtlinien eingehalten werden.

Um ein bestimmtes experimentelles Problem zu lösen, besteht die Aufgabe des analytischen Chemikers zunächst darin, verschiedene mögliche Messanordnungen vorzuschlagen und nach mehreren Gesichtspunkten zu evaluieren. Dabei wird ein qualitatives Vorgehen festgelegt. Typische Fragen, die dabei diskutiert werden:

- Wie wird die Probe genommen?
- Soll eine Aufarbeitung der Probe erfolgen?
- Ist eine chromatographische Trennmethode einzusetzen?
- Wie wird detektiert?
- Wie wird der Detektor kalibriert?

Solche Fragen sind nicht Thema dieser Vorlesung. Es wird stattdessen angenommen, dieser Teil der Planung sei bereits abgeschlossen. Es liegt mithin ein detaillierter Plan zur Führung der Experimente vor.

Spätestens wenn die Messapparatur einmal aufgebaut ist und erste Messungen durchgeführt werden können, ist festzulegen, wie die numerischen Messwerte zu einem Schlussresultat aufgearbeitet werden sollen. Das Experiment ist mit einem quantitativen Modell zu beschreiben. Dabei werden folgende Fragen gestellt:

- Was ist der funktionale Zusammenhang zwischen den relevanten physikalischen Grössen?
- Gibt es Störeinflüsse, die vernachlässigt werden?
- Was kann eine Vernachlässigung schlimmstenfalls bewirken?
- Welche Quellen experimenteller Fehler sind zu berücksichtigen?
- Welchen Gesetzmässigkeiten folgen die Fehler?
- Wie äussern sich Abweichungen von den angenommenen Eigenschaften des Experiments?

Diese Fragen zu behandeln, ist Gegenstand des ersten Teils der Vorlesung. Dabei wird ein systematisches Vorgehen dargelegt, mit dessen Hilfe die anstehenden Schritte in der richtigen Reihenfolge durchgeführt werden können.

## 2. Ausgewählte Literatur

M. G. Kendall, A. Stuart, The advanced theory of statistics, (three-volume edition), Ed.4, Griffin, London; High Wycombe 1977-1983

A. Stuart, K. Ord, Kendall's Advanced theory of statistics, originally by M. G. Kendall, Ed.5, Griffin, London 1987-

D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. D. Jong, P. J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics & Qualimetrics, Elsevier, Amsterdam 1997, 1998

N. R. Draper, H. Smith, Applied regression analysis, Ed. 2, Wiley, New York 1981

P. C. Meier, R. E. Zünd, Statistical Methods in Analytical Chemistry, Wiley-Interscience, New York 1993  
(Chemical Analysis, vol. 123)

S. N. Deming, S. L. Morgan, Experimental design, Elsevier, Amsterdam 1993  
(Data handling in science and technology; vol. 11)

K. Doerffel, K. Eckschlag, Chemometrische Strategien in der Analytik, VEB, Leipzig 1990

L. Sachs, Angewandte Statistik: Anwendung statistischer Methoden, Elfte Auflage, Springer, Berlin 2004

Wissenschaftliche Tabellen Geigy, Band 3: Statistik, CIBA-GEIGY AG, Basel 1980

W. H. Press, B. P. Flannery, S. A. Teucholsky, W. T. Vetterling, Numerical Recipes, Cambridge University Press, Cambridge 1989

D. Huff, How to lie with statistics, Gollancz, London 1954

### 3. Zusammenhänge zwischen Variablen

Wählt man zwei beliebige messbare Grössen aus dem täglichen Leben, so lässt sich mit grösster Wahrscheinlichkeit ein Zusammenhang zwischen ihnen feststellen. Der Zusammenhang mag schwach sein, aber vollständige Unabhängigkeit findet man kaum. Auf den ersten Blick ist das eine absurde Behauptung, die zum Widerspruch reizt. Was hat die Intensität der eigenen Kopfschmerzen mit der Unfallhäufigkeit auf Autobahnen zu tun?

Klar erkennbare Zusammenhänge sind oft kausal. Je länger man die Pfanne über der Flamme stehen lässt, desto heisser wird das Kaffeewasser. Eine Variable hängt von der anderen ab, nicht umgekehrt. Ein Physiker könnte den Zusammenhang sogar formelmässig beschreiben. Dabei würde er eine Funktion definieren, in der die eine Variable als Argument erscheint, die andere Variable als Funktionswert. Das widerspiegelt die Kausalität. Der Funktionswert reagiert auf Änderungen der Argumente und kann nicht anderweitig beeinflusst werden. Es hat keinen Sinn, das Argument über eine Manipulation des Funktionswerts beeinflussen zu wollen. Die Zeit kann nicht über die Temperatur des Wassers beeinflusst werden.

Für die Richtung der Kausalität ist nicht in erster Linie eine Sammlung von Naturgesetzen verantwortlich, sondern der Aufbau eines Experiments. Die Zustandsgleichung eines Gases verknüpft die Grössen Temperatur, Volumen, Druck und Stoffmenge über einen klar formulierten Zusammenhang. Keine der Variablen ist vor einer anderen irgendwie ausgezeichnet. Betrachtet man aber ein Experiment, bei dem ein Kolben in einem thermostatisierten Zylinder verschoben wird, ist der Druck kausal vom Volumen abhängig. In einem anderen Experiment könnte man über eine chemische Reaktion ein Gas entstehen lassen. Dann würde z. B. das Volumen von der Stoffmenge kausal abhängen. Die Richtung der Kausalität kann in diesem Sinn auch umgekehrt werden. In zwei unterschiedlichen Experimenten können zwei Variablen entgegengesetzte Rollen spielen, obwohl das gleiche Naturgesetz für ihren Zusammenhang verantwortlich ist.

Kausale Zusammenhänge zwischen zufällig ausgewählten Variablen sind äusserst selten. Davon rührt wohl unser Widerstand her, an einen Zusammenhang zwischen beliebigen Variablen zu glauben. Es gibt im wesentlichen zwei allgegenwärtige Ursachen für nichtkausale Zusammenhänge:

- Zusammenhang über eine dritte Variable
- Zusammenhang über eine Nebenbedingung

Viele Variablen hängen irgendwie von der Zeit ab. Einige dieser Zusammenhänge kann man als kausal bezeichnen, andere sind schwierig zu beschreiben, obwohl offensichtlich. Wählt man nun zwei solche Variablen aus, so ergibt sich fast zwangsläufig ein Zusammenhang zwischen ihnen, jedenfalls wenn die Zeit spielen kann. Wann immer zwei Variablen irgendwie von einer dritten abhängen, sind sie gekoppelt. Typische Drittvariablen sind die Zeit, die Temperatur, das Geschlecht (bei medizinischen Untersuchungen). Es ist schwer, in der analytischen Chemie eine Variable zu finden, die weder von der Zeit noch von der Temperatur abhängt. Entsprechend schwierig ist es, zwei Variablen zu finden, die nicht irgendwie zusammenhängen.

Variablen können auch durch eine Nebenbedingung gekoppelt sein. Wenn die Summe von mehreren Grössen zusammen 100% ergeben muss, hängen die einzelnen Grössen mehr oder minder voneinander ab. Im extremen Fall von zwei Variablen, die so gekoppelt sind, ist der Zusammenhang sogar exakt bekannt. Bei fünf Variablen könnte eine experimentell festgestellte Korrelation leicht als "Entdeckung" missinterpretiert werden.

### 3.1. Beschreibung von Zusammenhängen

Es besteht ein prinzipieller Unterschied zwischen einem Modell zur Beschreibung der Flugbahnen von Himmelskörpern im Sonnensystem einerseits, und dem Einfluss der Konjunktur auf den Energieverbrauch in einem Industrieland andererseits. Im einen Fall existieren gesicherte Erkenntnisse über den funktionalen Zusammenhang der relevanten Grössen. Die Natur hält sich an Gesetze, die im Rahmen der gegenwärtig besten experimentellen Resultate als gesichert gelten können. Darauf aufbauend lassen sich die Zusammenhänge zwischen Variablen mit mathematischer Exaktheit als Funktion beschreiben. Im anderen Fall ist der Zusammenhang so ausserordentlich komplex, dass eine derartige Beschreibung von vornherein ausser Betracht fällt. Trotzdem existieren klar erkennbare Zusammenhänge, die man zu quantifizieren wünscht. Man ist dabei allerdings auf weniger griffige Beschreibungen angewiesen. Man findet Möglichkeiten zur Beschreibung von diffusen Zusammenhängen im Bereich der mathematischen Statistik.

Im Bereich der analytischen Chemie sind viele Zusammenhänge genügend gut bekannt, um eine Funktion als Beschreibung zu finden. Man kennt also die kausalen Zusammenhänge. Sollten experimentelle Daten dennoch Abweichungen vom angenommenen Verhalten zeigen, ist das ausschliesslich auf Vernachlässigungen und Näherungen in der Beschreibung zurückzuführen, sowie insbesondere auf Messfehler. Im ersten Teil der Vorlesung werden nur solche Fälle behandelt.

## 4. Verhalten der Messfehler

### 4.1. Qualitative Beschreibung von Messfehlern

Alle Messwerte sind fehlerbehaftet. Man unterscheidet drei Arten von Messfehlern:

- grobe Fehler (gross errors)
- systematische Fehler (systematic errors)
- zufällige Fehler (random errors)

Grobe Fehler beruhen meist auf "Unfällen". Falsch abgelesene Werte, vergessene Kommas und defekte Apparaturen führen zu unsinnigen Werten. Grobe Fehler sind meist harmlos, da sie spätestens bei der Auswertung als solche erkannt werden.

Systematische Fehler verfälschen alle Messwerte in der gleichen Weise. So kann etwa ein ungünstiger Ablesewinkel dafür verantwortlich sein, dass alle Messwerte zu hoch ausfallen. Systematische Fehler beeinflussen die Richtigkeit (accuracy) von Messresultaten.

Man verwechsle systematische Fehler nicht mit Spezifikationsfehlern (specification errors), die ebenfalls systematischen Charakter haben können. Dabei sind aber nicht die Messwerte selbst fehlerbehaftet, sondern es gibt spürbare Abweichungen von deren vorausgesetztem Verhalten. Das Modell ist dann verbesserungswürdig.

Zufällige Fehler sind unvermeidliche Schwankungen um einen Mittelwert. Typische Quellen zufälliger Fehler sind Rauschen, Inhomogenitäten und viele kleine Störeinflüsse. Zufällige Fehler beeinflussen die Präzision (precision) von Messresultaten. Grössen, die mit zufälligen Fehlern behaftet sind, nennt man Zufallsvariablen (random variables).

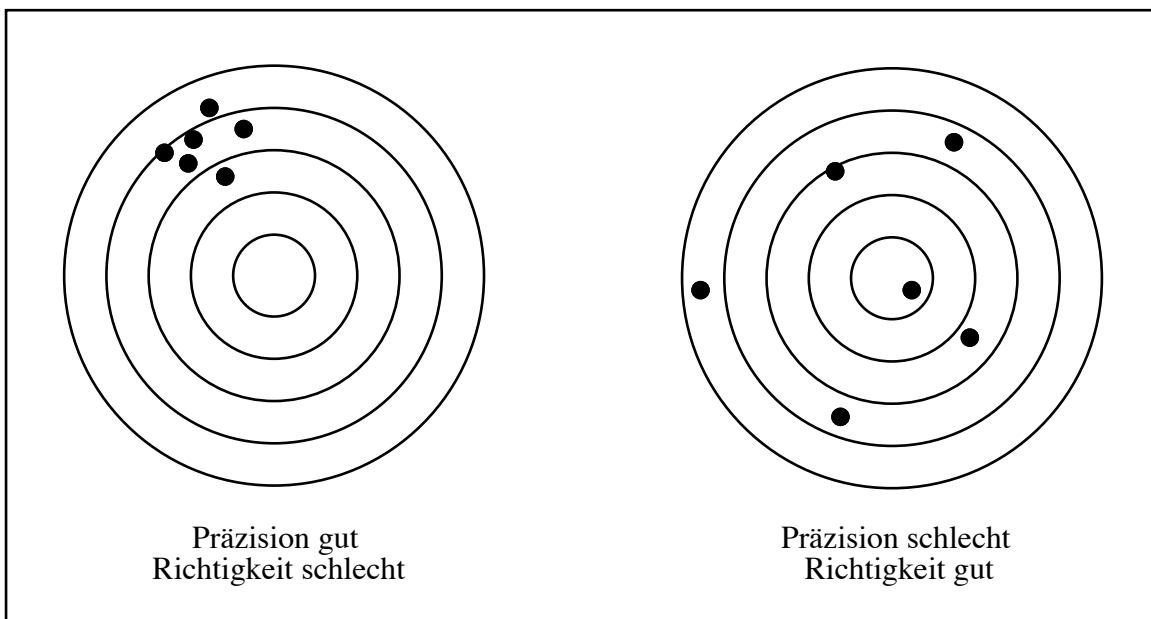


Fig. 4.1 Veranschaulichung der Begriffe Richtigkeit und Präzision

Oft werden auch andere Begriffe als Präzision und Richtigkeit gebraucht, die aber meist mit einer anderen Bedeutung in der Umgangssprache in Konflikt stehen. Dies gilt besonders für den Begriff "Genauigkeit", der hier nicht definiert wird. Die Verhältnisse sind nur in der englischen Fachsprache klar:

accuracy: Richtigkeit

precision: Präzision



## 4.2. Quantitative Erfassung von Messfehlern

Das Verhalten von systematischen Fehlern lässt sich nicht vorhersagen. Daher kann man sie nur korrigieren, wenn man Information "von aussen" über sie hat. Durch eine Bearbeitung der Messwerte ohne Zusatzinformation sind systematische Fehler nicht fassbar. Darum gibt es über sie auch keine Fehlertheorie.

Zufällige Fehler hingegen lassen sich bei blosser Betrachtung der Messwerte teilweise erkennen und entsprechend eliminieren. Das zufällige Verhalten der Fehler folgt gewissen Gesetzmässigkeiten, die man zur Formulierung einer Fehlertheorie verwenden kann. Je mehr man über das Verhalten der Fehler weiss, desto kraftvoller kann man die Methode zu ihrer Erkennung gestalten.

Wenn im folgenden von "Messfehlern" die Rede ist, sind prinzipiell nur zufällige Fehler gemeint.

Messwerte (wahrer Wert + Messfehler) sind Zufallsvariablen. Das Verhalten einer Zufallsvariablen lässt sich durch ihre Wahrscheinlichkeitsdichte-Verteilung (probability density distribution) oder kurz Dichtefunktion charakterisieren (siehe Fig. 4.2). Das bestimmte Integral unter der Verteilungskurve ist die Wahrscheinlichkeit, mit der ein Messwert zwischen den Integralgrenzen zu finden ist.

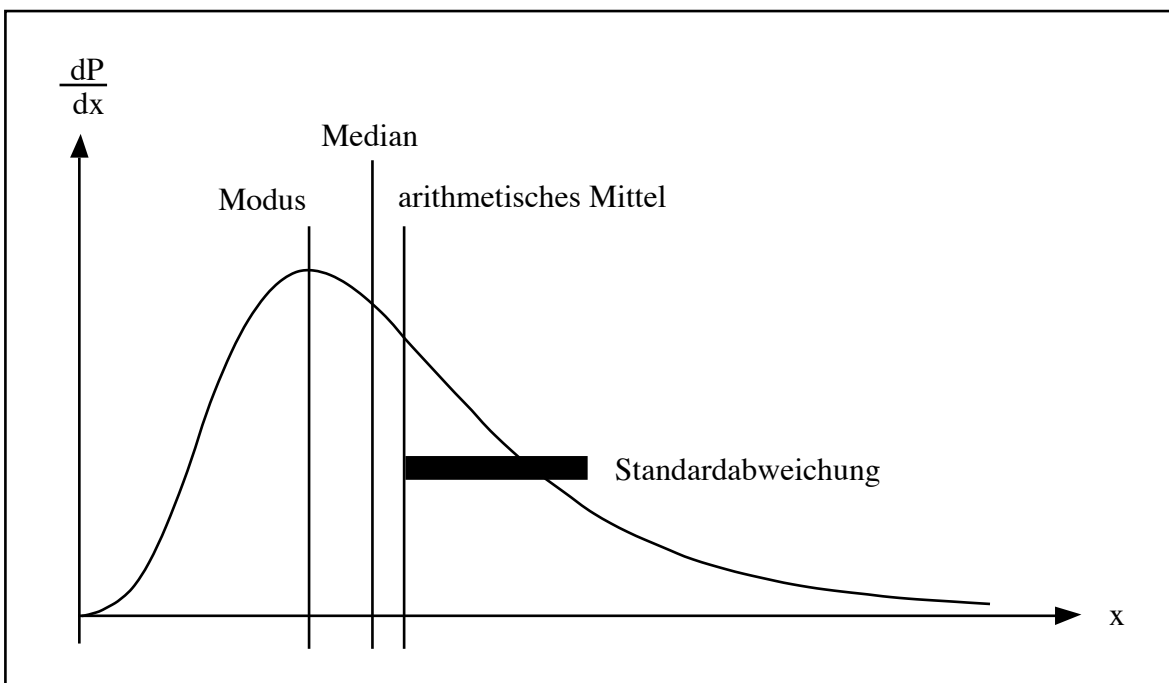


Fig. 4.2 Allgemeine Dichtefunktion

Eng verknüpft mit der Dichtefunktion ist das Konzept der Grundgesamtheit (population). Sämtliche Messwerte, die man unter gegebenen Bedingungen gewinnen könnte, bilden die Grundgesamtheit, ein hypothetisches Gebilde. Jedes Experiment umfasst nur eine begrenzte Anzahl Messwerte. In der Statistik hat sich dafür die Bezeichnung Stichprobe (sample) eingebürgert. Die Anzahl Messwerte bezeichnet man als Stichprobenumfang (sample size).

Wenn man die Dichtefunktion kennt, weiss man alles, was es über die Messwerte zu wissen gibt. Ziel eines Experiments ist es demnach, die Dichtefunktion der Messwerte näherungsweise zu bestimmen.

Für die in der analytischen Chemie üblichen Stichprobenumfänge von typischerweise weniger als 10 Messwerten scheint dieses Unterfangen hoffnungslos. Wie soll man mit wenigen Messwerten den Kurvenzug der Verteilungsfunktion rekonstruieren können?

Man mag einwenden, dass das auch gar nicht nötig sei. Das Schlussresultat einer Konzentrationsbestimmung wird höchst selten in der Form einer Dichtefunktion angegeben. Stattdessen hat es die Form eines Mittelwertes mit Fehlerschranke. Es ist aber höchst wünschenswert, der Fehlerschranke die Wahrscheinlichkeit zuzuordnen, mit der der wahre Wert zwischen den Schranken liegt. Man nennt diese Wahrscheinlichkeit statistische Sicherheit (confidence coefficient). Man will sie frei wählen. Dazu muss aber zu jeder statistischen Sicherheit eine Fehlerschranke definiert sein. Das ist genau die Information, die in der vollständigen Verteilungsfunktion steckt. Man kommt also letztlich nicht darum herum, die Verteilungsfunktion zu schätzen.

Es gibt Situationen, in denen man fast nichts über die Form der Verteilung weiss oder sich keinerlei Annahmen erlaubt. Hat man nur eine kleine Stichprobe, gibt es kaum eine Möglichkeit, eine fundierte Fehlerschranke mit zugehöriger Wahrscheinlichkeit anzugeben. In diesen Fällen beschränkt man sich auf die Angabe anderer Grössen.

Für jede Verteilung lassen sich einige Kenngrössen definieren, die die Verteilung in einem beschränkten Ausmass charakterisieren. Siehe Fig. 4.2. Diese Grössen haben eine greifbare Bedeutung und lassen sich aus einer Stichprobe schätzen. Man bezeichnet in der Statistik die Kenngrössen mit griechischen Buchstaben, deren Schätzungen mit dem entsprechenden lateinischen Buchstaben, oder man setzt auf das Symbol der Kenngrösse ein Dach (engl. hat).

Der Modus (mode) ist der Wert bei der grössten Wahrscheinlichkeitsdichte. Diese Grösse zieht zwar die Aufmerksamkeit auf sich, ist aber von untergeordneter Bedeutung. Eine Schätzung ist nur möglich, wenn man die Form der Verteilung kennt.

Der Median (median) einer Verteilung teilt die Wahrscheinlichkeit in zwei gleiche Hälften. Beidseits des Medians befinden sich also 50 % der Wahrscheinlichkeit. Zur Schätzung ordnet man die Messwerte nach ihrer Grösse. Bei ungeradem Stichprobenumfang ist der geschätzte Median der mittlere Wert, bei geradem Stichprobenumfang das arithmetische Mittel der beiden mittleren Messwerte.

Das arithmetische Mittel  $\mu$  (mean) einer kontinuierlichen Verteilung ist definiert als:

$$\mu = \int_{-\infty}^{\infty} x \frac{dP}{dx}(x) dx \quad (4.1)$$

Die Schätzung  $\bar{x}$  aus einer Stichprobe mit Umfang  $n$  lässt sich so berechnen:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.2)$$

Die Schätzung ist nur richtig, wenn die Messwerte nicht seriell korreliert sind, d. h., wenn ein Messwert nicht beeinflusst wird von den vorher gemessenen Werten. Im Allgemeinen kann sich ein Messgerät nicht an die vorhergehenden Messwerte erinnern. Es gibt aber Ausnahmen. In einem solchen Fall ist eine Zeitreihenanalyse (time series analysis) nötig, um den Mittelwert sauber schätzen zu können. Dieser Fall wird hier nicht behandelt.

Die Varianz (variance) ist ein Streuungsmass:

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{dP}{dx}(x) dx \quad (4.3)$$

Die Quadratwurzel der Varianz ist anschaulicher und wird als Standardabweichung (standard deviation) bezeichnet. Die Schätzung  $s_x$  aus einer Stichprobe:

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (4.4)$$

Auch diese Schätzung ist nur richtig, wenn die Messwerte nicht seriell korreliert sind.

Zu jeder geschätzten Standardabweichung gehört die Anzahl Freiheitsgrade  $v$  (number of degrees of freedom). Sie gibt an, mit wievielen unabhängigen Einzelmessungen die Standardabweichung geschätzt wurde. Hier ist

$$v = n - 1 \quad (4.5)$$

und nicht  $n$ , weil für  $s_x$  die Abweichungen vom Mittelwert  $\bar{x}$  verwendet wurden.  $\bar{x}$  wurde aber seinerseits aus der Stichprobe geschätzt.

Weitere Kennzahlen wie Schiefe oder Wölbung werden in der analytischen Chemie selten verwendet. Details entnehme man der Fachliteratur.

Häufig sind die Verteilungen eingipfelig und symmetrisch. Median, Mittelwert und Modus fallen dann zusammen (deren Schätzungen aber nicht notwendigerweise).

Wie soll man nun aber das Schlussresultat formulieren? Gewünscht wird je eine Angabe über Mittelwert und Streuung. Man könnte also das arithmetische Mittel und die Standardabweichung der Messwerte angeben. Die Standardabweichung der Einzelwerte (vgl. 4.4) ist charakteristisch für die Messmethode, aber kein geeignetes Streuungsmass für den Mittelwert, da es den Stichprobenumfang nicht angemessen berücksichtigt. Die Schätzung des Mittelwertes ist umso besser, je mehr Messwerte zur Verfügung stehen.

Würden mehrere Stichproben vom Umfang  $n$  genommen, so sähe man, dass die Mittelwerte  $\bar{x}$  daraus ebenfalls um den wahren Wert  $\mu$  schwanken. Man kann die Standardabweichung  $s_{\bar{x}}$  der Mittelwerte  $\bar{x}$  aber bereits aus einer einzigen Stichprobe schätzen:

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} \quad \text{und entsprechend gilt auch} \quad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \quad (4.6)$$

Das Schlussresultat könnte also aus dem arithmetischen Mittel und der Standardabweichung des Mittelwertes bestehen. Auf die Angabe einer zugehörigen Wahrscheinlichkeit müsste man verzichten, da man die Form der Verteilung nicht kennt. Das ist sehr unbefriedigend. Irgend eine Angabe über die Wahrscheinlichkeit, und sei sie noch so schwach, ist sehr wünschenswert. Erlaubt man sich wirklich keinerlei Annahmen über die Form der Verteilung, gibt es nur eine schwache Beziehung, die man zur Berechnung von Wahrscheinlichkeitsaussagen verwenden kann. Alle Verteilungen mit definierter Varianz erfüllen die Ungleichung von Bienaymé (1853), die aber meist Tschebyschew (1874) zugeschrieben wird:

$$P\left(\frac{|x-\mu|}{\sigma} > a\right) < \frac{1}{a^2} \quad \text{mit beliebigem } a \geq 1 \quad (4.7)$$

In Worten: Die Ungleichung  $\frac{|x-\mu|}{\sigma} > a$  ist mit einer Wahrscheinlichkeit von höchstens  $\frac{1}{a^2}$  erfüllt.

Dabei ist  $x$  ein Einzelwert aus der Grundgesamtheit,  $\mu$  der Mittelwert und  $\sigma$  die Standardabweichung. Fehlerschranken, die auf der Ungleichung von Bienaymé beruhen, sind für die Zwecke der analytischen Chemie unvernünftig weit, besonders wenn eine grosse statistische Sicherheit gewählt

wird. Die Verteilung von Messfehlern ist meist eingipflig und symmetrisch. Für solche Verteilungen gilt die stärkere Ungleichung von Gauss, die schon 1821 gefunden wurde:

$$P\left(\frac{|x-\mu|}{\sigma} > a\right) < \frac{4}{9a^2} \quad \text{mit beliebigem } a \geq 1 \quad (4.8)$$

Die Fehlerschranken werden also gegenüber der Ungleichung von Bienaymé um einen Faktor 1,5 reduziert. Leider sind die Beziehungen nur brauchbar, wenn die wahre Standardabweichung  $\sigma$  der Grundgesamtheit bekannt ist oder zumindest eine gute Schätzung dafür existiert. Das wird im Allgemeinen gerade dann nicht der Fall sein, wenn man sich keine Annahmen über die Form der Verteilung erlauben darf.

Diese Situation ist aber durchaus hypothetisch. Im Allgemeinen verfügt ein analytischer Chemiker sehr wohl über Information betreffend die Form der Verteilung. Durch die Wahl einer geeignet scheinenden Modellverteilung legt man die prinzipielle Form der Verteilung fest. Es gibt Methoden statistischer Art, anhand eines experimentellen Datensatzes nachträglich Abweichungen vom vorausgesetzten Verhalten zu erkennen.

### 4.3. Modellverteilungen

Die Vielfalt der möglichen Verteilungen wird durch Annahmen über die Form drastisch eingeschränkt. Zudem wird dadurch die Beschreibung der Verteilung kompakt. Die Verteilung lässt sich mit wenigen Kenngrößen vollständig beschreiben. Man bezeichnet solche Kenngrößen als Parameter (parameters).

#### 4.3.1. Normalverteilung

Eine häufig gebrauchte Modellverteilung ist die Normalverteilung (normal distribution) (siehe Fig. 4.3). Messwerte sind häufig annähernd normalverteilt. Die Normalverteilung ist zudem besonders beliebt, weil sie sich mathematisch vergleichsweise leicht handhaben lässt. Die Normalverteilung ist eine symmetrische Verteilung mit der Wahrscheinlichkeitsdichte:

$$\frac{dP}{dx} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (4.9)$$

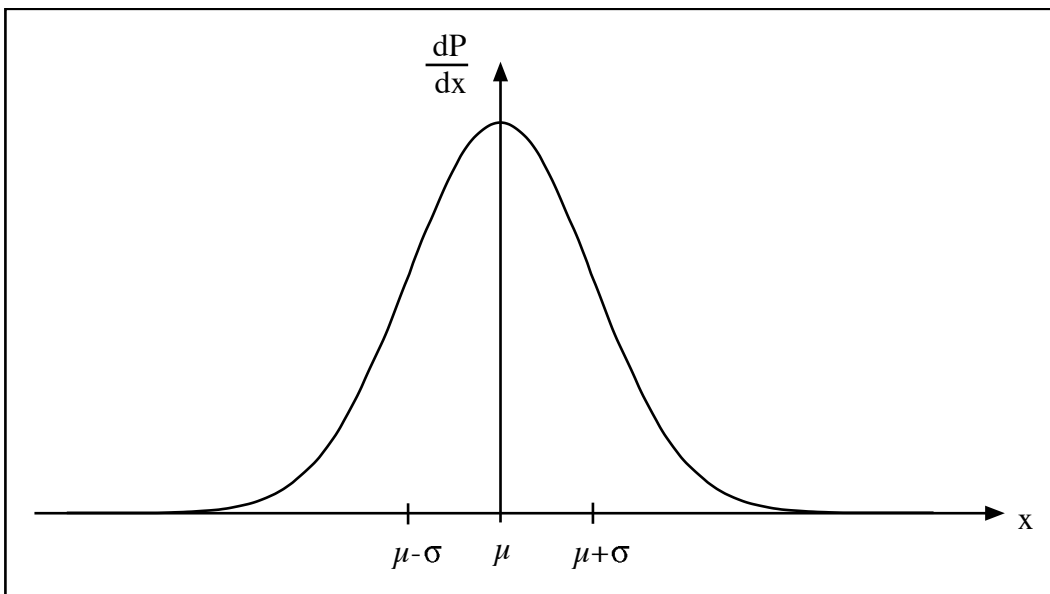


Fig. 4.3 Normalverteilung

Wie man sieht, ist die Verteilung durch zwei Parameter, Mittelwert und Standardabweichung, vollständig beschrieben.

Die Normalverteilung ist eine ausgesprochen enge Verteilung, d. h. die Wahrscheinlichkeitsdichte geht für grosse Abweichungen vom Mittelwert ausserordentlich rasch gegen null. Messfehler sind selbstverständlich nie streng normalverteilt. Die Abweichungen halten sich aber im Allgemeinen in engen Grenzen, wenn man das Modell nicht strapaziert. Die grösste Schwäche der Normalverteilung ist ihre Unfähigkeit, die Häufigkeit von Messfehlern in mehr als 3 Standardabweichungen Abstand vom Mittelwert zu beschreiben. Messfehler erscheinen wesentlich häufiger in den Aussenbezirken der Verteilung, als die Normalverteilung voraussagt (siehe auch 6.4.3., Ausreisserproblem). Die Verteilung von Messfehlern ist nur in seltsamen Fällen enger als die Normalverteilung. Auf der Normalverteilung beruhende Fehlerschranken sind daher öfters zu eng, besonders bei grosser statistischer Sicherheit. Aus diesem Grund werden hier noch drei weitere symmetrische Modellverteilungen diskutiert, die nicht das extreme Verhalten der Normalverteilung haben.

### 4.3.2. Alternativen zur Normalverteilung

Student-Verteilungen (student distributions) oder t-Verteilungen sind mit der Normalverteilung verwandt (siehe 4.4). Sie werden sehr selten als Modellverteilungen gebraucht, weil ihre mathematische Handhabung sehr schwierig ist. Zu jeder t-Verteilung gehört eine Anzahl Freiheitsgrade (für deren Bedeutung siehe Kap. 4.5.). Die t-Verteilungen streben mit steigender Anzahl Freiheitsgrade gegen die Normalverteilung.

Die Verteilung mit nur einem Freiheitsgrad heisst auch Cauchy-Verteilung. Sie kann als extremes Gegenstück zur Normalverteilung angesehen werden. Die Verteilung ist so breit, dass die Standardabweichung nicht definiert ist. (Sie ist unendlich.) Die Stichproben-Standardabweichung strebt mit zunehmendem Stichprobenumfang gegen unendlich. Die Standardabweichung von Stichproben-Mittelwerten ist unabhängig vom Stichprobenumfang. Das Vergrössern der Stichprobe erhöht also die Präzision nicht. Die Cauchy-Verteilung hat noch andere bizarre Eigenschaften. Messfehler sind nicht Cauchy-verteilt. Die Ungleichungen von Gauss und Bienaymé sind auf die Cauchy-Verteilung nicht anwendbar, da deren Varianz nicht definiert ist.

Die t-Verteilungen mit 3 und mehr Freiheitsgraden haben eine definierte Varianz. Daher könnten sie als Modellverteilungen verwendet werden. Hier werden die t-Verteilungen mit 3 und 5 Freiheitsgraden als Beispiel diskutiert.

Fig. 4.4 zeigt die Form der Verteilungen im Vergleich. Die Standardabweichungen der Verteilungen sind gleich. Die Standardabweichung der Cauchy-Verteilung ist nicht definiert, daher wird willkürlich eine Breite gewählt, die für die 80%-Schranke den gleichen Wert ergibt wie bei der Normalverteilung.

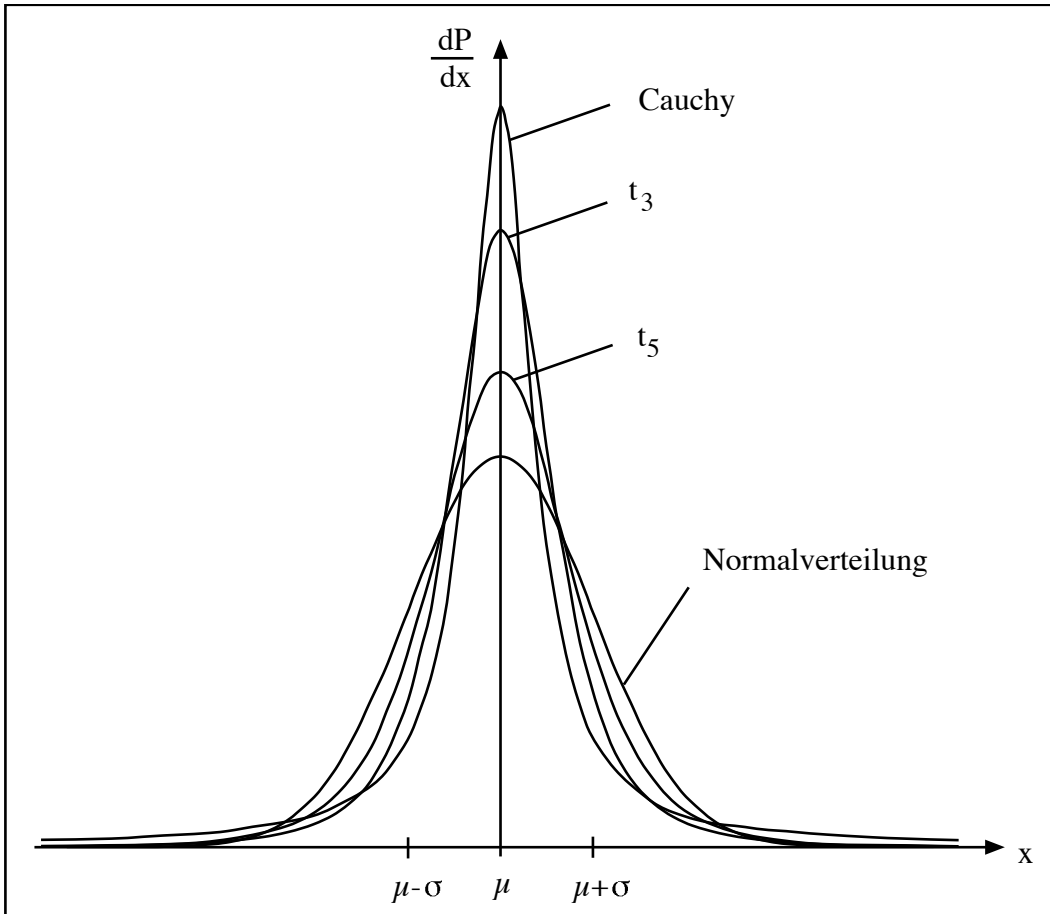


Fig. 4.4 Verschiedene symmetrische Modellverteilungen

Um ein Gefühl für den Zusammenhang zwischen der Grösse von Fehlerschranken, der statistischen Sicherheit und der zugrundeliegenden Verteilung zu vermitteln, sind in Tab. 4.1 einige Beispiele von Fehlerschranken für die Modellverteilungen in Fig. 4.4 zusammengestellt. Die Zahlen in den Reihen "Gauss" und "Bienaymé" sind die maximalen Fehlerschranken bei Anwendung der entsprechenden Ungleichung. Die Standardabweichung der Verteilungen (ausser der Cauchy-Verteilung) wird auf  $\sigma = 1$  festgelegt.

Verteilung	statistische Sicherheit						
	50 %	80 %	90 %	95 %	99 %	99,9 %	99,99 %
Normal	0,674	1,282	1,645	1,960	2,576	3,291	3,891
t5	0,563	1,143	1,561	1,991	3,123	5,321	8,658
t3	0,442	0,946	1,359	1,837	3,372	7,461	16,17
Gauss	0,943	1,490	2,108	2,981	6,666	21,08	66,66
Bienaymé	1,414	2,236	3,162	4,472	10,00	31,62	100,0
Cauchy	0,417	1,282	2,629	5,291	26,51	256,1	2651

Tab. 4.1 Halbe Fehlerbereiche für verschiedene Verteilungstypen. Siehe Text.

Es ist bemerkenswert, dass die Fehlerschranken im Bereich unterhalb von 90 % statistischer Sicherheit nicht allzu stark von der Annahme über die Form der Verteilung abhängen. Wer es sich leisten kann, wählt also kleine statistische Sicherheiten.

Das Ansinnen, eine Normalverteilung vorauszusetzen und dafür "sicherheitshalber" eine statistische Sicherheit von 99,9 % zu wählen, ist eine gefährliche Selbsttäuschung. Sollten die Messwerte in Wirklichkeit ungefähr t3-verteilt sein, entspricht die Fehlerschranke etwa einer Wahrscheinlichkeit

von 99 %. Man irrt sich also mit einer zehnmal so grossen Wahrscheinlichkeit als angenommen. Völlig unvernünftig wird es sein, eine statistische Sicherheit von 99,999 % zu wählen und eine Normalverteilung vorauszusetzen. Es gibt vermutlich kein Beispiel aus der analytischen Chemie, bei dem dieses Vorgehen auch nur annähernd sinnvolle Fehlerschranken ergeben würde.

Unter den Voraussetzungen einer eingipfeligen, symmetrischen Verteilung, der Abwesenheit von groben Fehlern, sowie gut bekannter Standardabweichung hat man bei der häufig verwendeten statistischen Sicherheit von 95 % die Gewähr, dass die Fehlerschranken der Normalverteilung mindestens etwa den halben möglichen Bereich abdecken.

### 4.3.3. Lognormalverteilung

Eine weitere häufig verwendete Modellverteilung ist die Lognormalverteilung. Sie heisst so, weil sie durch Logarithmieren in die Normalverteilung übergeht. Die Verteilung ist asymmetrisch (linksteil) und erstreckt sich nur über die positive Hälfte der Zahlenachse. Messwerte sind oft dann lognormalverteilt, wenn sie nicht negativ werden können und die Grösse der Fehler ähnlich ist wie die Messwerte selbst.

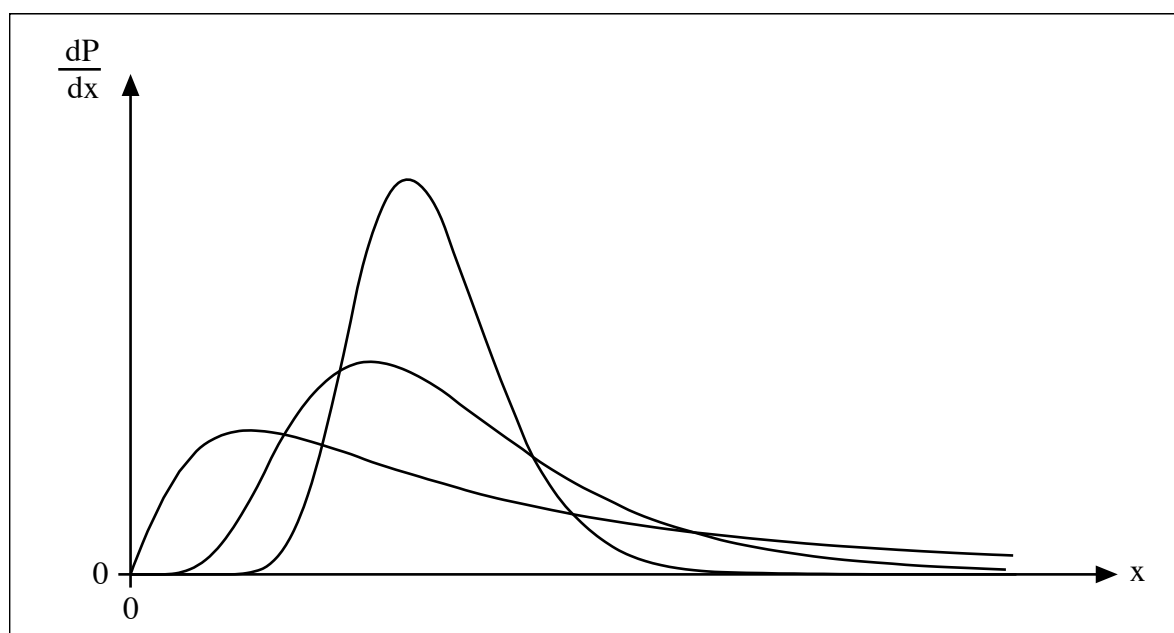


Fig. 4.5 Lognormalverteilungen

Durch Logarithmieren der Messwerte führt man lognormalverteilte Grössen in normalverteilte über. Man kann daraufhin mit der bequemen Normalverteilung weiterarbeiten und am Ende die berechneten Fehlerschranken zurücktransformieren.

## 4.4. Zentraler Grenzwertsatz

Messwerte sind gleichzeitig mit Fehlern aus vielen Quellen behaftet. Die Verteilungen der Einzelbeiträge überlagern sich zu einer Gesamtverteilung. Es lässt sich zeigen, dass bei einer Überlagerung von immer mehr Einzeleffekten die Verteilung der Summe gegen die Normalverteilung strebt. Das gilt für beliebige Einzelverteilungen. Das dieser Aussage zugrundeliegende Gesetz heisst Zentraler Grenzwertsatz (Central Limit Theorem). Dank dieses Prinzips darf man oft eine Normalverteilung für Messfehler voraussetzen.

## 4.5. Berechnung von Fehlerschranken aus geschätzten Verteilungen

Im Allgemeinen kennt man die Verteilungsfunktion der Messwerte nicht exakt. Es stellt sich also die Frage, wie man aus geschätzten Verteilungsfunktionen dennoch Fehlerschranken berechnen kann. In der ersten Hälfte des 20. Jahrhunderts wurden die Grundlagen gelegt, um dieses Problem für normalverteilte Grössen zu lösen. Für andere Verteilungen ist man auf sehr rechenintensive Monte-Carlo-Simulationen angewiesen, die erst seit kurzer Zeit möglich sind. Darauf kann hier nicht eingegangen werden.

Hier wird der einfachste Fall behandelt: Aus einer normalverteilten Grundgesamtheit wird eine Stichprobe vom Umfang  $n$  entnommen. Was für Fehlerschranken lassen sich für eine vorgegebene statistische Sicherheit angeben, und was ist deren Bedeutung?

Die Grösse  $\frac{x_i - \mu}{\sigma_x}$  streut gemäss einer Standardnormalverteilung (mit Mittelwert 0 und Standardabweichung 1). Ersetzt man die Parameter durch ihre Schätzungen, ist die resultierende Grösse nicht mehr normalverteilt. Der Brite William S. Gosset hat 1908 unter dem Pseudonym "Student" eine Herleitung für die Verteilung veröffentlicht. Die Verteilung heisst daher Student-Verteilung oder  $t$ -Verteilung.

$$\frac{x_i - \bar{x}}{s_x} \text{ ist verteilt gemäss } t_v \quad (4.10)$$

Für die formelmässige Beschreibung der Verteilung wird auf die Fachliteratur verwiesen. Eine Student-Verteilung hat nur einen Parameter, nämlich die Anzahl Freiheitsgrade  $v$ . Lässt man  $v$  gegen unendlich laufen, strebt die Verteilung gegen die Standardnormalverteilung. Die Verteilungen der

Grössen  $\frac{x_i - \mu}{\sigma_x}$  und  $\frac{x_i - \bar{x}}{s_x}$  sind in Fig 4.6 für 3 Freiheitsgrade dargestellt. Wie erwartet, streut die Grösse mit den geschätzten Parametern verbreitert, besonders in den Aussenbezirken der Verteilungen.

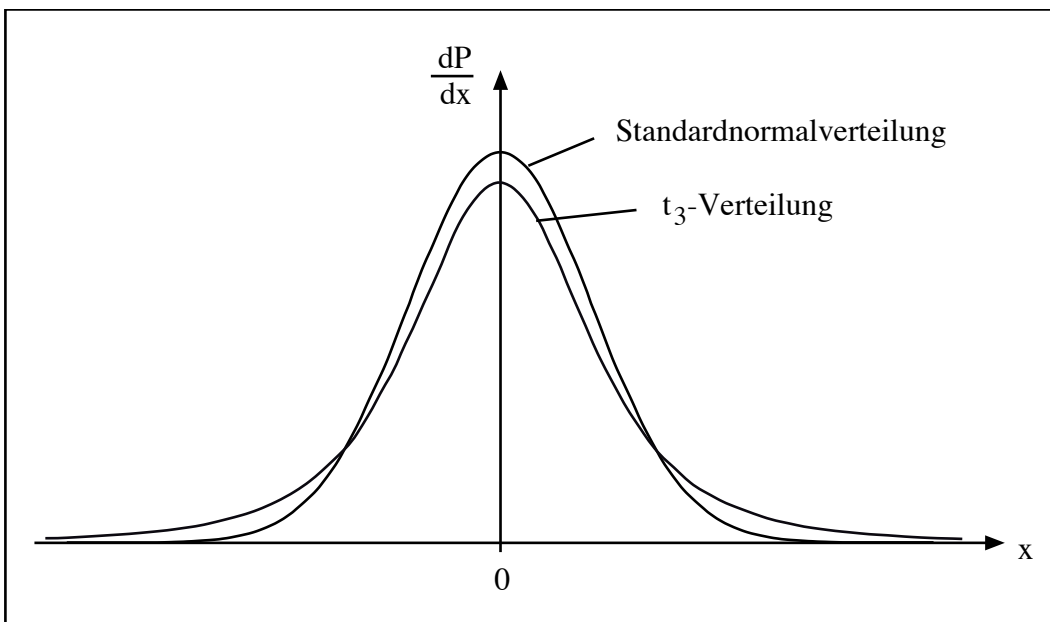


Fig. 4.6 Vergleich der Standardnormalverteilung mit der  $t$ -Verteilung



Es ist nun möglich, zwei Fehlerbereiche anzugeben, einen für die Einzelmessung und einen für den Mittelwert. Die Fehlerbereiche sind mit der statistischen Sicherheit  $P$  verknüpft. Der Fehlerbereich für die Einzelmessungen heisst Toleranzbereich  $TB_x$  (confidence interval of the individual measurement), der Fehlerbereich für den Mittelwert heisst Vertrauensbereich  $VB_{\bar{x}}$  (confidence interval of the mean).

$$TB_x = \bar{x} \pm t_{v,P} s_x \quad (4.11)$$

$$VB_{\bar{x}} = \bar{x} \pm t_{v,P} s_{\bar{x}} \quad (4.12)$$

Der Studentfaktor  $t_{v,P}$  wird aus der Student-Verteilung berechnet (siehe Fig. 4.7). Er ist umso kleiner, je kleiner die statistische Sicherheit  $P$  gewählt wird. Der Zusammenhang zwischen  $P$  und  $v$  ist zwar analytisch darstellbar, aber für eine praktische Rechnung ungeeignet. Man ist daher auf den Computer oder auf Tabellenwerke angewiesen. Der Studentfaktor ist zudem umso kleiner, je grösser die Anzahl Freiheitsgrade  $v$  ist, d. h. je zuverlässiger die Schätzung  $s_x$  wird. Hier ist

$$v = n - 1$$

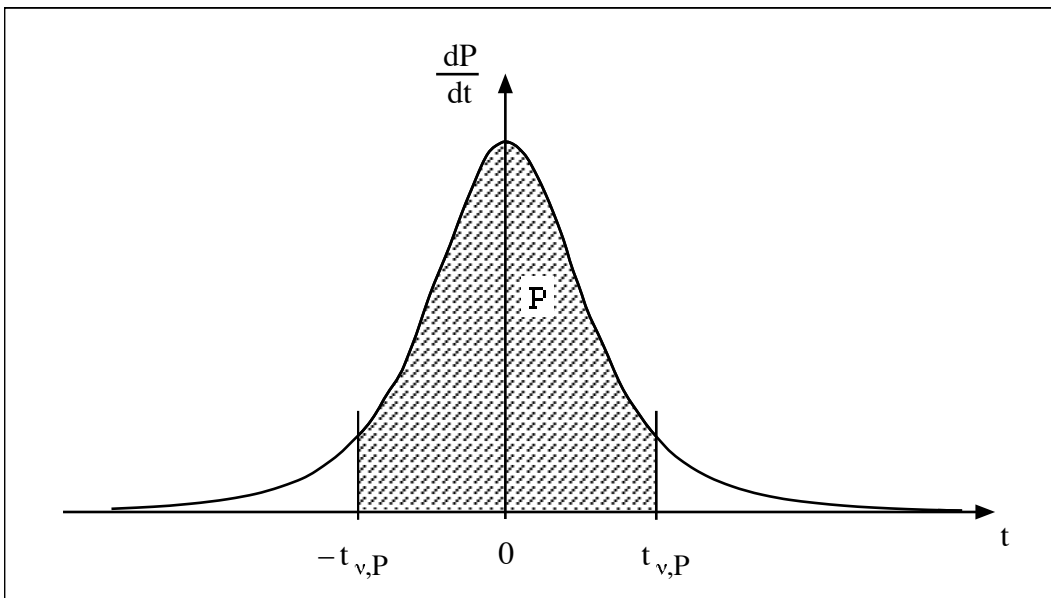


Fig. 4.7 Zusammenhang zwischen t-Faktor und statistischer Sicherheit  $P$

Die Mittelwerte zukünftiger Messreihen sind (bei konstantem Stichprobenumfang) auch wieder normalverteilt. Sind die Einzelwerte nicht normalverteilt, so sind es auch die Mittelwerte nicht, aber mit steigendem Stichprobenumfang nähert sich die Verteilung der Mittelwerte gemäss dem zentralen Grenzwertsatz trotzdem immer mehr einer Normalverteilung an, oftmals sogar mit schneller Konvergenz.

Die Bedeutung des Toleranz- und Vertrauensbereiches und der damit verknüpften statistischen Sicherheit  $P$  ist teilweise recht unanschaulich. Sie wird daher am Beispiel erläutert. Siehe Fig. 4.8.

Aus einer normalverteilten Grundgesamtheit mit den Parametern  $\mu = 0$  und  $\sigma = 1$  wird eine Stichprobe mit dem Umfang  $n = 6$  genommen:

-0,6    -0,3    0,3    0,9    1,2    1,5

Die Einzelwerte sind in Fig. 4.8 mit Pfeilen eingezeichnet. Man berechnet folgende Schätzungen:

$$\begin{aligned}\bar{x} &= 0,5 \\ s_x &= 0,841 \\ s_{\bar{x}} &= 0,344 \\ v &= 5 \quad t_5 (95 \%) = 2,571 \quad (\text{aus einer t-Tabelle}) \\ TB_{\bar{x}} &= 0,5 \pm 2,163 \quad [-1,663 ; 2,663] \\ VB_{\bar{x}} &= 0,5 \pm 0,883 \quad [-0,383 ; 1,383]\end{aligned}$$

Im oberen Teil der Fig. 4.8 ist die Verteilung der Einzelwerte eingezeichnet, die nach Voraussetzung die Parameter  $\mu = 0$  und  $\sigma_x = 1$  hat. Ebenfalls bekannt ist mithin die Verteilung der Stichprobenmittelwerte für den Stichprobenumfang 6, dargestellt im unteren Teil der Figur. Die Mittelwerte streuen ebenfalls um den Wert  $\mu = 0$ , haben aber eine kleinere Standardabweichung von

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{6}} = 0,408.$$

Die statistische Sicherheit  $P = 95 \%$  hat drei Bedeutungen:

1. Im Gegensatz zu diesem Beispiel kennt man im Allgemeinen den wahren Mittelwert  $\mu$  nicht. (Man bräuchte sonst keine Messungen mehr durchzuführen). Der wahre Mittelwert liegt mit einer Wahrscheinlichkeit von exakt 95 % innerhalb des Vertrauensbereiches. Der Vertrauensbereich eignet sich also für das Schlussresultat, bei dem eine Aussage über den wahren Mittelwert erwünscht ist. Diese Bedeutung der statistischen Sicherheit ist zweifellos die wichtigste und die am leichtesten zu verstehende.
2. Die beiden Schranken des 95%-Toleranzbereiches schneiden aus der Verteilung der Einzelwerte eine Fläche aus, die einer Wahrscheinlichkeit von 94,8 % entspricht. Mit dieser Wahrscheinlichkeit liegen zukünftige Messwerte innerhalb des Toleranzbereiches. Die Wahrscheinlichkeit ist also nicht exakt 95 %. Wie man aus der Figur leicht sieht, erwartet man das aber auch nicht. Die Wahrscheinlichkeit ist vielmehr eine Zufallsvariable. Sie liegt nur im Mittel vieler Stichproben bei 95 %. Der Toleranzbereich ist ein Charakteristikum für die Messmethode, weil er das Verhalten der einzelnen Messwerte beschreibt.
3. Nimmt man weitere Stichproben vom Umfang 6, so liegen deren Mittelwerte mit einer Wahrscheinlichkeit von 82,6 % zwischen den Schranken des 95%-Vertrauensbereiches. Die nominale Wahrscheinlichkeit von 95 %, die dem Bereich zugeordnet ist, stimmt also auch hier mit der wirklichen nicht überein.

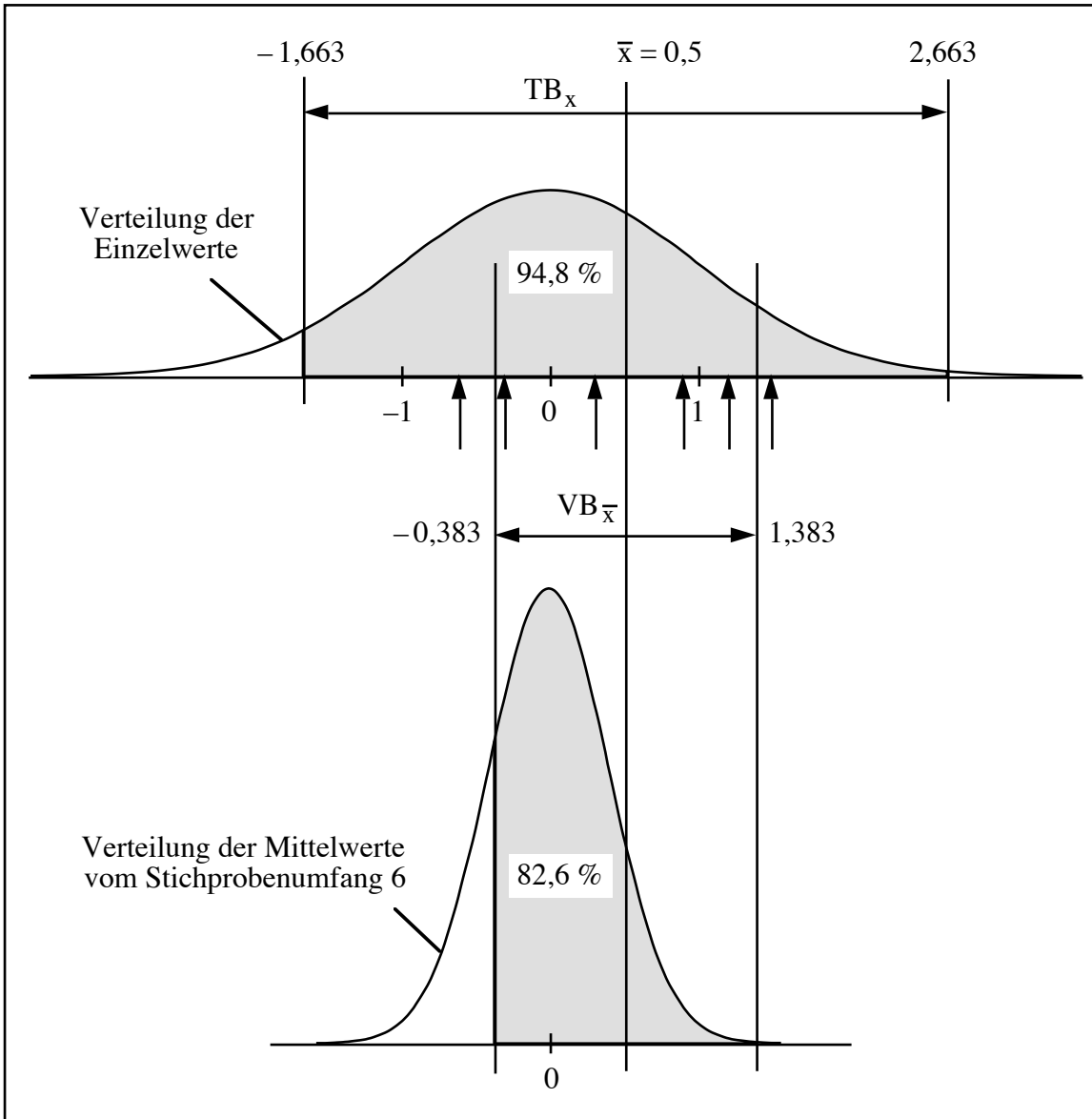


Fig. 4.8 Veranschaulichung der Fehlerschranken anhand eines Rechenbeispiels

## 4.6. Fehlerfortpflanzung

Nicht immer kann man jene Größen direkt messen, die einen letztlich interessieren. Oft muss man sie aus anderen, unmittelbar zugänglichen Größen berechnen. Damit auch für berechnete Werte Fehlerschranken angegeben werden können, muss man beschreiben, wie sich die Fehler der Ausgangsgrößen auf die berechneten Größen auswirken. Man nennt dieses Gebiet Fehlerfortpflanzung (error propagation).

Allgemein:

$$z = f(x_1, x_2, \dots, x_n) \quad (4.13)$$

$z$ : Zielgröße, die berechnet werden soll.

$f$ : beliebige bekannte Funktion

$x_1, x_2, \dots, x_n$ : fehlerbehaftete Größen

Von jedem  $x_i$  sei bekannt:

- $\hat{x}_i$  : Schätzwert für  $x_i$
- $s_i$  : geschätzte Standardabweichung von  $x_i$ , ein Mass für den zufälligen Fehler. Man berechnet sie mit einem statistischen Modell.
- $\Delta x_i$  : geschätzter systematischer Fehler von  $x_i$ . Diese Grösse ist schwer abzuschätzen. Man peilt sie über den Daumen oder setzt sie null, wenn einem nichts Besseres einfällt.

Als Schätzwert von  $z$  verwenden wir den Wert  $\hat{z} = f(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$

Zufällige und systematische Fehler pflanzen sich nach unterschiedlichen Gesetzen fort. Man behandelt sie daher getrennt und addiert am Schluss die beiden Beträge zum Fehler von  $z$ . Dabei tritt allerdings ein Problem auf. Auch wenn die zufälligen Fehler aller  $x_i$  gemäss einer Normalverteilung streuen, gilt dies im Allgemeinen nicht mehr für  $z$ . Zudem kann es vorkommen, dass die massgebende Anzahl Freiheitsgrade  $v_i$  für jedes  $s_i$  eine andere ist. Es gibt keine mathematisch einfache Lösung für dieses Problem. Unter Verlust von Information kann man die kleinste vorkommende Anzahl Freiheitsgrade wählen. Man kann den Verlust in Grenzen halten, wenn man sich mit einer kleinen statistischen Sicherheit begnügt.

Als Schlussresultat eignet sich

$$VB_{\hat{z}} = \hat{z} \pm (s_z t + \Delta z) \quad (4.14)$$

$s_z$  und  $\Delta z$  berechnet man mit Hilfe der nachfolgenden Verfahren.

#### 4.6.1. Fortpflanzung systematischer Fehler

Allgemein: Taylor-Reihe (nur lineare Terme):

$$\Delta z = \frac{\partial f}{\partial x_1} \Delta x_1 + \frac{\partial f}{\partial x_2} \Delta x_2 + \dots \quad (4.15)$$

Meist kennt man das Vorzeichen eines systematischen Fehlers nicht. In so einem Fall ersetzt man  $\frac{\partial f}{\partial x_i} \Delta x_i$  durch  $\left| \frac{\partial f}{\partial x_i} \Delta x_i \right|$ .

Beispiele:

- Addition:  $z = x_1 + x_2$

$$\frac{\partial f}{\partial x_1} = 1 = \frac{\partial f}{\partial x_2}$$

$$\Delta z = |1 \cdot \Delta x_1| + |1 \cdot \Delta x_2| = |\Delta x_1| + |\Delta x_2|$$

Man addiert die Fehler.

- Multiplikation:  $z = x_1 \cdot x_2$

$$\frac{\partial f}{\partial x_1} = x_2 \quad \frac{\partial f}{\partial x_2} = x_1$$

$$\Delta z = |x_2 \Delta x_1| + |x_1 \Delta x_2|$$

Nach Division durch  $z = x_1 \cdot x_2$ :

$$\frac{\Delta z}{z} = \left| \frac{\Delta x_1}{x_1} \right| + \left| \frac{\Delta x_2}{x_2} \right|$$

Man addiert die relativen Fehler.

- Division:  $z = \frac{x_1}{x_2}$

$$\frac{\partial f}{\partial x_1} = \frac{1}{x_2} \quad \frac{\partial f}{\partial x_2} = -\frac{x_1}{x_2^2}$$

$$\Delta z = \left| \frac{1}{x_2} \Delta x_1 \right| + \left| \frac{x_1}{x_2^2} \Delta x_2 \right|$$

Nach Division durch  $z = \frac{x_1}{x_2}$ :

$$\frac{\Delta z}{z} = \left| \frac{\Delta x_1}{x_1} \right| + \left| \frac{\Delta x_2}{x_2} \right|$$

Man addiert die relativen Fehler wie bei der Multiplikation.

- Potenz:  $z = x^a$     a: exakt

$$\frac{\partial f}{\partial x} = a x^{a-1}$$

$$\Delta z = |a x^{a-1} \Delta x|$$

Nach Division durch  $z = x^a$ :

$$\frac{\Delta z}{z} = |a x^{a-1} \Delta x x^{-a}| = \left| a \frac{\Delta x}{x} \right|$$

Man multipliziert den relativen Fehler mit dem Betrag des Exponenten.

#### 4.6.2. Fortpflanzung zufälliger Fehler

Allgemein: Taylor-Reihe:

$$s_z^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 s_1^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 s_2^2 + \dots + 2 \left(\frac{\partial f}{\partial x_1}\right) \left(\frac{\partial f}{\partial x_2}\right) \rho_{1,2} s_1 s_2 + \dots \quad (4.16)$$

Von der Reihe werden nur die Terme linear in  $s^2$  und die paarweisen Kreuzterme berücksichtigt. Falls die  $x_i$  nicht korreliert sind, also die Korrelationskoeffizienten  $\rho_{i,j} = 0$  sind, vereinfacht sich die Reihe erheblich. Schätzungen für Parameter, die mit der gleichen Stichprobe bestimmt werden, sind im Allgemeinen korreliert. Die Korrelationskoeffizienten können meist aus dem Modell berechnet werden. Schätzungen aus unterschiedlichen Stichproben sind grundsätzlich unkorreliert. Der Faktor 2 vor den Kreuztermen berücksichtigt die Symmetrie  $\rho_{i,j} = \rho_{j,i}$ . Es sind also nur Terme für  $i < j$  zu berücksichtigen.

Die Grösse  $\rho_{1,2} s_1 s_2$  wird auch als Kovarianz (covariance) bezeichnet. Leider hat sich kein Symbol für die Grösse eingebürgert. Man bezeichnet die Kovarianz zwischen a und b mit  $\text{Cov}(a,b)$ , wobei unklar bleibt, ob es sich um den Parameter oder seine Schätzung handelt.

Oft geht man von einer Matrixdarstellung aus, die im Umgang mit Computerprogrammen besonders bequem ist. Sei  $\mathbf{R}$  die symmetrische Matrix der Korrelationskoeffizienten und  $\mathbf{S}$  die Diagonalmatrix der Standardabweichungen:

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{12} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \dots & 1 \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_p \end{pmatrix} \quad (4.17)$$

Man erhält daraus die symmetrische Varianz-Kovarianz-Matrix  $\mathbf{VC}$ :

$$\mathbf{VC} = \mathbf{S R S} \quad (4.18)$$

(In Kapitel 6 wird  $\mathbf{VC}$  durch die Matrix  $s^2 (\mathbf{P}^T \mathbf{P})^{-1}$  dargestellt.)

Definiert man noch den Vektor  $\mathbf{d}$  der Ableitungen:

$$\mathbf{d} = \left( \frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \dots \quad \frac{\partial f}{\partial x_p} \right)^T \quad (4.19)$$

so erhält man analog Gleichung (4.16):

$$s_z^2 = \mathbf{d}^T \mathbf{VC} \mathbf{d} \quad (4.20)$$

Beispiele (Anwendung von (4.16)):

- Addition:  $z = x_1 + x_2$

$$\frac{\partial f}{\partial x_1} = 1 = \frac{\partial f}{\partial x_2}$$

$$s_z^2 = (1)^2 s_1^2 + (1)^2 s_2^2 + 2(1)(1) \rho_{1,2} s_1 s_2$$

Wenn  $\rho_{1,2} = 0$  ist, addiert man die Varianzen.

- Multiplikation:  $z = x_1 x_2$

$$\frac{\partial f}{\partial x_1} = x_2 \quad \frac{\partial f}{\partial x_2} = x_1$$

$$s_z^2 = (x_2)^2 s_1^2 + (x_1)^2 s_2^2 + 2(x_1)(x_2) \rho_{1,2} s_1 s_2$$

Nach Division durch  $z^2 = x_1^2 x_2^2$ :

$$\left(\frac{s_z}{z}\right)^2 = \left(\frac{s_1}{x_1}\right)^2 + \left(\frac{s_2}{x_2}\right)^2 + 2 \rho_{1,2} \frac{s_1}{x_1} \frac{s_2}{x_2}$$

Wenn  $\rho_{1,2} = 0$  ist, addiert man die relativen Varianzen.

- Division:  $z = \frac{x_1}{x_2}$

$$\frac{\partial f}{\partial x_1} = \frac{1}{x_2} \quad \frac{\partial f}{\partial x_2} = -\frac{x_1}{x_2^2}$$

$$s_z^2 = \left(\frac{1}{x_2}\right)^2 s_1^2 + \left(\frac{-x_1}{x_2^2}\right)^2 s_2^2 + 2 \frac{1}{x_2} \frac{-x_1}{x_2^2} \rho_{1,2} s_1 s_2$$

Nach Division durch  $z^2 = \frac{x_1^2}{x_2^2}$ :

$$\left(\frac{s_z}{z}\right)^2 = \left(\frac{s_1}{x_1}\right)^2 + \left(\frac{s_2}{x_2}\right)^2 - 2 \rho_{1,2} \frac{s_1}{x_1} \frac{s_2}{x_2}$$

Wenn  $\rho_{1,2} = 0$  ist, addiert man die relativen Varianzen wie bei der Multiplikation.

- Potenz:  $z = x^a$      $a$ : exakt

$$\frac{\partial f}{\partial x} = a x^{a-1}$$

$$s_z^2 = (a x^{a-1} s_x)^2$$

Nach Division durch  $z^2 = x^{2a}$  :

$$\left(\frac{s_z}{z}\right)^2 = (a x^{a-1} s_x x^{-a})^2 = a^2 \left(\frac{s_x}{x}\right)^2$$

$$\frac{s_z}{z} = \left| a \frac{s_x}{x} \right|$$

Man multipliziert die relative Standardabweichung mit dem Betrag des Exponenten und nicht etwa die relative Varianz, wie man aus Analogieüberlegungen schliessen könnte. Zufällige Fehler pflanzen sich hier wie systematische fort. Eine Potenz darf nicht wie die Multiplikation zweier unkorrelierter Faktoren behandelt werden, da die Faktoren bei einer Potenz vollständig korreliert sind.



## 5. Grundlagen statistischer Tests

Mit einem statistischen Test will man eine unbekannte Grundgesamtheit, der man eine Stichprobe entnommen hat, mit einer bekannten oder hypothetischen Grundgesamtheit vergleichen. Man kann mit einem Test nur Unterschiede, nicht aber Übereinstimmungen feststellen. Die Hypothese, dass zwei Grundgesamtheiten übereinstimmen, nennt man Nullhypothese  $H_0$  (null hypothesis). Die Schätzung eines Parameters wird natürlich mit dem hypothetischen Wert nie übereinstimmen. Man kann aber mit Hilfe eines Modells berechnen, mit welcher Wahrscheinlichkeit die beobachtete Differenz rein zufällig auftritt, wenn man die Gültigkeit der Nullhypothese voraussetzt. Man verwirft die Nullhypothese, wenn diese Wahrscheinlichkeit kleiner ist als eine (im voraus) willkürlich festgelegte Schranke.

Mit anderen Worten: Durch das Einführen einer zusätzlichen Voraussetzung (der Nullhypothese) gelangt man zu Beziehungen, mit deren Hilfe man die Wahrscheinlichkeit für das Eintreten bereits beobachteter Vorgänge berechnen kann. Wenn diese Wahrscheinlichkeit zu klein erscheint, z. B. unterhalb 5 % liegt, vermutet man einen Widerspruch. Es widerstrebt einem zu glauben, dass ein bereits beobachteter Vorgang mit einer ehemals so kleinen Wahrscheinlichkeit eingetreten sein sollte. Man verwirft dann also die Nullhypothese, damit man die Seltsamkeiten nicht glauben muss, die aus ihr folgen.

Die Nullhypothese allein ist noch nicht ausreichend zur Beschreibung eines Tests. Es muss noch festgelegt werden, welche Alternativhypothese  $H_A$  (alternative hypothesis) man akzeptiert, wenn die Nullhypothese verworfen wird. Das erscheint auf den ersten Blick unnötig. Wenn z. B. als Nullhypothese  $H_0: \mu = 0$  festgelegt wird, dann scheint die Alternativhypothese  $H_A: \mu \neq 0$  die einzig vernünftige zu sein. Es könnte aber sein, dass einen der Fall  $\mu < 0$  nicht interessiert. In diesem Fall wählt man als Alternative  $H_A: \mu > 0$ . Dies wird als einseitiger Test (one-tailed test) bezeichnet. Das Gegenstück heisst entsprechend zweiseitiger Test (two-tailed test).

Müsste man dann aber nicht als Nullhypothese  $H_0: \mu \leq 0$  wählen? Man könnte das tun, hätte aber gewiss grosse Schwierigkeiten, aus der Ungleichung brauchbare Beziehungen herzuleiten, die schliesslich zu einer Wahrscheinlichkeitsaussage führen. Nur das Gleichheitszeichen ist eine genügend kraftvolle Zusatzaussage.

Weiter ins Detail führende Überlegungen zum Thema Alternativhypothese werden am besten anhand von Beispielen durchgeführt. Siehe später.

### 5.1. Risiko falscher Entscheide

Bei Entscheidungen, die aufgrund statistischer Tests gefällt werden, können zwei Fehler auftreten:

Fehler 1. Art (type I error):

Man verwirft die Nullhypothese fälschlicherweise. Man glaubt also, einen Unterschied zwischen den Grundgesamtheiten gefunden zu haben, in Wirklichkeit gibt es aber keinen. Die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, heisst  $\alpha$ .

Fehler 2. Art (type II error):

Man akzeptiert die Nullhypothese fälschlicherweise. Man findet also keinen Unterschied zwischen den Grundgesamtheiten, in Wirklichkeit gibt es aber einen. Die Wahrscheinlichkeit, einen Fehler 2. Art zu begehen, heisst  $\beta$ .

Die statistische Sicherheit  $1 - \alpha$  kann man frei wählen. Dies ist die Wahrscheinlichkeit, mit der eine wahre Nullhypothese auch tatsächlich beibehalten wird.

Die **Macht**  $1 - \beta$  (power) eines Tests ist jene Wahrscheinlichkeit, mit der eine falsche Nullhypothese auch tatsächlich abgelehnt wird. Die Macht hängt zwar eng mit der statistischen Sicherheit zusammen, es existiert aber keine einfache Beziehung zwischen den beiden. Die Macht eines Tests lässt sich nur für feste Alternativhypothesen berechnen. Sie ist umso grösser, je grösser der Stichprobenumfang ist und je kleiner die statistische Sicherheit gewählt wird.

Man unterscheidet zwischen den beiden Verhaltensweisen des Kritikers und des Entdeckers:

- Dem Kritiker erscheint das unberechtigte Verwerfen der Nullhypothese gefährlich. Er fordert also eine hohe statistische Sicherheit und nimmt in Kauf, ab und zu einen Unterschied nicht zu entdecken. Juristen sind typische Kritiker. Sie fordern 100 % (statistische) Sicherheit.
- Der Entdecker tendiert dazu, die Nullhypothese zu verwerfen. Er begnügt sich mit einer kleinen statistischen Sicherheit, auch wenn er dadurch öfter einen Unterschied zu finden glaubt, wo keiner ist. Erich von Däniken als grösster aller Entdecker wählt jeweils 100 % Macht, auch wenn ihm dabei die statistische Sicherheit gänzlich abhanden kommt.

Das Verhalten ist Bestandteil der Entscheidungsstrategie. Vgl. Kapitel 5.4.

## 5.2. Beispiel eines zweiseitigen statistischen Tests

Im Rahmen der Qualitätskontrolle wird in einem Vitaminpräparat die Konzentration von Vitamin A gemessen und mit einem Sollwert  $c_{\text{soll}} = 100 \text{ mg/kg}$  verglichen. Die verwendete Messmethode ist schon so lange im Einsatz, dass die Verteilung der Messwerte sehr genau bekannt ist. Man weiss, dass man eine Normalverteilung annehmen darf. Zudem kennt man die Standardabweichung der Einzelwerte  $\sigma_c = 1 \text{ mg/kg}$ . Die wirkliche Konzentration  $c_{\text{real}}$  sei  $101 \text{ mg/kg}$ , was der Experimentator natürlich nicht weiss. Vielmehr besteht Grund zur Annahme, es sei  $c_{\text{real}} = c_{\text{soll}}$ . Die Konzentration wird ein einziges Mal gemessen. Wird es dem Experimentator gelingen, den kleinen Unterschied festzustellen?

Die Nullhypothese ist aus der Aufgabestellung gegeben:  $c_{\text{real}} = c_{\text{soll}}$ , oder

$$H_0 : c_{\text{real}} - c_{\text{soll}} = 0$$

Sowohl zu kleine als auch zu grosse Vitamin-Konzentrationen sollen erkannt werden. Die Alternativhypothese ist also

$$H_A : c_{\text{real}} - c_{\text{soll}} \neq 0$$

Die statistische Sicherheit  $1 - \alpha$  wird auf 95 % festgelegt. Damit ergibt sich eine Situation, die in Figur 5.1 graphisch dargestellt ist.

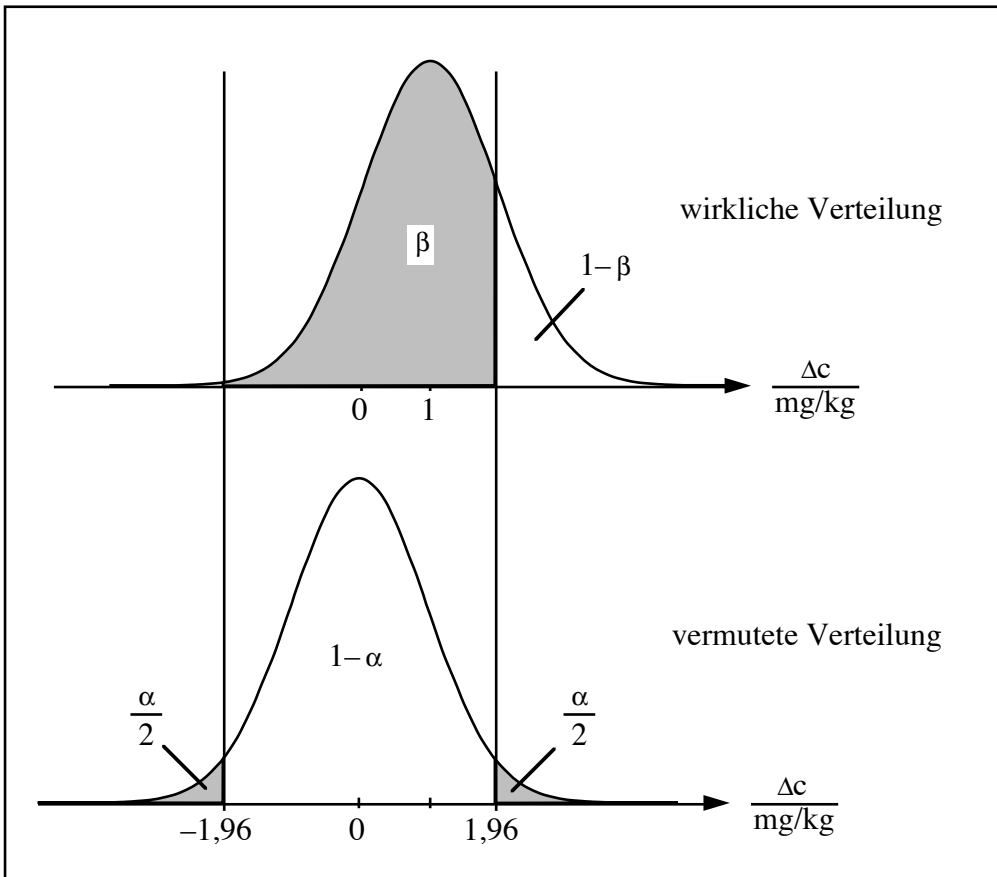


Fig. 5.1 Beispiel eines zweiseitigen statistischen Tests

Da die Mittelwerte der Verteilungen in Wirklichkeit nicht gleich sind, wird der Experimentator entweder die Nullhypothese richtigweise ablehnen, oder einen Fehler 2. Art machen.

Das Verhalten des Experimentators ersieht man aus dem unteren Teil der Figur 5.1. Er kann vom oberen Teil nichts wissen. Er akzeptiert die Nullhypothese, wenn der Messwert zwischen die beiden vertikalen Linien zu liegen kommt. Wenn die Nullhypothese richtig wäre, läge dieser Fall mit einer Wahrscheinlichkeit von 95 % vor.

Was dieses Entscheidungsverhalten bewirkt, ersieht man aus dem oberen Teil der Figur. Mit einer Wahrscheinlichkeit von etwa 83 % fällt der Messwert zwischen die vertikalen Linien. Dies ist die Wahrscheinlichkeit, einen Fehler 2. Art zu machen, also  $\beta$ . Die Macht des Tests  $1 - \beta$ , also die Wahrscheinlichkeit, die falsche Nullhypothese auch wirklich zu verwerfen, beträgt demnach 17 %.

Vermutlich wird der Unterschied also nicht erkannt. Die Differenz der Mittelwerte ist zu klein, gemessen an der Breite der Verteilung.

Der Experimentator kann die Macht des Tests erhöhen, indem er sich mit einer kleineren statistischen Sicherheit begnügt. Dadurch rücken die vertikalen Linien näher zusammen. Testet er auf dem 80%-Niveau, steigt die Macht des Tests auf etwa 40 %. Dies ist eine eher bescheidene Verbesserung angesichts der damit verbundenen Verringerung der statistischen Sicherheit. Nur ein sehr progressiver Entdecker würde das Risiko tragen wollen.

Eine kraftvollere, aber aufwendigere Methode, die Macht des Tests zu erhöhen, besteht im Vergrößern des Stichprobenumfangs. Die Mittelwerte von 4 Messwerten streuen gemäss Gleichung (4.6) nur noch mit der halben Standardabweichung. Die Breite der Verteilungen in Figur 5.1 wird dadurch halbiert, die Mittelwerte bleiben hingegen am selben Ort. Damit kann man den Unterschied der Mittelwerte mit einer Wahrscheinlichkeit von etwa 52 % entdecken.

Eine kritische Betrachtung dieses Beispiels lässt einige Unschönheiten erkennen. Man weiss mit Sicherheit, dass die wirkliche Konzentration mit dem Sollwert nicht übereinstimmt. Die exakte Übereinstimmung eines (ungerundeten) Messwerts mit einem Sollwert ist schlechterdings unmöglich. Man müsste also die Nullhypothese in jedem Fall verwerfen, eine Messung könnte man sich ersparen. Dies zeigt, dass die Nullhypothese nicht vernünftig formuliert ist. Alarm sollte erst geschlagen werden, wenn die Abweichung zwischen wirklicher Konzentration und Sollwert eine bestimmte Schranke überschreitet. Also sollte man Hypothesen dieser Art formulieren:

$$H_0 : |c_{\text{real}} - c_{\text{soll}}| \leq \Delta c_{\text{max}}$$

$$H_A : |c_{\text{real}} - c_{\text{soll}}| > \Delta c_{\text{max}}$$

Wie schon erwähnt, eignet sich aber eine Ungleichung nicht als Nullhypothese. Wie man in so einer Situation vorgeht, zeigt das nächste Beispiel, das einen etwas anders gelagerten Fall behandelt.

### 5.3. Beispiel eines einseitigen statistischen Tests

Die maximal erlaubte Konzentration der Dithiocarbamat-Fungizide in Kopfsalat ist gemäss Schweizerischer Lebensmittelverordnung auf  $c_{\text{max}} = 2 \text{ mg/kg}$  (als  $\text{CS}_2$ ) festgesetzt. Die in der amtlichen Sammlung vorgeschriebene Bestimmungsmethode ist sehr gut bekannt. Man weiss, dass man eine Normalverteilung der Messwerte annehmen darf. Zudem kennt man die Standardabweichung der Einzelwerte  $\sigma = 0,1 \text{ mg/kg}$  so genau, dass man sie als exakt bekannt annehmen kann. In einem gegebenen Fall sei die Konzentration  $2,3 \text{ mg/kg}$ , was der Experimentator natürlich nicht weiss. Die Konzentration wird ein einziges Mal gemessen. Die statistische Sicherheit wird auf 99 % festgelegt, da der Experimentator die Position des Kritikers einnimmt. Er will also nur dann Alarm schlagen, wenn er sich seiner Sache wirklich sicher ist. Wie ist der statistische Test zu formulieren?

Alarm ist zu schlagen, wenn  $c_{\text{real}} \geq c_{\text{max}}$  ist. Da die Ungleichung nicht als Nullhypothese brauchbar ist, betrachten wir den schlimmstmöglichen Fall. Eine Überschreitung des gesetzlichen Höchstwertes wird man umso besser nachweisen können, je grösser die Überschreitung ist. Am meisten Schwierigkeiten werden sich einstellen, wenn die wirkliche Konzentration gerade den Höchstwert erreicht. Diese Bedingung wird als Nullhypothese verwendet:

$$H_0 : c_{\text{real}} - c_{\text{max}} = 0$$

Niemand glaubt wirklich, dass die Nullhypothese richtig ist. Messwerte erfüllen prinzipiell keine Gleichungen dieser Art. Es handelt sich um einen hypothetischen Grenzfall.

Mit dem Test will man eine Überschreitung erkennen. Die Alternativhypothese lautet demnach:

$$H_A : c_{\text{real}} - c_{\text{max}} > 0$$

Dass die Nullhypothese auch dann falsch ist, wenn der Höchstwert unterschritten wird, interessiert hier nicht. Einen Test mit dieser Eigenheit nennt man einseitig.

Da der gesetzliche Höchstwert im gegebenen Fall tatsächlich überschritten ist, wird der Experimentator entweder die Nullhypothese richtigerweise ablehnen, oder einen Fehler 2. Art begehen.

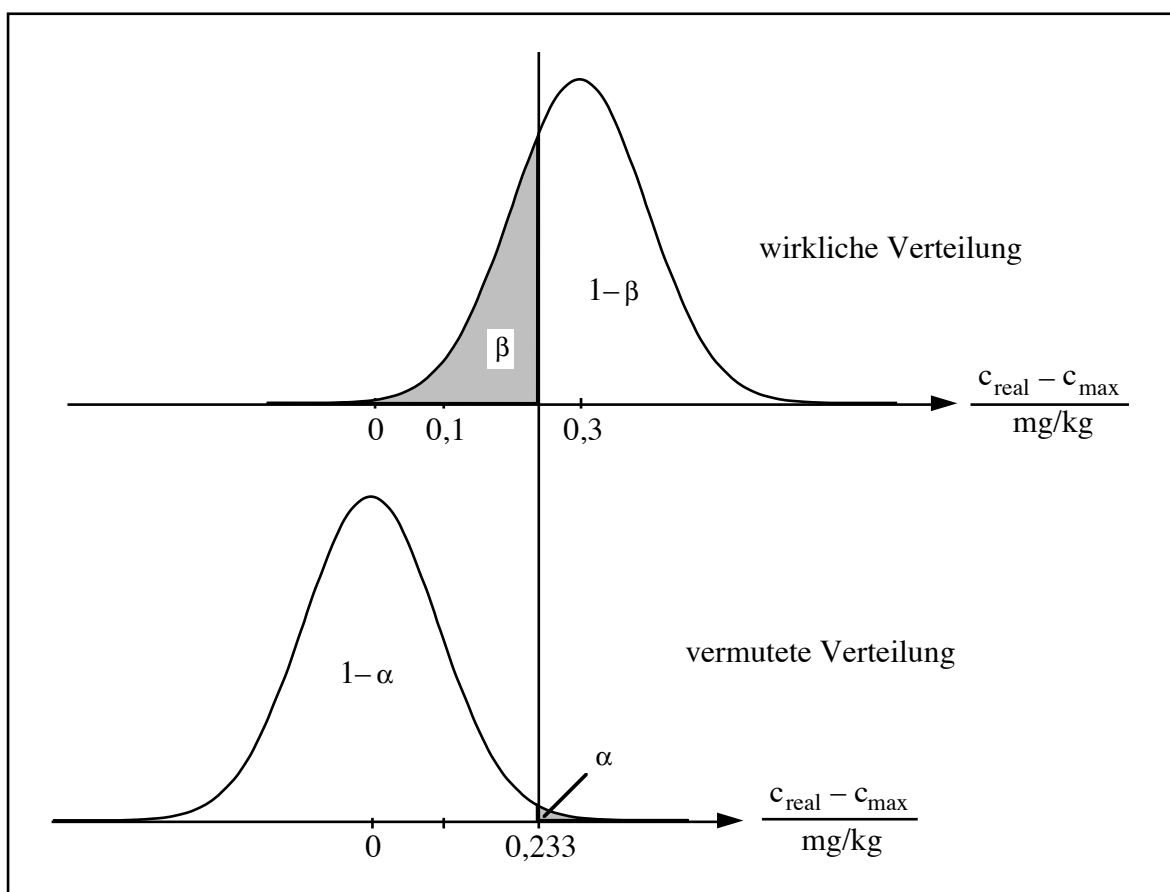


Fig. 5.2 Beispiel eines einseitigen statistischen Tests

Die Situation ist in Fig. 5.2 graphisch dargestellt. Aufgetragen ist die Überschreitung des gesetzlichen Höchstwertes  $c_{\text{real}} - c_{\text{max}}$ . Das Verhalten des Experimentators ist aus dem unteren Teil der Figur ersichtlich. Über den oberen Teil kann er nichts wissen. Er akzeptiert die Alternativhypothese, wenn der Messwert rechts von der vertikalen Linie zu liegen kommt. Die Lage der Vertikalen ergibt sich aus der Strategie des Kritikers. Der gefährliche Fehler besteht im versehentlichen Erkennen einer Überschreitung, die nicht vorhanden ist. Dieser Fall soll nur mit 1 % Wahrscheinlichkeit auftreten. Der Wert von 0,233 mg/kg ergibt sich aus der Forderung nach 99 % statistischer Sicherheit. Einer Tabelle der standardisierten Normalverteilung (mit  $\mu = 0$  und  $\sigma = 1$ ) entnimmt man für  $\alpha = 1 \%$  den Wert 2,33. Damit ergibt sich für eine Verteilung mit  $\sigma = 0,1$  mg/kg der Wert 0,233 mg/kg als 99%-Schranke.

Wenn die Nullhypothese richtig wäre, käme der Messwert mit einer Wahrscheinlichkeit von 1 % rechts der vertikalen Linie zu liegen. Wenn der Höchstwert in Wirklichkeit unterschritten wäre, läge diese Wahrscheinlichkeit tiefer. Der Experimentator irrt sich bei diesem Vorgehen also mit einer Wahrscheinlichkeit von höchstens 1 %, wenn er die Nullhypothese akzeptiert.

Was dieses Entscheidungsverhalten bewirkt, ersieht man aus dem oberen Teil der Figur. Mit einer Wahrscheinlichkeit von etwa 25 % fällt der Messwert auf die linke Seite der vertikalen Linie. Dies ist die Wahrscheinlichkeit, einen Fehler 2. Art zu begehen, also  $\beta$ . Die Macht des Tests  $1 - \beta$ , also die Wahrscheinlichkeit, die richtige Alternativhypothese auch wirklich zu akzeptieren, beträgt demnach 75 %.

Die vorhergehenden Beispiele statistischer Tests dienen der Erläuterung des Vorgehens sowie der Veranschaulichung einiger Begriffe. Es handelt sich um besonders anschauliche Beispiele, da die Verteilung der Messwerte selbst unmittelbar für die Tests verwendet werden konnte. Im Allgemeinen ist das nicht der Fall. Was tut man etwa, wenn die Standardabweichung der Messwerte in den vorangehenden Beispielen nicht bekannt ist?

Eine Möglichkeit besteht darin, die Standardabweichung durch  $n$  Wiederholmessungen zu schätzen. Dadurch kann man den Mittelwert  $\bar{x}$ , die Standardabweichung der Einzelwerte  $s_x$  und des Mittelwerts  $s_{\bar{x}}$  schätzen. Unter der Voraussetzung, dass die Grösse  $\frac{x_i - \mu}{\sigma_x}$  gemäss einer Standardnormalverteilung streut, kennt man auch die Verteilung der Grösse  $\frac{\bar{x} - \mu}{s_{\bar{x}}}$ . Sie ist nämlich eine t-Verteilung mit  $n-1$  Freiheitsgraden. Damit kennt man wieder eine exakte Verteilung, mit der man genauso verfährt, wie in den vorangehenden Beispielen gezeigt.

Ganz allgemein leitet man aus der Nullhypothese Zusatzbeziehungen her. In diese Formeln setzt man die Messwerte ein und berechnet damit eine Kenngrösse  $x$  oder Teststatistik (test statistic). Sie ist eine Zufallsvariable, hat also eine Verteilung. Die zugehörige Wahrscheinlichkeit ist das bestimmte Integral unter der Verteilungskurve, wobei die Teststatistik eine der beiden Integralgrenzen bildet. Die andere Grenze ist meist 0 oder  $-\infty$ . Siehe Fig. 5.3 für ein Beispiel.

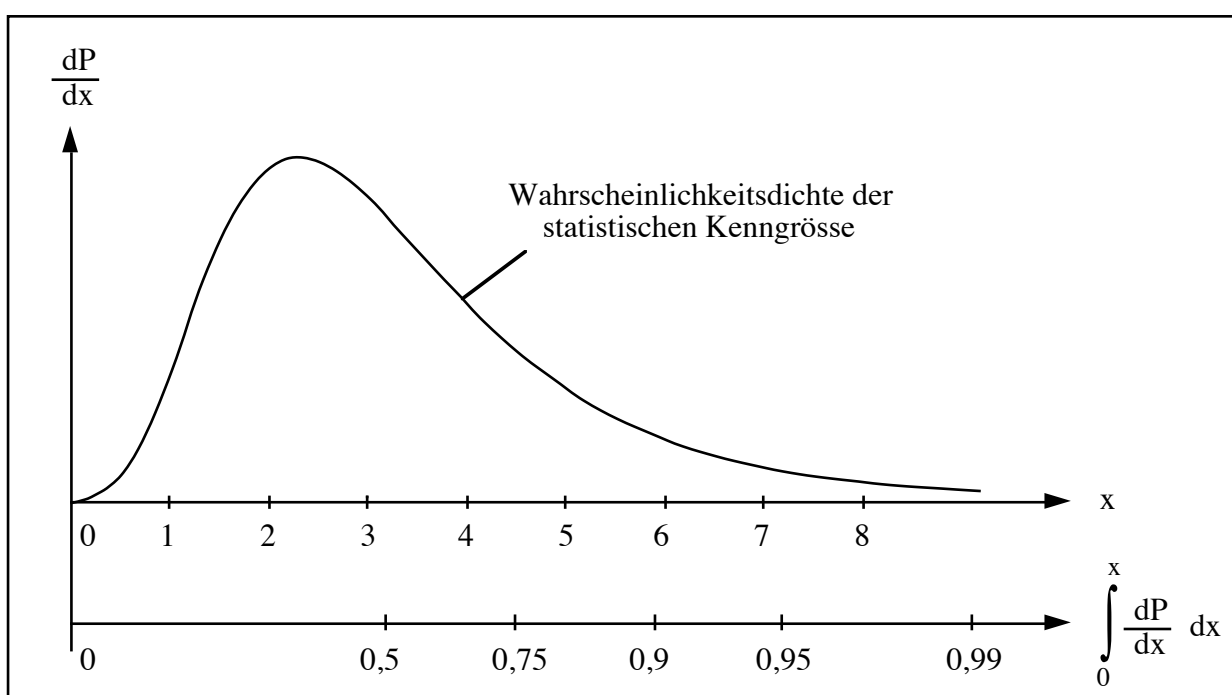


Fig. 5.3 Zusammenhang zwischen Teststatistik und Wahrscheinlichkeit

Der Zusammenhang zwischen Teststatistik und zugehöriger Wahrscheinlichkeit ist meist sehr kompliziert. Traditionsgemäss gehört daher zu einem Test eine Tabelle, die den Zusammenhang wiedergibt. Die statistischen Modelle, die heute in der analytischen Chemie am häufigsten angewendet werden, wurden zu Beginn dieses Jahrhunderts entwickelt. Zu jener Zeit war das Berechnen numerisch komplexer Funktionen schwierig. Das Bereitstellen einer Tabelle machte die Berechnungen überflüssig. Umfangreiche Tabellen galten früher als Zeichen besonderer Qualität. Heute sind sie ein Ärgernis, da die Handhabung von Tabellen in einem Computer mühsam ist. Wesentlich bequemer ist ein Algorithmus zur Berechnung der Wahrscheinlichkeit aus der Teststatistik und umgekehrt. Für viele statistische Tests ist der Algorithmus leicht zu programmieren. (Siehe Literaturverzeichnis, Press et al., Numerical Recipes. Es handelt sich bei diesem Werk um eines der besten Bücher über Numerik.)

Die Skala der statistischen Sicherheit ist wesentlich aussagekräftiger als jene der Teststatistik (siehe untere Skala in Fig. 5.3). Es ist also vernünftig, Werte der Teststatistik auf die Skala der Wahrscheinlichkeit umzurechnen. Mit einer Tabelle ist das meist nur durch Interpolieren möglich. Daher wird traditionell der umgekehrte Weg beschriftet.

Das "historische" Vorgehen beim Durchführen eines Tests:

1. Statistische Sicherheit festlegen.
2. Mithilfe der Tabelle aus der statistischen Sicherheit auf den zugehörigen Wert der Teststatistik umrechnen.
3. Teststatistik aus den Messdaten berechnen.
4. Die beiden Werte für die Teststatistik vergleichen. Entscheiden.

Ein zeitgemässeres Vorgehen:

1. Statistische Sicherheit festlegen.
2. Teststatistik aus den Messdaten berechnen.
3. Mit einem Computerprogramm aus der Teststatistik den zugehörigen Wert der statistischen Sicherheit berechnen.
4. Die beiden Wahrscheinlichkeiten vergleichen. Entscheiden.

In Lehrbüchern wird auch heute noch fast ausschliesslich das historische Vorgehen behandelt. In Kapitel 9 ist eine Sammlung von häufig gebrauchten statistischen Tests zusammengestellt. Dabei wird ebenfalls nach dem historischen Vorgehen verfahren. Die entsprechende Modifikation der Tests ist trivial.

#### 5.4. Entscheidungsstrategie

Das Fällen von Entscheidungen aufgrund unscharfer Werte ist nicht trivial. Im täglichen Leben fällt es uns eher leicht, da die Tragweite der meisten Entscheidungen bescheiden ist. In der analytischen Chemie hingegen muss das Entscheidungsverhalten strengen Kriterien genügen. Die Ausformulierung eines statistischen Tests folgt unmittelbar aus der Entscheidungsstrategie. Ein Beispiel soll das veranschaulichen:

Durch Messungen soll festgestellt werden, ob die Konzentration einer Verunreinigung in einem Lebensmittel einen gesetzlich festgelegten Grenzwert überschreitet. Die Konzentrationsbestimmung führt zu einem unscharfen Wert. Die gemessene Konzentration hat die Form einer Verteilung. Daraus muss ein Wert ausgewählt werden, der danach mit dem Grenzwert verglichen werden kann. Eine mögliche Situation ist in Fig. 5.4 dargestellt.

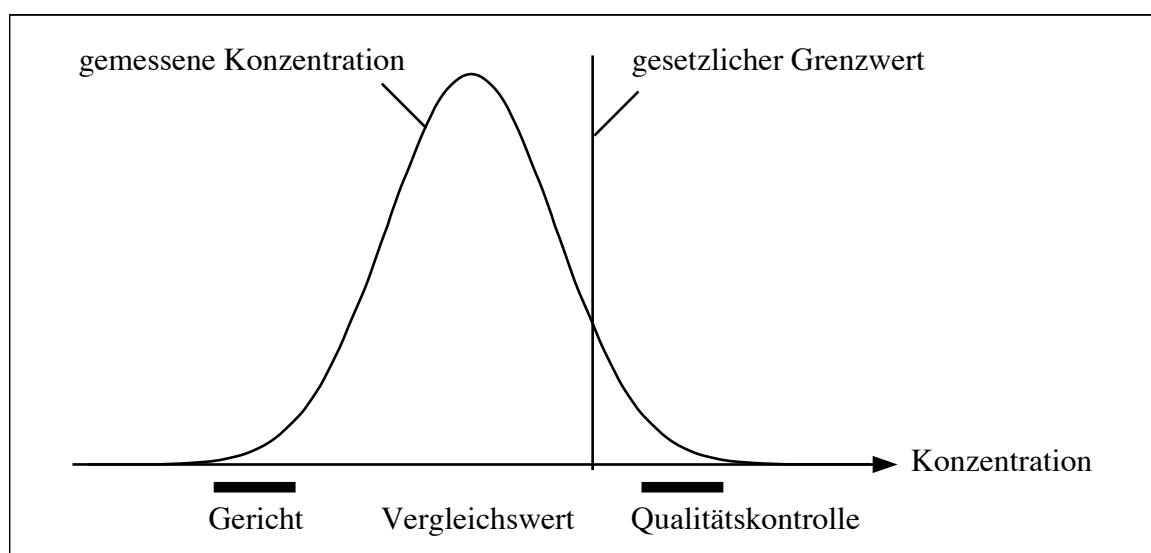


Fig. 5.4 Wahl des Vergleichswertes bei gegensätzlichen Entscheidungsstrategien

In dieser Situation befindet sich für die gemessene Konzentration beidseits des Grenzwertes so viel Wahrscheinlichkeit, dass die Entscheidung nicht von vornherein klar ist. Insbesondere scheint eine Überschreitung des Grenzwertes möglich. Es muss nun aus der Verteilung ein Wert  $c_{\text{Vergleich}}$  ausgewählt werden, der mit dem Grenzwert verglichen werden soll. Man kann folgende Nullhypothese definieren:

$$H_0 : c_{\text{Vergleich}} - c_{\text{Grenz}} = 0$$

Es wird also wieder der Fall gewählt, dass die Konzentration mit dem Grenzwert übereinstimmt. Dies ist die Situation, die bereits in Kapitel 5.3. besprochen wurde. Als Alternativhypothese wird die Überschreitung des Grenzwertes angenommen:

$$H_A : c_{\text{Vergleich}} - c_{\text{Grenz}} > 0$$

Die Wahl des Vergleichswertes hängt nun aber sehr stark von der Situation des Entscheidenden ab. Im Umgang mit unscharfen Werten können zwei Fehlentscheide auftreten.

- Es wird nicht erkannt, dass der Grenzwert überschritten ist.
- Es wird fälschlicherweise angenommen, der Grenzwert sei überschritten.

Es ist nicht von vornherein klar, welcher Fall der schlimmere ist. Die Entscheidungsstrategie muss darauf ausgerichtet sein, den schlimmsten Fall mit einer möglichst kleinen Wahrscheinlichkeit eintreten zu lassen.

Die Abteilung Qualitätskontrolle des Produzenten ist für die Einhaltung der gesetzlichen Bestimmungen verantwortlich. Das Erkennen von Qualitätsmängeln ist sehr wichtig. Lieber werden weitere Messungen angeordnet, oder es wird gar ein einwandfreies Produkt vernichtet, als dass ein verunreinigtes Produkt zum Verkauf freigegeben wird. Der Vergleichswert wird dann aus dem rechts liegenden Teil der Verteilung gewählt. Der Bereich ist in Fig. 5.4 durch einen schwarzen Balken markiert. Man könnte z. B. die einseitige 99%-Schranke wählen, also oberhalb des Vergleichswertes 1 % Wahrscheinlichkeit belassen. Dadurch wird schon beim leisesten Verdacht Alarm ausgelöst. In der vorliegenden Situation würde die Alternativhypothese klar angenommen.

Der Kantonschemiker, der vor der Entscheidung steht, ob ein Produkt wegen verletzter gesetzlicher Bestimmungen aus dem Verkehr gezogen werden soll, wird wohl etwas weniger ungestüm handeln. Er wird einen Kompromiss finden müssen zwischen der Verpflichtung zum Schutz der Konsumenten und der Gefahr eines peinlichen Irrtums. Eine Behörde wird in dieser Situation eher eine gegenteilige Strategie verfolgen, wie im nächsten Fall:

Das Gericht, das nach einer angefochtenen Beanstandung über eine Verurteilung zu entscheiden hat, ist zu einer zurückhaltenden Strategie verpflichtet. Von Gesetzes wegen gilt jedermann als unschuldig, dessen Schuld nicht bewiesen ist. (Was in diesem Zusammenhang ein "Beweis" ist, sei dahingestellt.) Der Vergleichswert wird aus dem untersten Bereich der Verteilung gewählt. Dadurch wird eine Überschreitung erst dann angezeigt, wenn kaum mehr ein Zweifel besteht. In der vorliegenden Situation würde die Alternativhypothese keinesfalls angenommen.

Wie dieses Beispiel zeigt, können zwei Entscheidungen aufgrund des gleichen Datenmaterials aber unterschiedlicher Entscheidungsstrategie gegensätzlich ausfallen.



## 6. Schrittweises Vorgehen beim Entwurf eines Modells

### 6.1. Charakterisierung der Variablen und ihrer Zusammenhänge

Sämtliche für das Experiment wichtigen Grössen werden aufgezählt. Von jeder Variablen wird festgestellt, was für eine Rolle sie im Experiment spielt. Einige Variablen werden auf einem bestimmten Wert festgehalten, andere werden systematisch variiert. Es mag Variablen geben, über die man nur sehr wenig weiss. Man nimmt dann meistens an, ihr Einfluss auf andere Grössen sei klein (sonst wüsste man mehr über sie). Besonders wichtig ist die Charakterisierung der systematisch geänderten Variablen.

In der statistischen Literatur haben sich die Begriffe abhängige Variable (dependent variable) und unabhängige Variable (independent variable) eingebürgert. Diese Sprachregelung wurde etwas unglücklich gewählt. Eine Variable kann nicht für sich allein "abhängig" sein. Abhängig wovon? Man würde solche Variablen besser als "unkontrolliert" bezeichnen. Der Begriff "unabhängig" ist vollends irreführend. Intuitiv könnte man glauben, eine "unabhängige" Variable hänge von keiner anderen Variablen ab. Das kann zwar sein, ist aber keine Voraussetzung. Manchmal ist der Zusammenhang sogar exakt bekannt. Besser wäre hier die Bezeichnung "kontrolliert".

Experimente in der quantitativen Analyse zielen im Allgemeinen darauf ab, in einem physikalischen System eine Konzentration oder eine damit verwandte Grösse zu bestimmen. Konzentrationen sind aber kaum jemals direkt messbar. Man muss einen Umweg beschreiten. In der Photometrie misst man die Intensität von Licht, in der Potentiometrie misst man elektrische Spannungen, in der Gravimetrie bestimmt man Massen, etc. Die direkt gemessene Grösse ist von der Konzentration in bestimmter Weise abhängig. Diese Tatsache nutzt man aus, um auf die Konzentration zu schliessen. Sobald man die Konzentration berechnet hat, interessiert man sich nicht mehr für die Werte der direkt messbaren Grösse. Diese Situation ist für die analytische Chemie typisch.

Obwohl der funktionale Zusammenhang zwischen der Konzentration und der direkt messbaren Grösse im Prinzip bekannt ist, fehlt im Allgemeinen noch ein Teil der Information. Gewisse in den Gleichungen vorkommende Koeffizienten sind meist unbekannt. Man behilft sich mit einer Kalibration (calibration). Die Konzentration wird in systematischer und bekannter Weise variiert und die direkt zugängliche Grösse gemessen. Danach wendet man das Messverfahren auf Systeme mit unbekannter Konzentration an. Die zuvor bestimmte Kalibrationsfunktion erlaubt nun die Umrechnung von der gemessenen Grösse auf die Konzentration.

Man bezeichnet die direkt messbare Grösse als "abhängig". Meistens gibt es nur eine Grösse dieser Art, bei einem Methodenvergleich allenfalls zwei. Alle anderen Variablen nennt man "unabhängig". Die "abhängigen" sind mit den "unabhängigen" Variablen über einen kausalen Zusammenhang verknüpft. Eine Erhöhung der Konzentration verursacht eine bestimmte Änderung der gemessenen Grösse.

Über einen allfälligen Zusammenhang zwischen den "unabhängigen" Variablen wird überhaupt nichts ausgesagt. Das ist auch nicht nötig. Die Variablen brauchen nicht wirklich unabhängig zu sein. Die hier behandelten statistischen Verfahren setzen kein bestimmtes Verhalten voraus.

Nachdem die abhängigen Variablen identifiziert sind, werden alle bekannten Abhängigkeiten zwischen den relevanten Grössen aufgeschrieben. Dabei kann man sich auf den Einfluss der unabhängigen Variablen auf die abhängigen beschränken. Jede Abhängigkeit lässt sich als Gleichung formulieren. Man findet hier Naturgesetze und Ingenieurgleichungen, Neben-, Rand- und Anschlussbedingungen, Näherungen und Annahmen.

Die Freiheit beim Modellbau ist in diesem Stadium praktisch grenzenlos. Alles ist erlaubt, solange man weiss, was man tut. Es kann sogar sinnvoll sein, einen Zusammenhang zu formulieren, von dem man weiss, dass er eigentlich falsch ist. So ist es in speziellen Situationen durchaus zulässig, einen komplizierten Zusammenhang zu "linearisieren", indem einfach eine lineare Funktion ange-

passt wird. Siehe Fig. 6.1. Wenn die Messwerte in einem engen Bereich liegen, in dem sich die Krümmung der Kurve kaum auswirkt, ist die Beschreibung durch eine Gerade ebenso gut. Man darf dann natürlich über den Bereich ausserhalb der Messwerte nichts aussagen.

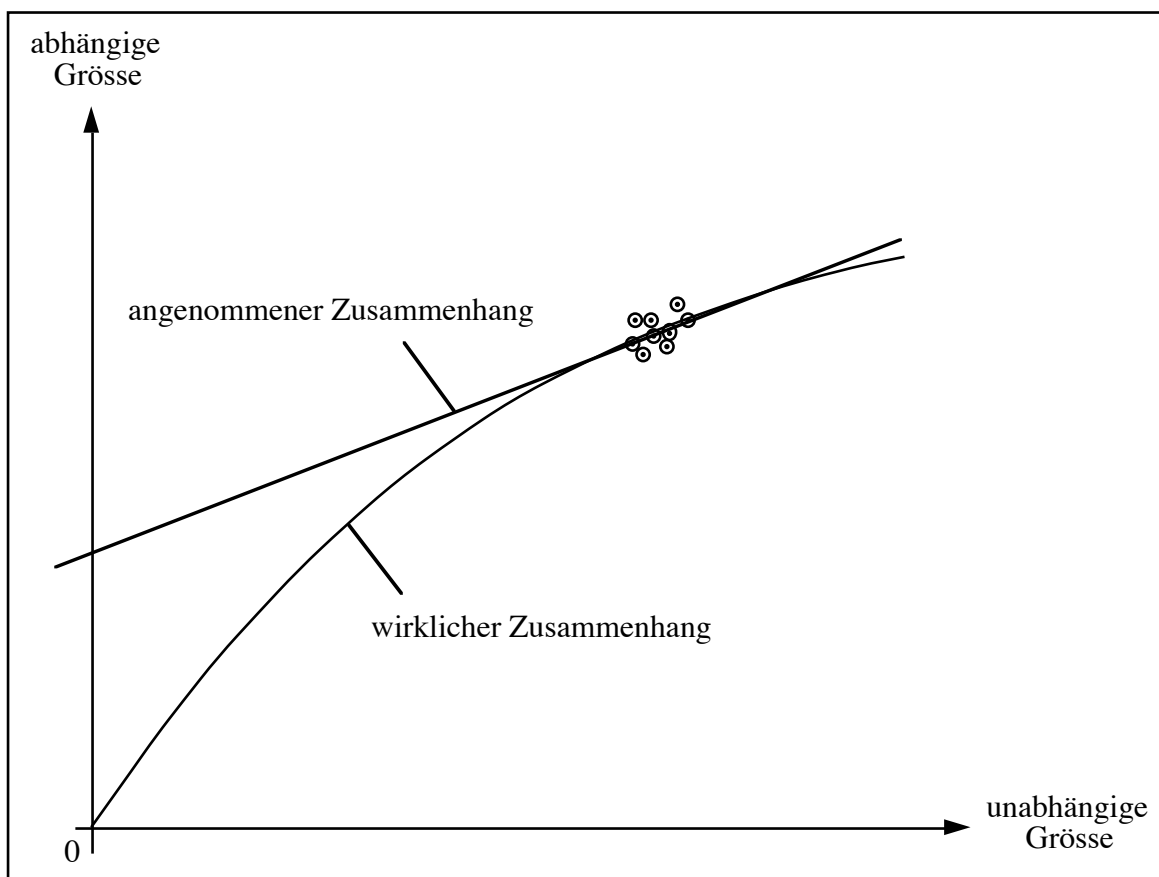


Fig. 6.1 Wahl eines "falschen" aber sinnvollen Zusammenhangs

Falls man die Freiheit hat, den funktionalen Zusammenhang relativ frei wählen zu können, sollten grundsätzlich so wenige anzupassende Parameter wie möglich eingeführt werden.

Manchmal hat eine unabhängige Variable einen nicht verstandenen Einfluss auf die abhängigen Variablen. Man kann den Zusammenhang also nicht ohne weiteres in Formeln fassen. Es kann auch vorkommen, dass der Zusammenhang im Prinzip bekannt aber derart kompliziert ist, dass man bei einer Beschreibung die Komplexität des Modells unverhältnismässig aufblähen würde. In diesem Fall kann man die Variable experimentell konstant halten. Dadurch erspart man sich die Beschreibung des Zusammenhangs.

Das entstehende Gleichungssystem wird nun nach den abhängigen Variablen aufgelöst. Auf der linken Seite der Gleichungen steht also nur noch eine abhängige Variable, die rechten Seiten enthalten nur unabhängige Variablen. Dies lässt sich hoffentlich bewerkstelligen. Wenn nicht, ist das auch nicht tragisch, die numerische Behandlung ist aber erschwert.

## 6.2. Charakterisierung der Messfehler

Nur die abhängigen Variablen werden hier als fehlerbehaftet betrachtet. Messfehler sind Zufallsvariablen, streuen also um einen Mittelwert. Es stellt sich nun die Frage, wie Mittelwert und Streuung der Messfehler von den anderen Variablen abhängen. In diesem Stadium des Modellbaus werden zum ersten Mal statistisch relevante Grössen betrachtet.

Der Mittelwert von Messfehlern ist im Allgemeinen null. Damit gleichbedeutend ist die Aussage, dass die Schätzungen der gemessenen Grössen verzerrungsfrei (free of bias) sind. Man schätzt also tatsächlich den entsprechenden wahren Wert der Grösse, nicht irgend einen danebenliegenden. Man nimmt also grundsätzlich an, der wahre Mittelwert der Messfehler sei identisch null in allen Variablen. (Manchmal gilt das erst nach einer Transformation der gemessenen Grössen.)

Für die Streuung der Messfehler liegen die Dinge nicht so einfach. Messfehler setzen sich aus kleinen Beiträgen zusammen, die aus den unterschiedlichsten Quellen stammen können. Dabei kann man zwei Typen von Fehlerquellen unterscheiden, die besonders häufig sind:

- Der absolute Fehler  $\Delta x$  einer Grösse  $x$  ist unabhängig von  $x$ .  
In diese Kategorie gehören z. B. Ablesefehler und Detektorrauschen.
- Der relative Fehler  $\Delta x/x$  einer Grösse  $x$  ist unabhängig von  $x$ .  
In diese Kategorie gehören z. B. Verdünnungsfehler.

Die abhängige Grösse wird im Allgemeinen mit Fehlern beider Sorten behaftet sein. Man versucht also, die wichtigsten Fehlerquellen zu charakterisieren. Dabei hilft die Überlegung, wie sich die Messfehler beim Nullpunkt (wahrer Wert der Messwerte = 0) verhalten. Oft ist der Wert null eine unüberschreitbare Schranke für die Messwerte. Ein Gaschromatogramm zeigt keine negativen Peaks. Ein Photometer mit Fluoreszenzdetektor kann keine negative Intensität anzeigen. Eine mittlere Zerfallszeit kann nicht negativ sein. Wenn der fragliche Anteil des Fehlers beim Nullpunkt (praktisch) verschwindet, ist er von der zweiten Sorte, wenn er unabhängig vom wahren Wert ist, von der ersten Sorte.

Fast immer lässt sich aber der Hauptanteil des Fehlers einer der beiden Fehlerarten zuordnen. Man nimmt dann an, dass alle Fehler sich wie der Hauptanteil verhalten, dass es also nur Fehler eines Typs gibt. Wenn das nicht der Fall ist, wird der Modellbau sehr kompliziert. Dieser Fall wird hier nicht behandelt. Man kann jetzt also zwei Fälle unterscheiden:

1. Fall: Die Messfehler sind von der ersten Sorte. Man nimmt im Allgemeinen für die Messfehler eine Normalverteilung an. Die Standardabweichung ist unabhängig vom Messwert.
2. Fall: Die Messfehler sind von der zweiten Sorte. Man nimmt im Allgemeinen an, die Messwerte (wahrer Wert  $\mu$  + Messfehler) seien lognormalverteilt. Die Standardabweichung ist proportional zum wahren Wert  $\mu$  der Messwerte.

Im zweiten Fall wird man die Messwerte sinnvollerweise transformieren. Wenn die Messwerte logarithmiert werden, streuen die Logarithmen gemäss einer Normalverteilung mit konstanter Standardabweichung und Mittelwert  $\log(\mu)$ . Die Grössen  $\log(\text{Messwert}) - \log(\mu)$  streuen also um den Wert null mit konstanter Standardabweichung, haben demnach das gleiche Verhalten wie die Messfehler im ersten Fall, womit der zweite Fall auf den ersten zurückgeführt ist.

Die Annahme einer Normalverteilung ist üblich, wenn die Verteilung offensichtlich symmetrisch ist und wenn nur eine kleine Stichprobe zur Verfügung steht. Bei schiefen Verteilungen oder sehr umfangreichen Stichproben können auch alternative Verteilungen angenommen werden.

Es wird vorausgesetzt, dass die unabhängigen Variablen nicht fehlerbehaftet sind. In der Praxis ist das natürlich nicht der Fall. Es stellt sich die Frage, wie sich Abweichungen vom exakten Verhalten äussern. Durch zufällige Fehler verursachte Schwankungen in den unabhängigen Variablen übertragen sich auf die abhängige Grösse und führen dort zu zusätzlichen Schwankungen. In Fig. 6.2 ist die Situation für den eindimensionalen Fall dargestellt.

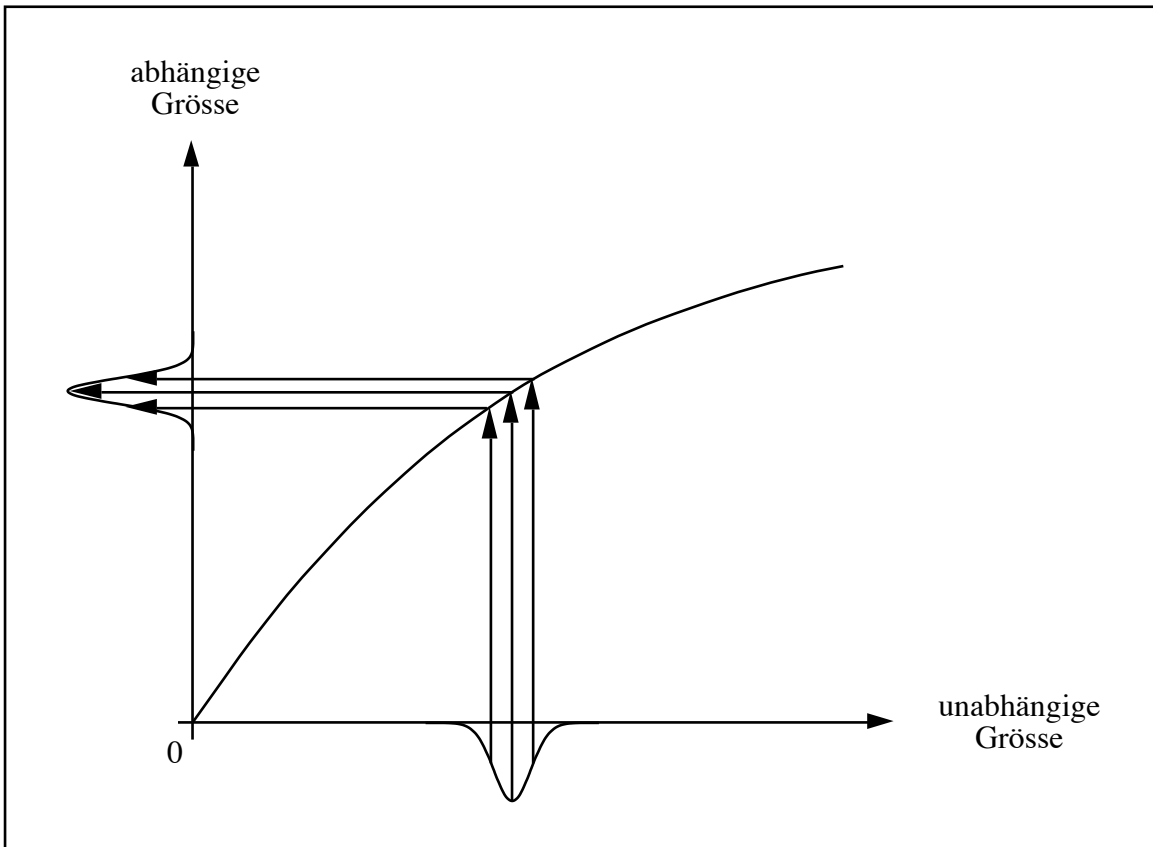


Fig. 6.2 Effekt von Messfehlern in den unabhängigen Variablen

Wie man sieht, sollte die Krümmung der Kurve im "Übertragungsbereich" der Messfehler klein sein, da sonst die Verteilung der resultierenden Zusatzeffekte asymmetrisch wird.

Wesentlich wichtiger aber ist die Forderung, dass die Zusatzeffekte klein sind im Vergleich zu den eigentlichen Messfehlern der abhängigen Variablen. Die Quadrate von zufälligen Fehlern, also die Varianzen sind additiv, nicht die Standardabweichungen. Nennt man die abhängige Variable  $y$  und die unabhängigen Variablen  $x_1, x_2, \dots, x_n$ , so lässt sich die Forderung quantitativ so ausdrücken:

$$\sigma_y^2 > \sum_{i=1}^n \left( \sigma_{x_i} \frac{\partial y}{\partial x_i} \right)^2 \quad (6.1)$$

Sie lässt sich im Allgemeinen leicht erfüllen. Wenn ein Zusatzeffekt 10 % des eigentlichen Messfehlers der abhängigen Variablen beträgt, wird die Standardabweichung des totalen Fehlers nur 2 % über jener der "originalen" Messfehler liegen:

$$\sigma_{\text{neu}} = \sqrt{\sigma_{\text{alt}}^2 + (0,1 \sigma_{\text{alt}})^2} = \sqrt{1,01 \sigma_{\text{alt}}^2} \approx 1,02 \sqrt{\sigma_{\text{alt}}^2} = 1,02 \sigma_{\text{alt}}$$

Der Zusatzeffekt ist also in diesem Fall völlig unbedeutend. Aufgrund des Zentralen Grenzwertsatzes darf man zudem annehmen, dass die Überlagerung der vielen kleinen Störeinflüsse die Annahme einer Normalverteilung nicht kompromittiert.

Wenn die Verteilung der Messfehler festgelegt ist, hat man das Verhalten der Messwerte vollständig beschrieben. Das Modell ist gewählt.

### 6.3. Mathematische Behandlung des Modells

Oft wird in der analytischen Chemie ein Gerät modelliert, mit dem Konzentrationen bestimmt werden können. Das Messgerät liefert ein einziges Messsignal. Hier wird daher nur der Fall einer einzigen abhängigen Variablen als Funktion einer beliebigen Anzahl unabhängiger Variablen und Parameter behandelt. Um das Modell in eine saubere mathematische Form bringen zu können, sind einige Größen zu definieren.

$$y = f(x_1, x_2, \dots, \beta_1, \beta_2, \dots, \beta_p) + \varepsilon = f(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon \quad (6.2)$$

dabei bedeuten

y: abhängige Variable (Messwert)  
 f: beliebige bekannte Funktion  
 x: Vektor der unabhängigen, exakt bekannten Variablen  
 $\boldsymbol{\beta}$ : Vektor der p unbekannt anzupassenden Parameter  
 $\varepsilon$ : Messfehler

An die Funktion f werden keine besonderen Ansprüche gestellt, die in der Praxis nicht ohnehin erfüllt sind. Eine der unabhängigen Variablen  $x_i$  ist möglicherweise ausgezeichnet als jene Grösse, die das Messgerät letztlich bestimmen soll, also z. B. eine Konzentration, die über den Umweg der abhängigen Variablen y zugänglich gemacht wird. Die p Parameter  $\beta_1, \dots, \beta_p$  sind vorerst unbekannte Größen, die vermutlich über eine Kalibration bestimmt werden müssen. Die Messfehler  $\varepsilon$  streuen gemäss einer Normalverteilung mit Mittelwert null und Standardabweichung  $\sigma$ . Es wird vorausgesetzt, dass die Messfehler unkorreliert sind.

In einem konkreten Fall stehen n Messwerte  $y_1, y_2, \dots, y_n$  (Vektor  $\mathbf{y}$ ) mit den zugehörigen Werten für die unabhängigen Variablen  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  zur Verfügung. Die Schätzung  $\mathbf{b}$  der Parameter  $\boldsymbol{\beta}$  erhält man, indem man die Funktion

$$\Phi = \sum_{i=1}^n (y_i - f(\mathbf{x}_i, \mathbf{b}))^2 \quad (6.3)$$

durch Variation des Vektors  $\mathbf{b}$  minimiert. Dieses Vorgehen ist unter dem Namen "Methode der kleinsten Quadrate" (eigentlich "... der kleinsten Fehlerquadratsumme") bekannt. Wie dies numerisch geschieht, ist für jede Situation individuell abzuklären. Dabei hat man grundsätzlich keine Gewähr, dass man das globale Minimum der Funktion  $\Phi$  gefunden hat. Ein wichtiger Spezialfall wird weiter unten behandelt.

Mithilfe der geschätzten Parameter  $\mathbf{b}$  kann man nun die Messwerte um den geschätzten Messfehler korrigieren:

$$\hat{\mathbf{y}} = (f(\mathbf{x}_1, \mathbf{b}), f(\mathbf{x}_2, \mathbf{b}), \dots, f(\mathbf{x}_n, \mathbf{b}))^T \quad (6.4)$$

Die Schätzungen  $\mathbf{e}$  der Messfehler  $\boldsymbol{\varepsilon}$  sind:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \quad (6.5)$$

Die geschätzten Messfehler bezeichnet man als Residuen (residuals). Die Schätzung  $s^2$  der Fehlervarianz  $\sigma^2$  ist:

$$s^2 = \frac{\mathbf{e}^T \mathbf{e}}{n - p} = \frac{\mathbf{e}^T \mathbf{e}}{v} \quad (6.6)$$

$$v = n - p \quad (6.7)$$

Die Grösse  $v$  bezeichnet man als Anzahl Freiheitsgrade. Dies ist die Anzahl unabhängiger Messwerte, die für die Schätzung der Varianz zur Verfügung stehen. Sind es genau  $p$  Messwerte, können die Parameter daraus zwar bestimmt werden, es bleiben aber keine Messwerte mehr übrig, um eine Streuung zu schätzen Vgl. Gleichung (4.5).

Der Zusammenhang der Messfehler mit der Unsicherheit der Parameter ist im allgemeinen Fall kompliziert. Um die Voraussetzung der Normalverteilung für die Messfehler möglichst gut auszunutzen, wird die Fehlerquadratsumme  $\Phi$  in eine Taylor-Reihe in den Parametern entwickelt, wobei nur Terme erster Ordnung und die paarweisen Kreuzterme berücksichtigt werden. Vgl. Gleichung (4.16). Dies bedeutet unter anderem, dass bei der Variation eines Parameters die wirkliche Funktion durch die Tangente ersetzt wird, wie in Fig. 6.3 dargestellt.

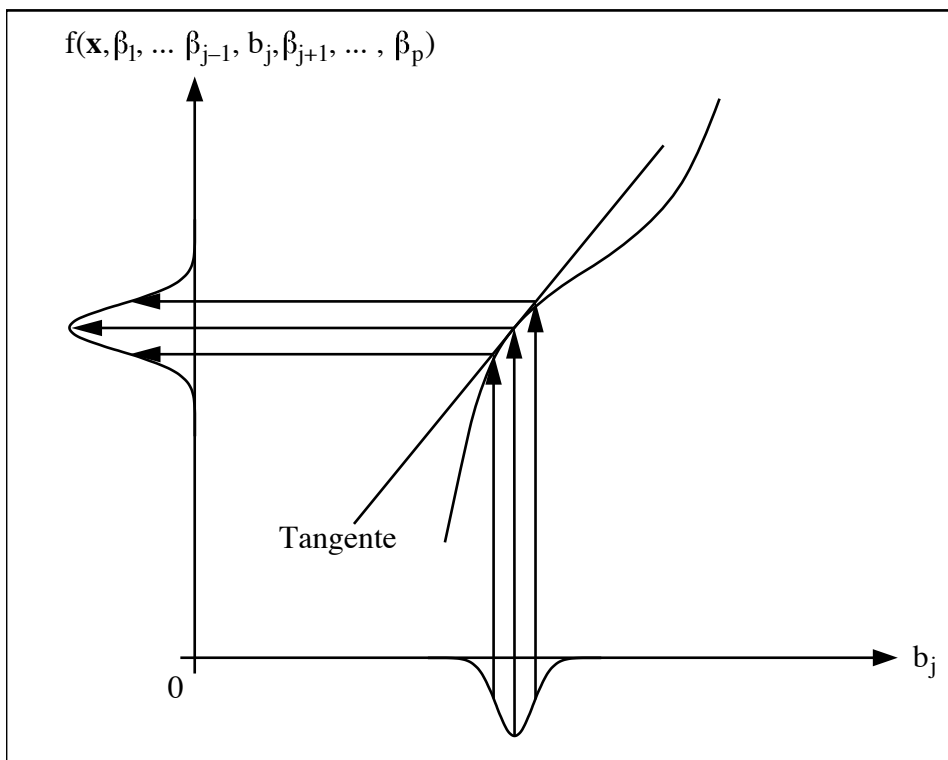


Fig. 6.3 Linearisierung der Funktion  $f$ : Ersatz durch die Tangente

Damit diese Näherung akzeptabel ist, dürfen die Fehler in den Variablen nicht übermässig gross sein. Die Überprüfung der Bedingung ist gerade dann nicht trivial, wenn sie nicht erfüllt ist. Zwar kann man gemäss Fig. 6.3 die Krümmung der Funktion überprüfen, aber die exakten Parameter sind nicht bekannt. Wenn stattdessen die stark fehlerbehafteten Schätzungen eingesetzt werden, ist dies kritisch.

Für die Berechnung der Fehler in den Parametern bildet man die  $n \times p$  Matrix  $\mathbf{P}$  mit den Elementen:

$$P_{ij} = \frac{\partial f(\mathbf{x}_i, \boldsymbol{\beta})}{\partial \beta_j} \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, p \quad (6.8)$$

Dabei stellt der Vektor  $\mathbf{x}_i$  die Realisationen der Variablen  $\mathbf{x}$  beim  $i$ -ten Experiment dar. Daraus wird die Varianz-Kovarianzmatrix  $\sigma^2(\mathbf{P}^T \mathbf{P})^{-1}$  der geschätzten Parameter berechnet.

Falls die Modellfunktion nur in impliziter Form vorliegt,

$$0 = h(y-\varepsilon, \mathbf{x}, \boldsymbol{\beta}) \quad (6.9)$$

kann man die Matrixelemente von  $\mathbf{P}$  folgendermassen berechnen:

$$P_{ij} = - \frac{\partial h(y-\varepsilon, \mathbf{x}_i, \boldsymbol{\beta})}{\partial \beta_j} \left( \frac{\partial h(y-\varepsilon, \mathbf{x}_i, \boldsymbol{\beta})}{\partial (y-\varepsilon)} \right)^{-1} \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, p \quad (6.10)$$

Dabei ist  $y-\varepsilon$  der Erwartungswert der Messwerte, also eine exakte Grösse.

Da die exakten Parameter  $\boldsymbol{\beta}$  nicht bekannt sind, werden die Schätzungen  $\mathbf{b}$  eingesetzt. Es ist prinzipiell unklar, wie gut diese Näherung ist. Das ist im Einzelfall abzuklären, was aber nicht trivial ist.

An dieser Stelle wird ein wichtiger Spezialfall behandelt. Man spricht von einem linearen Modell, falls sich die Funktion  $f$  so schreiben lässt:

$$f(\mathbf{x}, \boldsymbol{\beta}) = \sum_{j=1}^p \beta_j g_j(\mathbf{x}) \quad (6.11)$$

In den beliebigen Funktionen  $\mathbf{g}$  kommen die Parameter nicht vor. Dies gilt auch für die Matrixelemente von  $\mathbf{P}$ :

$$P_{ij} = \frac{\partial f(\mathbf{x}_i, \boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \sum_{j=1}^p \beta_j g_j(\mathbf{x}) = g_j(\mathbf{x}_i) \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, p \quad (6.12)$$

Die Matrix  $\mathbf{P}$  ist also ebenfalls exakt bekannt. Das Modell lässt sich nun so schreiben:

$$\mathbf{y} = \mathbf{P}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (6.13)$$

Damit lässt sich auch die Fehlerquadratsumme kompakt beschreiben:

$$\Phi = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{P}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{P}\boldsymbol{\beta}) = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{P}\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{P}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{P}^T \mathbf{P}\boldsymbol{\beta} \quad (6.14)$$

Die Grössen  $\boldsymbol{\beta}^T \mathbf{P}^T \mathbf{y} = (\mathbf{y}^T \mathbf{P}\boldsymbol{\beta})^T = \mathbf{y}^T \mathbf{P}\boldsymbol{\beta}$  sind Skalare und daher mit ihren Transponierten identisch. Daraus folgt:

$$\Phi = \mathbf{y}^T \mathbf{y} - 2 \boldsymbol{\beta}^T \mathbf{P}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{P}^T \mathbf{P}\boldsymbol{\beta} \quad (6.15)$$

Die Parameter werden geschätzt, indem die Fehlerquadratsumme minimiert wird. Dazu werden die Ableitungen nach den Parametern null gesetzt und das entstehende Gleichungssystem gelöst. Hier wird nicht darauf eingegangen, wie die Ableitungen berechnet werden, sondern gleich das Resultat gegeben:

$$\frac{\partial \Phi}{\partial \boldsymbol{\beta}} = - 2 \mathbf{P}^T \mathbf{y} + 2 \mathbf{P}^T \mathbf{P}\boldsymbol{\beta} \quad (6.16)$$

Die Bestimmungsgleichungen für die Schätzungen  $\mathbf{b}$  lassen sich also so formulieren:

$$\mathbf{0} = \mathbf{P}^T \mathbf{P} \mathbf{b} - \mathbf{P}^T \mathbf{y} \quad (6.17)$$

$$\mathbf{P}^T \mathbf{P} \mathbf{b} = \mathbf{P}^T \mathbf{y}$$

Durch Linksmultiplizieren mit  $(\mathbf{P}^T \mathbf{P})^{-1}$  erhält man schliesslich

$$\mathbf{b} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{y} \quad (6.18)$$

Im Fall eines linearen Modells lassen sich die Schätzungen der Parameter also unmittelbar berechnen. Die geschätzten Messwerte erhält man so:

$$\hat{\mathbf{y}} = \mathbf{P} \mathbf{b} = \mathbf{P} (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{y} \quad (6.19)$$

Die weiteren Berechnungen sind für lineare und nichtlineare Modelle gleich. Bei nichtlinearen Modellen handelt es sich aber um Näherungen, deren Qualität nicht ohne weiteres festgestellt werden kann.

Die symmetrische Matrix  $s^2 (\mathbf{P}^T \mathbf{P})^{-1}$  heisst Varianz-Kovarianz-Matrix der Parameter. Sie entspricht der Matrix  $\mathbf{VC}$  in Gleichung (4.18). In der Diagonalen stehen die Varianzen der Parameter, die Ausserdiagonalelemente sind deren Kovarianzen.

Da die Parameter aus dem gleichen Datensatz geschätzt wurden, sind sie korreliert. Die symmetrische Matrix  $\mathbf{R}$  der paarweisen Korrelationskoeffizienten lässt sich aus  $\mathbf{P}$  berechnen. Dazu wird zuerst die Diagonalmatrix  $\mathbf{D}$  definiert.  $\mathbf{D}$  ist im wesentlichen die Matrix  $(\mathbf{P}^T \mathbf{P})^{-1}$ , jedoch werden die Ausserdiagonalelemente null gesetzt. Aus  $\mathbf{D}$  berechnet man die Diagonalmatrizen  $\mathbf{D}^{1/2}$  und  $\mathbf{D}^{-1/2}$ . Sie haben die Diagonalelemente  $\sqrt{d_{ii}}$  bzw.  $\sqrt{1/d_{ii}}$ , wobei  $d_{ii}$  die Diagonalelemente von  $\mathbf{D}$  sind. Damit kann  $\mathbf{R}$  berechnet werden:

$$\mathbf{R} = \mathbf{D}^{-1/2} (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{D}^{-1/2} \quad (6.20)$$

Die Elemente von  $\mathbf{R}$  sind dimensionslose Zahlen aus dem Intervall  $[-1 \dots 1]$ . In der Diagonalen stehen Einsen. Im Fall eines linearen Modells sind die Korrelationskoeffizienten exakt bekannt, da die Parameter nicht in  $\mathbf{R}$  erscheinen.

Die Matrix  $\mathbf{R}$  ist von besonderer Wichtigkeit, insbesondere wenn die Parameter interpretiert werden sollen. Die Matrix ist in jedem Fall genau zu studieren. Wenn der Betrag eines oder mehrerer Ausserdiagonalelemente den Wert von 0.9 wesentlich übersteigt, ist Vorsicht geboten. Die zugehörigen geschätzten Parameter sind stark korreliert. Sie können nicht ohne weiteres getrennt voneinander interpretiert werden. Wenn einer der Parameter stark verändert würde, könnte dies durch geeignete Änderung des anderen grösstenteils wieder rückgängig gemacht werden. Um eine bestimmte Charakteristik der Funktion  $f$  zu beschreiben, genügt eigentlich ein einziger Parameter. Da es nun aber zwei sind, ist nicht klar, welcher wieviel Anteil beschreibt. Die starke Korrelation mag ein Hinweis auf ein zu komplexes Modell sein, das zu viele Parameter enthält. Die Funktion  $f$  ist zu flexibel, um das Datenmaterial kompakt zu beschreiben. Man kann sich behelfen, indem das Datenmaterial gezielt vergrössert wird. Wenn das nicht sinnvoll ist oder das Modell nicht vereinfacht werden kann, ist gegen derartige Korrelationen praktisch nichts zu unternehmen. Man muss damit leben.

Nicht so kritisch ist die Situation, wenn die Parameter als solche nicht von grosser Bedeutung sind und nicht interpretiert werden sollen. Beispielsweise können Steigung und Achsenabschnitt einer linearen Funktion sehr stark korreliert sein. Wenn die Funktion nur bestimmt wurde, um als Kalibrationsgerade für weitere Messungen zu dienen, ist eine Korrelation der Parameter unbedeutend.



Individuelle Vertrauensbereiche für die Schätzungen der Parameter kann man so berechnen:

$$\mathbf{b} \pm t(v;1-\alpha) \mathbf{s}_{\mathbf{b}} = \mathbf{b} \pm t(v;1-\alpha) \mathbf{s} \mathbf{D}^{1/2} \mathbf{1}_p \quad (6.21)$$

Der gemeinsame Vertrauensbereich der Parameter kann als Hyperellipsoid dargestellt werden:

$$(\boldsymbol{\beta} - \mathbf{b})^T (\mathbf{P}^T \mathbf{P})^{-1} (\boldsymbol{\beta} - \mathbf{b}) \leq s^2 p F(v, p, 1-\alpha) \quad (6.22)$$

Die Bedeutung des gemeinsamen und der individuellen Vertrauensbereiche wird am besten an einem Beispiel veranschaulicht. In Fig. 6.4 ist ein Fall für  $p = 2$  dargestellt. Die beiden Parameterschätzungen sind negativ korreliert.

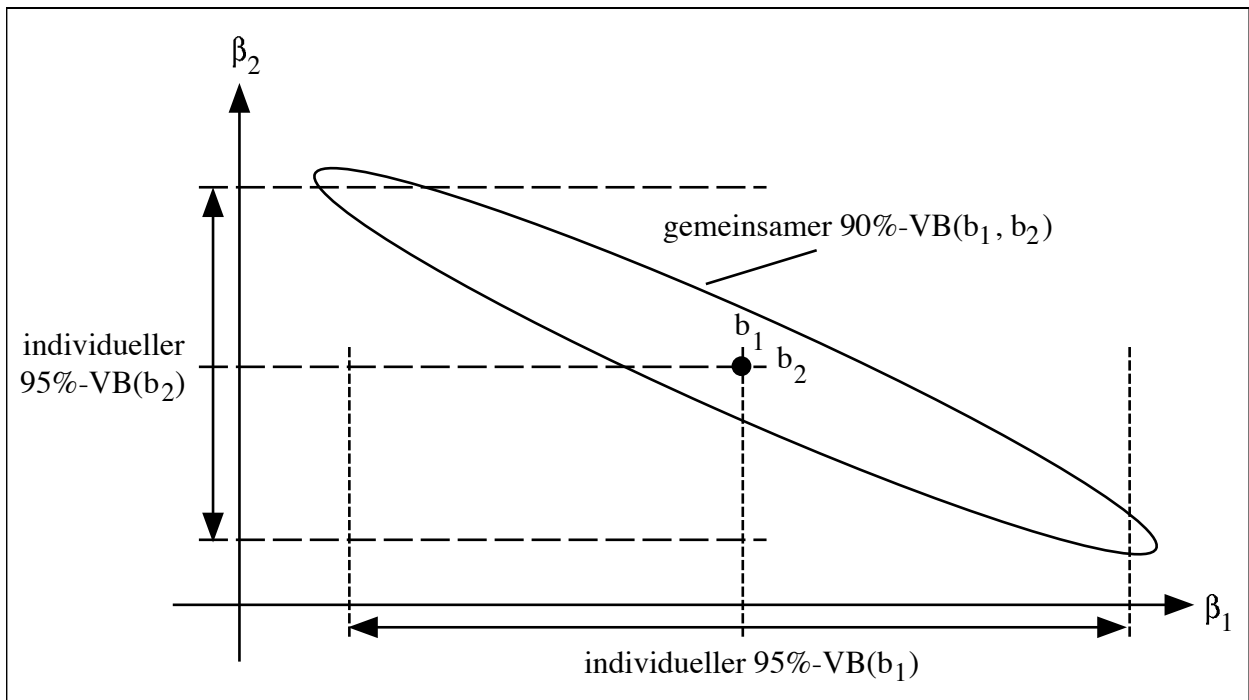


Fig. 6.4 Vergleich des gemeinsamen mit den individuellen Vertrauensbereichen. Siehe Text.

Es mag etwas seltsam anmuten, dass die Achsen mit  $\beta_1$  und  $\beta_2$  angeschrieben sind. Die wahren Parameter können nicht variabel sein. Jedoch handelt es sich hier gewissermassen um einen Rollentausch. Die Schätzungen haben in dieser Darstellung eine feste Position. Nun wird gefragt, wo sich die wahren Parameter befinden könnten.

Die individuellen Fehlerbereiche sind als Doppelpfeile entlang der Achsen bezeichnet. Sie sind zu verwenden, wenn es der Sinn des Experimentes war, die Parameter zu schätzen und einzeln zu beurteilen. Es ist zwar wichtig zu wissen, dass es eine Korrelation gibt, aber bei der Beurteilung spielt sie eine untergeordnete Rolle. Wenn z. B. zwei Gleichgewichtskonstanten bestimmt wurden, ist es für deren Beurteilung unerheblich, ob die Schätzungen korreliert sind.

Der gemeinsame Vertrauensbereich ist als Ellipse dargestellt. Parameter-Kombinationen ausserhalb des Bereiches sind unwahrscheinlich, obwohl die Parameter allenfalls innerhalb der individuellen Vertrauensbereiche liegen, die bei starker Korrelation grundsätzlich aufgebläht sind.

Der geschätzte Messwert  $\hat{y}$  bei ausgewählten Werten  $\mathbf{x}_0$  entspricht einer nichtlinearen Transformation der geschätzten Parameter  $\mathbf{b}$ , nämlich  $\hat{y} = f(\mathbf{x}_0, \mathbf{b})$ . Daher lässt sich seine Varianz  $s_{\hat{y}}^2$  im Allgemeinen nicht in geschlossener Form angeben. Man verwendet stattdessen die Formel für die Fort-

pflanzung zufälliger Fehler. Vgl. Gleichungen (4.16) und (4.20). Man setzt in den Ableitungen anstelle der Parameter  $\boldsymbol{\beta}$  deren Schätzwerte  $\mathbf{b}$  ein. Da die Parameter aus der gleichen Stichprobe geschätzt wurden, sind sie korreliert. Man benötigt also noch die paarweisen Korrelationskoeffizienten  $\rho_{i,j}$ . Die Ausserdiagonalelemente der Matrix  $\mathbf{R}$  stellen eine Näherung für die Korrelationskoeffizienten dar, im Fall eines linearen Modells sind sie exakt.

Wendet man die Gleichung (4.19) an,

$$\mathbf{d} = \left( \frac{\partial f(\mathbf{x}_0, \boldsymbol{\beta})}{\partial \beta_1} \quad \frac{\partial f(\mathbf{x}_0, \boldsymbol{\beta})}{\partial \beta_2} \quad \dots \quad \frac{\partial f(\mathbf{x}_0, \boldsymbol{\beta})}{\partial \beta_p} \right)^T \quad (6.23)$$

erhält man gemäss (4.20)

$$s_{\hat{y}}^2 = s^2 \mathbf{d}^T (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{d} \quad (6.24)$$

wobei wiederum anstelle der Parameter  $\boldsymbol{\beta}$  deren Schätzungen  $\mathbf{b}$  eingesetzt werden. Mithilfe dieses Ausdrucks lassen sich nun drei Fehlerbereiche für  $s_{\hat{y}}$  angeben.

Vertrauensbereich für die Schätzung des wahren Mittelwertes bei  $\mathbf{x}_0$ :

$$VB_{\hat{y}} = f(\mathbf{x}_0, \mathbf{b}) \pm t(v; 1-\alpha) \sqrt{s_{\hat{y}}^2} \quad (6.25)$$

Toleranzbereich für die Voraussage eines einzelnen Messwertes bei  $\mathbf{x}_0$ :

$$TB_{\hat{y}} = f(\mathbf{x}_0, \mathbf{b}) \pm t(v; 1-\alpha) \sqrt{s^2 + s_{\hat{y}}^2} \quad (6.26)$$

Vertrauensbereich für die Voraussage eines Mittelwertes von  $m$  Messwerten bei  $\mathbf{x}_0$ :

$$VB_{\hat{y}} = f(\mathbf{x}_0, \mathbf{b}) \pm t(v; 1-\alpha) \sqrt{\frac{s^2}{m} + s_{\hat{y}}^2} \quad (6.27)$$

Die Streuungen der Residuen sind nicht einheitlich, sondern abhängig von der Wahl der Variablen  $x_i$ . Um die Residuen vergleichen zu können, werden sie mit einem Faktor multipliziert, den man auch als Länge des Hebelarms oder einfach als Hebelarm (leverage) bezeichnet:

$$L = \frac{1}{\sqrt{1 - \mathbf{d}^T (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{d}}} \quad (6.28)$$

Die mit dem Hebelarm multiplizierten Residuen bezeichnet man als normalisiert (normalized). Sie streuen mit der konstanten Standardabweichung  $s$ . Normalisierung ist besonders nützlich, wenn Hebelpunkte vorhanden sind (siehe weiter unten, z.B. Fig. 6.4).

Manchmal wünscht man, aus gemessenen  $y$ -Werten unter Verwendung einer vorgängigen Kalibration auf den zugehörigen Vektor  $\hat{\mathbf{x}}_0$  zu schliessen. Als Schätzung für den Mittelwert kann man den Wert der inversen Modellfunktion verwenden. Dabei können sich aber vielfältige Schwierigkeiten ergeben. Die inverse Funktion braucht nicht definiert zu sein, oder es kann unendlich viele gleichwertige Schätzungen geben. Es ist völlig offen, wie das Problem im Einzelfall zu lösen ist. Zur Berechnung des Vertrauensbereiches kann man die Methoden der Fehlerfortpflanzung verwenden. Am besten wird das Vorgehen anhand eines einfachen Beispiels erläutert. Vgl. Kapitel 7.

## 6.4. Test des Modells und des Datensatzes

Nachdem das Modell gewählt ist, unterzieht man es einer kritischen Prüfung. Unzulänglichkeiten werden manchmal erst sichtbar, wenn eine erste Auswertung vollständig abgeschlossen ist. Man braucht dazu also einen Datensatz. Zu überprüfen sind die grundlegenden Annahmen über den funktionalen Zusammenhang zwischen den Variablen und das Verhalten der Messfehler. Eine Möglichkeit, Abweichungen vom vorausgesetzten Verhalten aufzuspüren, ist eine Residuenanalyse.

### 6.4.1. Residuenanalyse

Messfehler enthalten keine Information. Sie streuen in zufälliger Art um den Mittelwert null. Da die Residuen die geschätzten Messfehler sind, müssen sie dieses Verhalten widerspiegeln. Abweichungen von den Modellannahmen schlagen sich oft als systematische Zusatzeffekte in den Residuen nieder. Es ist daher eine gute Angewohnheit, die Residuen auf systematische Effekte irgendeiner Art zu untersuchen und damit die Anwendbarkeit des Modells zu prüfen. Am besten trägt man die Residuen graphisch gegen sinnvoll ausgewählte Variablen auf. Was immer für Variablen man wählt, die Residuen sollen in völlig zufälliger Weise um den Mittelwert null streuen. Für die beiden häufigsten Abweichungen ist in Fig. 6.3 ein Beispiel gezeigt.

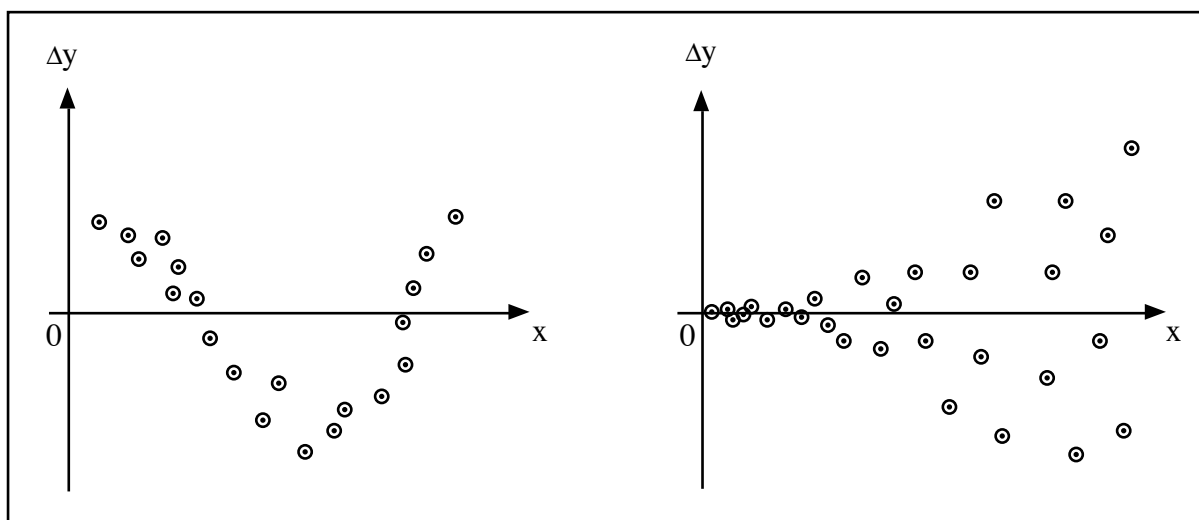


Fig. 6.3 Beispiele für die beiden häufigsten pathologischen Residuenplots

Links in Fig. 6.3 zeigen die Residuen einen systematischen Gang in ihrem Mittelwert, rechts in ihrer Standardabweichung. Im Fall links dürfte der funktionale Zusammenhang zwischen der gemessenen Grösse und einer wichtigen unabhängigen Variablen unzweckmässig modelliert worden sein. Im Fall rechts entspricht das angenommene Verhalten der Messfehler nicht den Tatsachen. Es handelt sich in beiden Fällen um Spezifikationsfehler. Die Auswertung sollte mit einem verbesserten Modell wiederholt werden.

### 6.4.2. Systematische Fehler

Abweichungen vom vorausgesetzten Verhalten können auch durch systematische Fehler verursacht sein. Dabei handelt es sich also um Fehler, die nicht dem Modell, sondern den Messwerten selbst anhaften. Systematische Fehler können in letzter Konsequenz nur mit unabhängigen Experimenten erkannt werden. Daraus kann man möglicherweise eine Hypothese formulieren und Messwerte von zurückliegenden Experimenten nachträglich entsprechend korrigieren. Ein klassisches Beispiel:

Man weiss aufgrund von physikalischen Überlegungen, dass die Kalibrationskurve durch den Nullpunkt des Koordinatensystems läuft. Gleichzeitig vermutet man aber einen gleichbleibenden Störeinfluss auf die Messwerte durch eine Kontamination der Kalibrationsproben durch unsaubere Chemikalien. Das konstante Zusatzsignal kann man durch ein entsprechendes Modell auffangen.

Man wird dabei aber (mit statistischen Methoden) nie nachweisen können, dass es sich wirklich um systematische und nicht um Spezifikationsfehler handelt.

### 6.4.3. Ausreisser

Nicht nur das Modell ist kritisch zu beleuchten, sondern auch der Datensatz. Oft entdeckt man in der Stichprobe einen Wert, der sich von den anderen wesentlich zu unterscheiden scheint. Er zeichnet sich durch ein besonders grosses Residuum aus. Man bezeichnet einen solchen Wert als Ausreisser (outlier). Es stellt sich die Frage, was man in dieser Situation tun soll. Es gibt im wesentlichen drei Sorten Ausreisser:

- Durch grobe Fehler verfälschte Werte lassen sich meist nachträglich als fremd erkennen und können korrigiert oder entfernt werden.
- Daten aus einer fremden Grundgesamtheit, deren Fremdartigkeit man aber nicht ohne weiteres nachweisen kann, müssten in jedem Fall entfernt werden.
- Atypische Vertreter aus der gleichen Grundgesamtheit wie die anderen Daten stören, da sie das Bild der Grundgesamtheit ernstlich verfälschen können. Sie sind nicht repräsentativ für die Grundgesamtheit.

Eigentlich ist eine Unterscheidung zwischen den ersten beiden Punkten unnötig. Auch mit groben Fehlern behaftete Messwerte stammen letztlich aus einer fremden Grundgesamtheit, gehören also in die zweite Kategorie. Der Unterschied liegt eher im Ausmass der Fehler und in der Leichtigkeit, mit der grobe Fehler im Allgemeinen erkannt werden können.

Es stehen mehrere Massnahmen zur Behandlung von Ausreissern zur Auswahl:

- man tut gar nichts
- man entfernt den Ausreisser
- man ändert den Ausreisser ab
- man gibt dem Ausreisser ein kleineres Gewicht
- man ändert das statistische Modell

Es gibt keine allgemeingültigen Regeln, was man mit einem Ausreisser tun soll. Die Versuchung ist gross, den Störenfried stillschweigend zu entfernen. Das kann in gewissen Fällen die richtige Strategie sein, ist es aber meist nicht. Zuerst prüft man den verdächtigen Wert auf Herz und Nieren. Nur wenn es Argumente nichtstatistischer Art gibt, die auf einen verfälschten Wert hindeuten, entfernt man ihn aus dem Datensatz und wiederholt die Auswertung. Wenn sich solche Hinweise nicht finden lassen, kann man die Auswertung mit und ohne Ausreisser durchführen und die Resultate vergleichen. Dadurch erkennt man die Konsequenzen des Ausreissers. Die Behandlung von Ausreissern ist prinzipiell abhängig von der Entscheidungsstrategie (siehe 5.4.).

**Manchmal ist ein Ausreisser der interessanteste Wert der Stichprobe.  
Fast alle grossen Entdeckungen waren einmal Ausreisser irgendeiner Art.**

### 6.4.4. Hebelpunkte

Der Experimentator ist fast völlig frei in der Gestaltung der Kalibration. Dadurch können Kalibrationen sehr unterschiedliche Qualität haben, wie folgendes extremes Beispiel zeigt:

Fig. 6.4 zeigt die Situation eines linearen Zusammenhangs zwischen einer abhängigen und einer unabhängigen Variablen. Es ist leicht ersichtlich, dass die Steigung der geschätzten Geraden in erheblichem Ausmass von einem einzigen Messpunkt abhängt. Man bezeichnet einen solchen Datenpunkt als Hebelpunkt (leverage point). Es handelt sich um einen Punkt weit weg von den anderen Punkten. Hebelpunkte sind aber nicht mit Ausreissern zu verwechseln. Ausreisser sind

Werte der abhängigen Variablen, die sich durch (betragsmässig) besonders grosse Residuen auszeichnen; sie liegen weit weg von ihrem geschätzten Mittelwert.

Dieses Konzept ist nicht übertragbar auf unabhängige Variablen, da diese keine Zufallsvariablen sind. Sie streuen nicht um einen wahren Mittelwert, sondern können vom Experimentator nach beliebigen Kriterien, und nach Modellannahme fehlerfrei, eingestellt werden. Wenn ein Punkt weit weg von anderen Punkten liegt, wurde das vom Experimentator (vermutlich) explizit so gewählt.

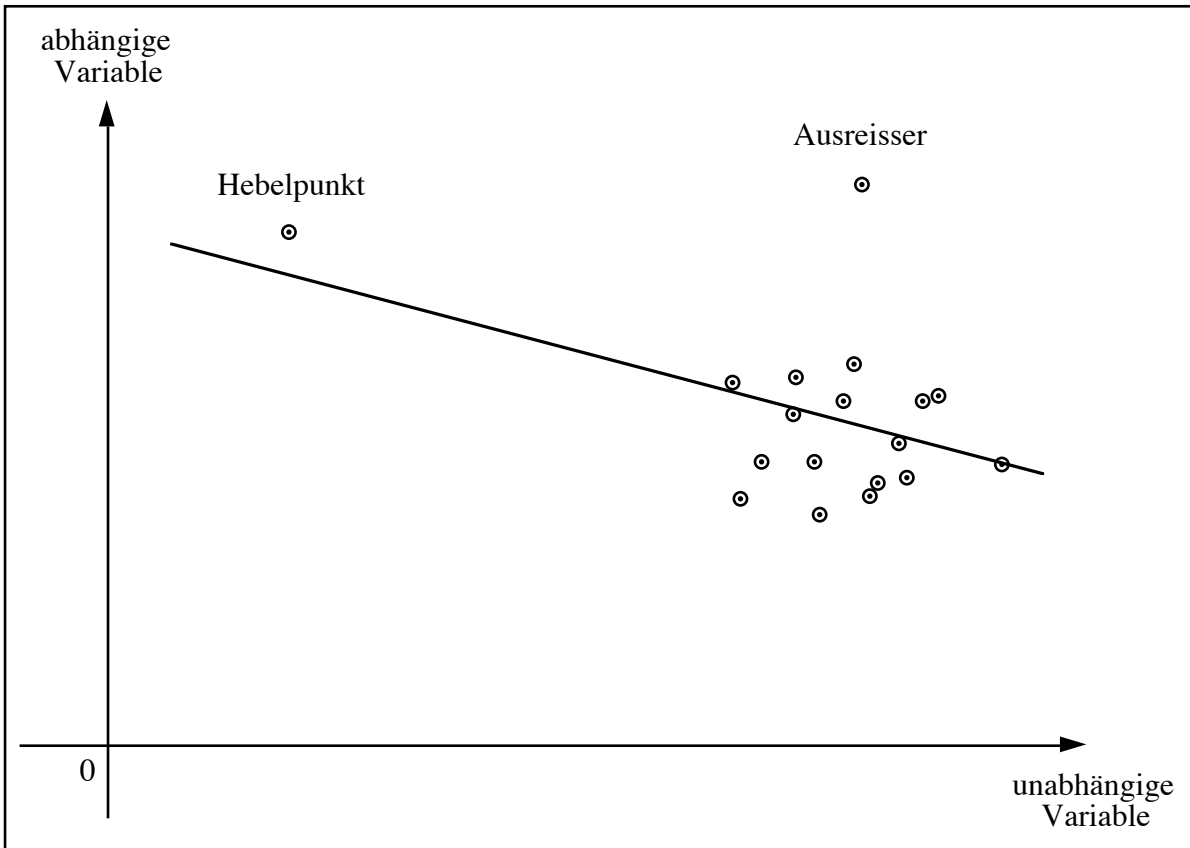


Fig. 6.4 Unzulängliches Modell mit Hebelpunkt

Um den Unterschied zwischen Ausreissern und Hebelpunkten zu verdeutlichen, ist in Fig. 6.4 auch noch ein Ausreisser eingezeichnet.

Hebelpunkte sind im Gegensatz zu Ausreissern weitgehend harmlos, sofern man über ihre Existenz unterrichtet ist. Sie verfälschen die Ergebnisse in keiner Weise, wenn das Modell die Wirklichkeit gut widerspiegelt. Dennoch ist bei der Anwesenheit von Hebelpunkten besondere Vorsicht angezeigt. Abweichungen von den angenommenen Modelleigenschaften können einen verheerenden Einfluss auf die Richtigkeit der Ergebnisse haben. Es könnte z. B. sein, dass der Hebelpunkt aus einer fremden Grundgesamtheit stammt, weil bei seiner Bestimmung das Messgerät nicht richtig funktionierte. Es handelte sich dabei also um einen Messwert, der mit einem groben Fehler behaftet ist. Im Allgemeinen ist ein solcher Wert ein Ausreisser, er hat ein besonders grosses Residuum. Die Residuen von Hebelpunkten sind aber ganz allgemein kleiner als im Mittel, daher ist der grobe Fehler schwer zu erkennen. Besser ist es, zur Beurteilung die normierten Residuen zu verwenden, die mit dem Hebelarm multipliziert sind. Sie haben die gleiche Standardabweichung. Im gezeigten Beispiel reagieren die Schätzungen für den Achsenabschnitt und die Steigung der Geraden ausserordentlich stark auf den Hebelpunkt. Das Modell ist dadurch wenig robust.

Es wird nun angenommen, dass es zusätzliche Information über den funktionalen Zusammenhang gibt. Aufgrund physikalischer Überlegungen wisse man, dass die Gerade durch den Nullpunkt des Koordinatensystems läuft. Der Zusammenhang ist mithin nicht nur linear, sondern sogar propor-

tional. Fig. 6.5 zeigt die neue Situation. Die zusätzliche Information erweist sich als sehr wertvoll. Der solitäre Punkt kann nicht mehr als Hebelpunkt bezeichnet werden, da er keinen übermässigen Einfluss auf die Schätzung irgendeines Lageparameters hat. Es stellt sich zudem heraus, dass der ehemalige Hebelpunkt ein Ausreisser ist.

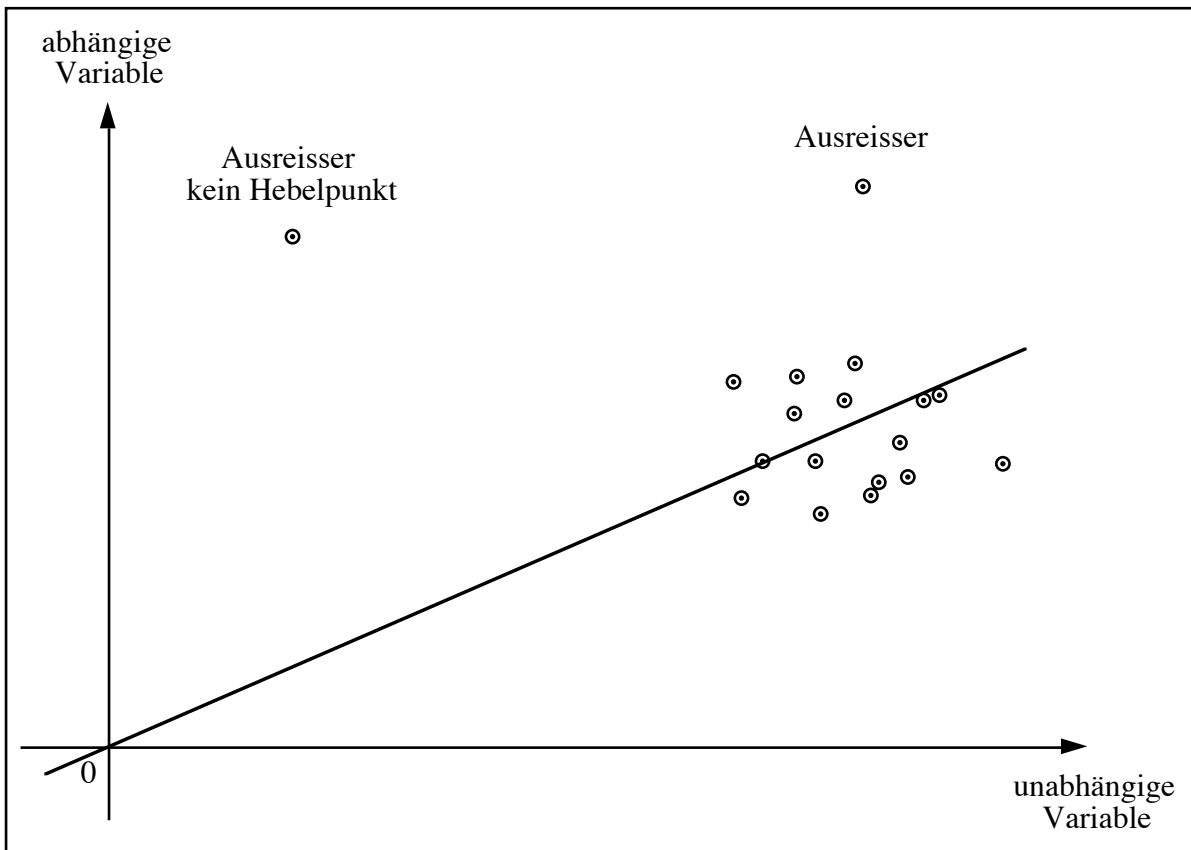


Fig. 6.5 Stärkeres Modell mit enttarntem zweitem Ausreisser

Dieses Beispiel zeigt, wie wichtig es ist, die ganze bekannte Information über den Zusammenhang der Variablen ins Modell einzubauen. Ziel sollte es sein, möglichst wenige anzupassende Parameter im Modell zu haben.

Im gezeigten Beispiel gibt es nur eine einzige unabhängige Variable. Das Erkennen von Hebelpunkten durch Visualisierung mit Graphiken ist im Fall mehrerer unabhängiger Variablen ungleich schwieriger als im trivialen eindimensionalen Fall. Man ist auf das numerische Kriterium in Gleichung (6.28) angewiesen. Durch seine Betrachtung kann die Tücke der Hebelpunkte teilweise entschärft werden.

#### 6.4.5. Übermässige Flexibilität des Modells

Immer wieder kommt es vor, dass ein Modell zwar alle Anforderungen erfüllt, aber möglicherweise für die zu behandelnden Datensätze zu flexibel ist. Es gibt die begründete Vermutung, dass anzupassende Parameter überflüssig sind oder anderweitig zusammenhängen. Einige Beispiele:

- Die Funktion zur Beschreibung des Potentials einer pH-Elektrode enthält einen Term für die Kreuz-Empfindlichkeit gegenüber  $\text{Na}^+$ . Ist dieser Term wirklich nötig? Wenn nicht, würde das die Kalibration der Elektrode wesentlich vereinfachen.

- Ein Modell zur Beschreibung eines Messgeräts enthält einen Term, der für das Messgerät charakteristisch ist. Nachdem das Modell Dutzende Male auf Datensätze angewandt wurde, ist ersichtlich, dass der Term nur sehr wenig variiert. Darf man in Zukunft den Mittelwert aller bisherigen Schätzungen verwenden und den Parameter nicht mehr anpassen?
- Ein Messgerät liefert ein Signal, das von einer Konzentration linear abhängig ist. Es wird jeden Morgen nach dem Einschalten kalibriert. Es ist offensichtlich, dass der Achsenabschnitt von Tag zu Tag variiert. Es gibt aber die begründete Vermutung, dass die Steigung wesentlich stabiler ist. Ist es zulässig, zusammen mit der neuen Kalibration auch jene der 9 vorausgegangenen Arbeitstage zu verwenden und 10 Achsenabschnitte aber nur eine einzige Steigung anzupassen? Das würde bedeuten, den 10-fachen Datensatz zur Verfügung zu haben, was die Qualität der Kalibration enorm vergrössern würde.

Unnötige Flexibilität eines Modells sollte vermieden werden. Sie ist bestenfalls unnötig, schlimmstenfalls schädlich.

Jeder Parameter, der den Wert Null oder etwas Vergleichbares schätzt, "verbraucht" dazu Messwerte, verschwendet also Ressourcen.

Unzulänglichkeiten im Datensatz können die Parameter so beeinflussen, dass sie die Wirklichkeit verzerrt wiedergeben, was aber möglicherweise niemandem auffällt. Bei einem weniger flexiblen Modell hätte vielleicht ein Blick auf die Residuenplots genügt, um zu erkennen, dass etwas nicht stimmen kann.

Ausreisser können maskiert werden, indem z.B. ein unnötiger "Krümmungsterm" signifikant von null verschieden werden kann, obwohl er sonst von null nicht zu unterscheiden ist.

Es gibt eine Möglichkeit, anhand eines Datensatzes zwei Modelle unterschiedlicher Flexibilität miteinander zu vergleichen. Die Quadratsumme  $\Phi$  in Gleichung (6.14) vergrössert sich offensichtlich, wenn ausgehend von einem "vollen Modell" zusätzliche Restriktionen für die Parameter eingeführt werden, die Flexibilität also eingeschränkt wird. Es lässt sich verstehen, in welchem Ausmass  $\Phi$  ansteigt, wenn beide Modelle alle Anforderungen erfüllen.

Dies kann ausgenützt werden, um einen statistischen Test zu formulieren, der die beiden Quadratsummen vergleicht. Sollte die Beschränkung der Flexibilität zu einem unzulänglichen Modell führen, äussert sich das in einer zusätzlichen Vergrösserung der Quadratsumme, die vom Test erkannt werden kann. Damit aber der Test anwendbar ist, müssen folgende Anforderungen erfüllt sein.

- Das zugrundeliegende "volle" Modell erfüllt die Anforderungen, die zu Beginn des Kapitels 6.3 formuliert wurden.
- Das "volle" Modell ist genügend flexibel.
- Das "volle" Modell ist linear gemäss Gleichung (6.11). Ist das nicht der Fall, ist der Test approximativ. Dabei ist nicht ohne Weiteres zu erkennen, inwiefern der Test noch brauchbar ist.
- Das abgeleitete "reduzierte" Modell stellt lineare Beziehungen zwischen den Parametern des "vollen" Modells her. Die Beschränkung der Flexibilität lässt sich also durch folgende Form ausdrücken:

$$a_0 = \sum_{i=1}^p a_i \cdot \beta_i \quad (6.29)$$

Dabei sind die  $a_i$  exakt bekannte Zahlen.

Es können beliebig viele Zusammenhänge gemäss (6.29) definiert werden. Für eine kompakte Schreibweise formuliert man die Koeffizienten  $a_i$  (ausser  $a_0$ ) als  $p$ -dimensionalen Zeilenvektor. Die  $r$  unabhängigen Zusammenhänge fasst man in der Matrix  $\mathbf{T}$  zusammen. Sie hat  $r$  Zeilen und  $p$  Spalten. Die Zusammenhänge können nun so formuliert werden:

$$\mathbf{a}_0 = \mathbf{T}\boldsymbol{\beta} \quad (6.30)$$

mit dem  $r$ -dimensionalen Spaltenvektor  $\mathbf{a}_0$ . Er ist praktisch immer der Nullvektor. Gleichung (6.30) stellt gleichzeitig die Nullhypothese für den folgenden F-Test dar. Die Alternativhypothese ist  $H_A: \mathbf{a}_0 \neq \mathbf{T}\boldsymbol{\beta}$ . Bei einem F-Test ist die Teststatistik der Quotient zweier Varianzen, die miteinander verglichen werden sollen. Zu jeder Varianz gehört eine Anzahl Freiheitsgrade.

Im Nenner steht die Varianz  $s^2$  gemäss Gleichung (6.6) für das volle Modell mit der Anzahl Freiheitsgrade nach (6.7).

Der Zähler wird folgendermassen berechnet, wobei wieder die Grössen für das volle Modell verwendet werden:

$$s_Z^2 = s^2 \frac{(\mathbf{Tb} - \mathbf{a}_0)^T [\mathbf{T}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{T}^T]^{-1} (\mathbf{Tb} - \mathbf{a}_0)}{r} \quad (6.31)$$

Zur Varianz des Zählers gehören  $r$  Freiheitsgrade. Die Teststatistik berechnet sich also nach

$$F_0 = \frac{s_Z^2}{s^2} = \frac{(\mathbf{Tb} - \mathbf{a}_0)^T [\mathbf{T}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{T}^T]^{-1} (\mathbf{Tb} - \mathbf{a}_0)}{r} \quad (6.32)$$

Übersteigt  $F_0$  den F-Wert für die gewählte statistische Sicherheit  $\alpha$ , gilt also  $F_0 > F_{\alpha, r, v}$ , wird man die Nullhypothese verwerfen. Die Beschränkung der Flexibilität hat dann also das Modell kompromittiert.

Es handelt sich hier um einen einseitigen Test. Man will ausschliesslich eine Erhöhung der Zählervarianz feststellen.



## 6.5. Entscheidung aufgrund des Schlussresultates

Die Messungen wurden wahrscheinlich durchgeführt, um eine Entscheidung zu fällen. Typische Fragen, die mithilfe der Experimente beantwortet werden sollen, sind etwa:

Unterscheiden sich die Resultate zweier Messmethoden systematisch?  
 Ist eine gesetzlich festgelegte Höchstkonzentration überschritten?  
 Ist ein potentiell Medikament genügend wirksam?  
 Gibt es eine Abhängigkeit zwischen zwei Variablen?

Um solche Fragen zu beantworten, wird man einen statistischen Test entwerfen.

Wer Entscheidungen zu fällen hat, soll das Ergebnis klar formulieren und begründen, also nicht nur die ausgewerteten Resultate veröffentlichen, sondern auch die daraus zu ziehenden Schlüsse. Der Grund für diese Forderung liegt in der Tatsache begründet, dass Aussenstehende nicht alle Tücken und Schwierigkeiten des Modellbaus nachvollziehen können. Der mit dem Experiment und dem Modellbau unmittelbar Vertraute verfügt über Informationen, zu denen Andere nicht ohne weiteres einen Zugang haben. Diese Zusatzinformationen stammen vor allem aus zwei Quellen:

- Erfahrungen im Fachgebiet erlauben Abschätzungen der Grösse der zu erwartenden zufälligen Fehler und der Form deren Verteilung. Oft hat man auch Erfahrungen über den Einfluss vergessener oder unzureichend erfasster Variablen. Letztlich ist dies also Wissen über die Anwendbarkeit des gewählten statistischen Modells in der Praxis.
- Allgemeines statistisches Fachwissen gibt einem eine Ahnung von der Zuverlässigkeit statistischer Verfahren und deren Anfälligkeit gegenüber Abweichungen von den vorausgesetzten Eigenschaften.

Dieses Wissen hat einen wesentlichen Einfluss auf das Entscheidungsverhalten des Experimentators. Andere Experten mögen mit dem gleichen Datenmaterial und dem gleichen Wissen zu anderen Entscheidungen gelangen. Dennoch ist das Urteil eines direkt Betroffenen wesentlich besser fundiert als das jedes Aussenstehenden. Es ist demnach wichtig, dass die Entscheidung mitgeteilt wird. Journalisten werden dankbar sein.

Um entscheiden zu können, braucht man nicht nur Messwerte und ein vernünftig gebautes Modell zu deren Behandlung, sondern auch eine klar formulierte Entscheidungsstrategie. Man stellt sich auf einen Standpunkt. Für Leute, die eine Entscheidung nachvollziehen wollen, muss dieser klar ersichtlich sein. Von diesem Standpunkt hängt unter anderem die Wahl einer statistischen Sicherheit ab. Ein Kritiker wird zudem mit einem Ausreisser im Datensatz anders verfahren als ein Entdecker. Die "sichere Seite" hängt unmittelbar von der Entscheidungsstrategie ab.

## 7. Herleitung der Linearen Regression

Die Lineare Regression ist ein wichtiger Spezialfall des Formalismus aus Kapitel 6. Daher wird die Lineare Regression als Beispiel im Detail durchgerechnet. Dabei wird die Formulierung mit Matrizen und Vektoren übernommen. Im Kapitel 8.3 wird die Lineare Regression in einer etwas anderen Darstellung nochmals vorgestellt. Die Formeln sind völlig äquivalent.

### 7.1. Formulierung der Modellfunktion

Die Funktion  $f$  in (6.2) hat im Fall der Linearen Regression die Form:

$$f(\mathbf{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x \quad (7.1)$$

Die Anzahl Parameter beträgt also  $p=2$ . Man bildet den 2-dimensionalen Spaltenvektor  $\boldsymbol{\beta}$ , indem man die Parameter in festgelegter Reihenfolge untereinander schreibt:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \quad (7.2)$$

### 7.2. Beschreibung der experimentellen Messwerte

Man legt eine eindeutige Reihenfolge der gemessenen Einzelwerte  $y_i$  fest. Der  $n$ -dimensionale Spaltenvektor  $\mathbf{y}$  wird gebildet, indem man die Werte der abhängigen Variablen in der festgelegten Reihenfolge untereinander schreibt:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (7.3)$$

Die Matrix  $\mathbf{P}$  hat 2 Spalten und  $n$  Zeilen. Die Matrixelemente werden gemäss (6.8) berechnet. Die Reihenfolge der Spalten ist jene der Parameter. Die Ableitungen nach  $\beta_0$  sind 1. Die erste Spalte der Matrix  $\mathbf{P}$  ist also ein  $\mathbf{1}$ -Vektor. Bei den Ableitungen nach  $\beta_1$  bleibt jeweils der Wert  $x_i$ .

$$\mathbf{P} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad (7.4)$$

### 7.3. Schätzungen der Parameter

Die Modellfunktion bei der Linearen Regression erfüllt die Anforderungen an Linearität gemäss (6.11). Die Schätzungen der Parameter können also nach (6.18) berechnet werden. Die Matrixmultiplikationen werden hier explizit durchgeführt.

1. Schritt: Berechnung von  $\mathbf{P}^T\mathbf{P}$ :

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \mathbf{P}$$

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix} \quad (7.5)$$

$$\mathbf{P}^T \quad \mathbf{P}^T\mathbf{P}$$

2. Schritt: Berechnung der Inversen von  $\mathbf{P}^T\mathbf{P}$ :

Generell ist die Inverse einer 2x2-Matrix A gegeben durch

$$\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}^{-1} = \frac{1}{a_{11} a_{22} - a_{12} a_{21}} \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix} \quad (7.6)$$

Angewandt auf die Matrix  $\mathbf{P}^T\mathbf{P}$  ergibt sich:

$$\begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}^{-1} = \frac{1}{n \sum x^2 - (\sum x)^2} \begin{pmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{pmatrix} \quad (7.7)$$

3. Schritt: Berechnung von  $\mathbf{P}^T\mathbf{y}$ :

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

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} \sum y \\ \sum xy \end{pmatrix} \quad (7.8)$$

$$\mathbf{P}^T \quad \mathbf{P}^T\mathbf{y}$$

4. Schritt: Multiplikation von  $(\mathbf{P}^T\mathbf{P})^{-1}$  mit  $\mathbf{P}^T\mathbf{y}$

$$\begin{pmatrix} \Sigma_y \\ \Sigma_{xy} \end{pmatrix} \mathbf{P}^T\mathbf{y}$$

$$\begin{pmatrix} \frac{\Sigma_x^2}{n \Sigma_x^2 - (\Sigma_x)^2} & \frac{-\Sigma_x}{n \Sigma_x^2 - (\Sigma_x)^2} \\ \frac{-\Sigma_x}{n \Sigma_x^2 - (\Sigma_x)^2} & \frac{n}{n \Sigma_x^2 - (\Sigma_x)^2} \end{pmatrix} \begin{pmatrix} \frac{\Sigma_x^2 \Sigma_y - \Sigma_x \Sigma_{xy}}{n \Sigma_x^2 - (\Sigma_x)^2} \\ \frac{-\Sigma_x \Sigma_y + n \Sigma_{xy}}{n \Sigma_x^2 - (\Sigma_x)^2} \end{pmatrix}$$

$$(\mathbf{P}^T\mathbf{P})^{-1} \quad (\mathbf{P}^T\mathbf{P})^{-1} \mathbf{P}^T\mathbf{y}$$

Daraus erhält man die Schätzung  $b_0$  des Achsenabschnittes  $\beta_0$ :

$$b_0 = \frac{\Sigma_x^2 \Sigma_y - \Sigma_x \Sigma_{xy}}{n \Sigma_x^2 - (\Sigma_x)^2} \quad (7.10)$$

sowie die Schätzung  $b_1$  der Steigung  $\beta_1$ :

$$b_1 = \frac{n \Sigma_{xy} - \Sigma_x \Sigma_y}{n \Sigma_x^2 - (\Sigma_x)^2} \quad (7.11)$$

## 7.4. Schätzung der Fehlervarianz

Den Vektor der geschätzten Messwerte  $\hat{\mathbf{y}}$  erhält man nach (6.19), den Vektor der Residuen nach (6.5). Die Schätzung der Fehlervarianz  $s^2$  berechnet man dann nach (6.6). Diese Rechnungen sind trivial. Die Anzahl Freiheitsgrade beträgt  $v = n - 2$ , weil 2 Lageparameter (Steigung und Achsenabschnitt) geschätzt werden.

## 7.5. Vertrauensbereich der Parameter

Die Schätzungen der Parameter sind Zufallsvariablen. Da die Parameter aus der gleichen Stichprobe geschätzt werden, sind sie im Allgemeinen korreliert.  $\sigma^2(\mathbf{P}^T\mathbf{P})^{-1}$  ist die (symmetrische) Varianz-Kovarianz-Matrix der Schätzungen. Die symmetrische Matrix  $\mathbf{R}$  der paarweisen Korrelationskoeffizienten ist nur von  $\mathbf{P}$  abhängig, demnach also (nach Modellannahme) exakt bekannt. Zu ihrer Berechnung wird zuerst die Diagonalmatrix  $\mathbf{D}$  definiert.  $\mathbf{D}$  hat die Diagonalelemente der Matrix  $(\mathbf{P}^T\mathbf{P})^{-1}$ . Aus  $\mathbf{D}$  berechnet man die Diagonalmatrix  $\mathbf{D}^{-1/2}$ . Sie hat die Diagonalelemente  $\sqrt{1/d_{ii}}$ , wobei  $d_{ii}$  die Diagonalelemente von  $\mathbf{D}$  sind. Damit kann  $\mathbf{R}$  berechnet werden:

$$\mathbf{R} = \mathbf{D}^{-1/2} (\mathbf{P}^T\mathbf{P})^{-1} \mathbf{D}^{-1/2} \quad (7.12)$$

Im Detail:

$$\mathbf{D} = \frac{1}{n \sum x^2 - (\sum x)^2} \begin{pmatrix} \sum x^2 & 0 \\ 0 & n \end{pmatrix} \quad (7.13)$$

$$\mathbf{D}^{1/2} = \frac{1}{\sqrt{n \sum x^2 - (\sum x)^2}} \begin{pmatrix} \sqrt{\sum x^2} & 0 \\ 0 & \sqrt{n} \end{pmatrix} \quad (7.14)$$

$$\mathbf{D}^{-1/2} = \sqrt{n \sum x^2 - (\sum x)^2} \begin{pmatrix} \frac{1}{\sqrt{\sum x^2}} & 0 \\ 0 & \frac{1}{\sqrt{n}} \end{pmatrix} \quad (7.15)$$

Für die Berechnung der Matrix  $\mathbf{D}^{-1/2} (\mathbf{P}^T\mathbf{P})^{-1} \mathbf{D}^{-1/2}$  ist es nicht notwendig, die Faktoren vor den einzelnen Matrizen zu berücksichtigen, da sie sich bei der Multiplikation wegheben. Es werden also nur die Matrizen selbst multipliziert.

1. Schritt:

$$\begin{pmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{pmatrix}$$

$$\begin{pmatrix} \frac{1}{\sqrt{\sum x^2}} & 0 \\ 0 & \frac{1}{\sqrt{n}} \end{pmatrix} \begin{pmatrix} \sqrt{\sum x^2} & \frac{-\sum x}{\sqrt{\sum x^2}} \\ \frac{-\sum x}{\sqrt{n}} & \sqrt{n} \end{pmatrix} \quad (7.16)$$

2. Schritt:

$$\begin{pmatrix} \frac{1}{\sqrt{\sum x^2}} & 0 \\ 0 & \frac{1}{\sqrt{n}} \end{pmatrix}$$

$$\begin{pmatrix} \sqrt{\sum x^2} & \frac{-\sum x}{\sqrt{\sum x^2}} \\ \frac{-\sum x}{\sqrt{n}} & \sqrt{n} \end{pmatrix} \begin{pmatrix} 1 & \frac{-\sum x}{\sqrt{n} \sum x^2} \\ \frac{-\sum x}{\sqrt{n} \sum x^2} & 1 \end{pmatrix} \quad (7.17)$$

Die Elemente von  $\mathbf{R}$  sind dimensionslose Zahlen aus dem Intervall  $[-1 \dots 1]$ . In der Diagonalen stehen Einsen. Im Fall der Linearen Regression gibt es nur ein einziges Ausserdiagonalelement:

$$\rho_{b_0 b_1} = \frac{-\sum x}{\sqrt{n} \sum x^2} \quad (7.18)$$

Individuelle Vertrauensbereiche für die Schätzungen der Parameter kann man so berechnen:

$$\mathbf{VB}_b = \mathbf{b} \pm t(n-p; 1-\alpha) s \mathbf{D}^{1/2} \mathbf{1}_p \quad (7.19)$$

Für die Lineare Regression:

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \mathbf{1}_2$$

$$\begin{pmatrix} \frac{\sqrt{\sum x^2}}{\sqrt{n \sum x^2 - (\sum x)^2}} & 0 \\ 0 & \frac{\sqrt{n}}{\sqrt{n \sum x^2 - (\sum x)^2}} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{\sum x^2}}{\sqrt{n \sum x^2 - (\sum x)^2}} \\ \frac{\sqrt{n}}{\sqrt{n \sum x^2 - (\sum x)^2}} \end{pmatrix}$$

$\mathbf{D}^{1/2} \qquad \mathbf{D}^{1/2} \mathbf{1}_2$

Damit erhält man für die Vertrauensbereiche:

$$\mathbf{VB}_{\mathbf{b}} = \mathbf{b} \pm \frac{s \cdot t(n-2; 1-\alpha)}{\sqrt{n \sum x^2 - (\sum x)^2}} \begin{pmatrix} \sqrt{\sum x^2} \\ \sqrt{n} \end{pmatrix} \quad (7.21)$$

## 7.6. VB für den Erwartungswert eines Messwertes bei ausgewählten x-Werten

Der Erwartungswert eines Messwertes  $\hat{y}$  bei ausgewählten Werten  $\mathbf{x}_0$  ist im Fall der Linearen Regression problemlos zu berechnen. Der Vektor  $\mathbf{d}$  nach (6.23) beträgt:

$$\mathbf{d} = \begin{pmatrix} 1 \\ x_0 \end{pmatrix} \quad (7.22)$$

Zur Berechnung von  $\mathbf{d}^T (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{d}$  wird der Faktor vor der Matrix  $(\mathbf{P}^T \mathbf{P})^{-1}$  der Übersichtlichkeit halber weggelassen.

1. Schritt:

$$\begin{pmatrix} \sum x^2 & -\sum x \\ -\sum x & n \end{pmatrix} \quad (7.23)$$

$$\begin{pmatrix} 1 & x_0 \end{pmatrix} \begin{pmatrix} \sum x^2 - x_0 \sum x & -\sum x + n x_0 \end{pmatrix}$$

2. Schritt:

$$\begin{pmatrix} 1 \\ x_0 \end{pmatrix}$$

$$\begin{pmatrix} \sum x^2 - x_0 \sum x & -\sum x + n x_0 \end{pmatrix} \begin{pmatrix} \sum x^2 - x_0 \sum x - x_0 \sum x + n x_0^2 \end{pmatrix}$$

Damit erhält man also für  $\mathbf{d}^T(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{d}$ :

$$\mathbf{d}^T(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{d} = \frac{\sum x^2 - 2x_0 \sum x + n x_0^2}{n \sum x^2 - (\sum x)^2} \quad (7.25)$$

Addiert man dazu die Gleichung

$$0 = \frac{1}{n} - \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n \sum x^2 - (\sum x)^2} \quad (7.26)$$

erhält man

$$\mathbf{d}^T(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{d} = \frac{1}{n} + \frac{\frac{(\sum x)^2}{n} - 2x_0 \sum x + n x_0^2}{n \sum x^2 - (\sum x)^2} \quad (7.27)$$



und daraus schliesslich

$$\mathbf{d}^T (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{d} = \frac{1}{n} + \frac{n \left( x_0 - \frac{\sum x}{n} \right)^2}{n \sum x^2 - (\sum x)^2} \quad (7.28)$$

Vertrauensbereich für die Schätzung des wahren Mittelwertes bei  $x_0$ :

$$\text{VB}_{\hat{y}} = b_0 + b_1 x_0 \pm t_{(n-2; 1-\alpha)} s \sqrt{\frac{1}{n} + \frac{n \left( x_0 - \frac{\sum x}{n} \right)^2}{n \sum x^2 - (\sum x)^2}}$$

Toleranzbereich für die Voraussage eines einzelnen Messwertes bei  $x_0$ :

$$\text{TB}_{\hat{y}} = b_0 + b_1 x_0 \pm t_{(n-2; 1-\alpha)} s \sqrt{1 + \frac{1}{n} + \frac{n \left( x_0 - \frac{\sum x}{n} \right)^2}{n \sum x^2 - (\sum x)^2}}$$

Vertrauensbereich für die Voraussage eines Mittelwertes von  $m$  Messwerten bei  $x_0$ :

$$\text{VB}_{\hat{y}} = b_0 + b_1 x_0 \pm t_{(n-2; 1-\alpha)} s \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{n \left( x_0 - \frac{\sum x}{n} \right)^2}{n \sum x^2 - (\sum x)^2}}$$

## 7.7. Vertrauensbereich für einen Wert $x$ aus Messungen von $y$

In der analytischen Chemie will man häufig von einem Messwert  $y$  auf das zugehörige  $x$  schliessen. Beispiel: Man hat eine Kalibrationsgerade aus  $n$  Wertepaaren  $x_i, y_i$  berechnet. Jetzt misst man bei unbekanntem, aber festem  $x$  einige  $y$ -Werte. Wie gross ist der zugehörige  $x$ -Wert?

Die  $m$  Einzelwerte werden zu einem Mittelwert zusammengefasst:

$$\bar{y} = \frac{1}{m} \sum y \quad (7.29)$$

Die Schätzung  $\hat{x}$  hängt nach Gleichung 7.1 mit  $\bar{y}$  zusammen, wobei die Parameter durch ihre Schätzungen ersetzt werden:

$$\bar{y} = b_0 + b_1 \hat{x}$$

Nach  $\hat{x}$  aufgelöst ergibt sich:

$$\hat{x} = \frac{\bar{y} - b_0}{b_1}$$

Alle drei Grössen auf der rechten Seite sind fehlerbehaftet. Geht man von Gleichung (4.16) aus, stellt sich die Frage nach den Korrelationskoeffizienten in den Kreuztermen. Die  $y$ -Werte in (4.29) schätzen einen eigenen Lageparameter  $\bar{y}$ , deshalb sind seine Fehler nicht mit jenen von Steigung und Achsenabschnitt korreliert. (Das gilt auch dann, wenn die  $y$ -Werte zur Schätzung der Gesamtvarianz  $s^2$  der Messfehler mitverwendet wurden.) Die Korrelationskoeffizienten mit  $\bar{y}$  verschwinden also und damit die Kreuzterme. Es bleibt aber die Korrelation zwischen Achsenabschnitt und Steigung. Damit ergibt sich:

$$s_{\hat{x}}^2 = \left( \frac{d\hat{x}}{d\bar{y}} \right)^2 s_{\bar{y}}^2 + \left( \frac{d\hat{x}}{db_0} \right)^2 s_{b_0}^2 + \left( \frac{d\hat{x}}{db_1} \right)^2 s_{b_1}^2 + 2 \left( \frac{d\hat{x}}{db_0} \right) \left( \frac{d\hat{x}}{db_1} \right) \rho_{b_0 b_1} s_{b_0} s_{b_1}$$

Der Korrelationskoeffizient gemäss (7.18) ist exakt bekannt. Um den Kreuzterm zum Verschwinden zu bringen, wird eine Koordinatentransformation durchgeführt. Statt den Werten  $x$  werden die transformierten Werte  $u$  verwendet:

$$u = x - \frac{\sum x}{n} \quad (7.30)$$

An dieser Stelle wird angenommen, die Schätzung der Parameter sei im neuen Koordinatensystem durchgeführt worden. Das betrifft insbesondere den Achsenabschnitt  $b_0$ . Die Steigung  $b_1$  ändert sich nicht. Es ist also

$$\bar{y} = b_0 + b_1 \hat{u} \quad (7.31)$$

$$\hat{u} = \frac{\bar{y} - b_0}{b_1}$$

Als Konsequenz der Transformation verschwindet der Zähler in (7.18) und damit der Korrelationskoeffizient.

$$\sum u = 0 \quad (7.32)$$

Nach der Transformation bleibt:

$$s_{\hat{u}}^2 = \left( \frac{d\hat{u}}{d\bar{y}} \right)^2 s_{\bar{y}}^2 + \left( \frac{d\hat{u}}{db_0} \right)^2 s_{b_0}^2 + \left( \frac{d\hat{u}}{db_1} \right)^2 s_{b_1}^2 \quad (7.33)$$

Die Ableitungen:

$$\begin{aligned} \frac{d\hat{u}}{d\bar{y}} &= \frac{d}{d\bar{y}} \left( \frac{\bar{y} - b_0}{b_1} \right) = \frac{d}{d\bar{y}} \left( \frac{\bar{y}}{b_1} - \frac{b_0}{b_1} \right) = \frac{d}{d\bar{y}} \left( \frac{\bar{y}}{b_1} \right) - \frac{d}{d\bar{y}} \left( \frac{b_0}{b_1} \right) \\ &= \frac{1}{b_1} - 0 = \frac{1}{b_1} \end{aligned}$$

$$\begin{aligned} \frac{d\hat{u}}{db_0} &= \frac{d}{db_0} \left( \frac{\bar{y} - b_0}{b_1} \right) = \frac{d}{db_0} \left( \frac{\bar{y}}{b_1} - \frac{b_0}{b_1} \right) = \frac{d}{db_0} \left( \frac{\bar{y}}{b_1} \right) - \frac{d}{db_0} \left( \frac{b_0}{b_1} \right) \\ &= 0 - \frac{1}{b_1} = -\frac{1}{b_1} \end{aligned}$$

$$\begin{aligned} \frac{d\hat{u}}{db_1} &= \frac{d}{db_1} \left( \frac{\bar{y} - b_0}{b_1} \right) = \frac{d}{db_1} \left( (\bar{y} - b_0) b_1^{-1} \right) = (\bar{y} - b_0) \frac{d}{db_1} \left( b_1^{-1} \right) \\ &= (\bar{y} - b_0) (-1) b_1^{-2} = -\frac{\bar{y} - b_0}{b_1^2} \end{aligned}$$

Ersetzt man  $\bar{y}$  durch die rechte Seite von (7.31), ergibt sich:

$$\frac{d\hat{u}}{db_1} = -\frac{\hat{u}}{b_1}$$

Die Varianz von  $\bar{y}$  ergibt sich aus der Varianz der Messfehler  $s^2$ :

$$s_{\bar{y}}^2 = \frac{s^2}{m}$$

Die Varianz der Parameter-Schätzungen erhält man aus (7.21), wobei  $x$  durch  $u$  zu ersetzen ist. Berücksichtigt man noch (7.32), erhält man:

$$s_{b_0}^2 = \frac{s^2}{n \sum u^2} \sum u^2 = \frac{s^2}{n}$$

$$s_{b_1}^2 = \frac{s^2}{n \sum u^2} n = \frac{s^2}{\sum u^2}$$

Setzt man diese Terme in (7.33) ein, erhält man

$$\begin{aligned}
 s_{\hat{u}}^2 &= \left(\frac{1}{b_1}\right)^2 \frac{s^2}{m} + \left(-\frac{1}{b_1}\right)^2 \frac{s^2}{n} + \left(-\frac{\hat{u}}{b_1}\right)^2 \frac{s^2}{\sum u^2} \\
 &= \frac{s^2}{b_1^2} \left( \frac{1}{m} + \frac{1}{n} + \frac{\hat{u}}{\sum u^2} \right) \quad (7.34)
 \end{aligned}$$

Nun wird die Koordinaten-Transformation (7.30) wieder rückgängig gemacht. Die Grösse  $u$  wird also wieder durch  $x$  ausgedrückt.

$$\begin{aligned}
 \hat{u} &= \hat{x} - \frac{\sum x}{n} \\
 \sum u^2 &= \sum \left( x - \frac{\sum x}{n} \right)^2 = \sum \left( x^2 - 2x \frac{\sum x}{n} + \left( \frac{\sum x}{n} \right)^2 \right) \\
 &= \sum x^2 - 2 \sum \left( x \frac{\sum x}{n} \right) + \sum \left( \frac{\sum x}{n} \right)^2 \\
 &= \sum x^2 - 2 \frac{\sum x}{n} \sum x + n \left( \frac{\sum x}{n} \right)^2 = \sum x^2 - \frac{(\sum x)^2}{n} \\
 \frac{\hat{u}}{\sum u^2} &= \frac{\hat{x} - \frac{\sum x}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{n \hat{x} - \sum x}{n \sum x^2 - (\sum x)^2}
 \end{aligned}$$

Setzt man diesen Term ein in (7.34), ergibt sich

$$s_{\hat{x}}^2 = \frac{s^2}{b_1^2} \left( \frac{1}{m} + \frac{1}{n} + \frac{n \hat{x} - \sum x}{n \sum x^2 - (\sum x)^2} \right)$$

Für den Vertrauensbereich erhält man

$$\text{VB}_{\hat{x}} = \hat{x} \pm \frac{s}{b_1} t_{(n-2; 1-\alpha)} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{n \hat{x} - \sum x}{n \sum x^2 - (\sum x)^2}} \quad (7.35)$$

## 8. Sammlung statistischer Modelle

### 8.1. Proportionaler Zusammenhang mit Lognormalverteilung

Annahmen:

Es besteht der Zusammenhang  $y = \beta x$   $x \geq 0$

Die  $y$ -Werte sind fehlerbehaftet, die  $x$ -Werte können exakt eingestellt werden (bei der Kalibration) oder sind unbekannt (bei der nachfolgenden Messung).

$\beta$  ist ein Parameter, den man schätzen will. Die Grössen  $\frac{y_i}{x_i}$  sind lognormalverteilt, d. h. ihre Logarithmen sind normalverteilt (mit konstanter Standardabweichung und ohne serielle Korrelation).

Die Streuungen zeigen folgendes Bild:

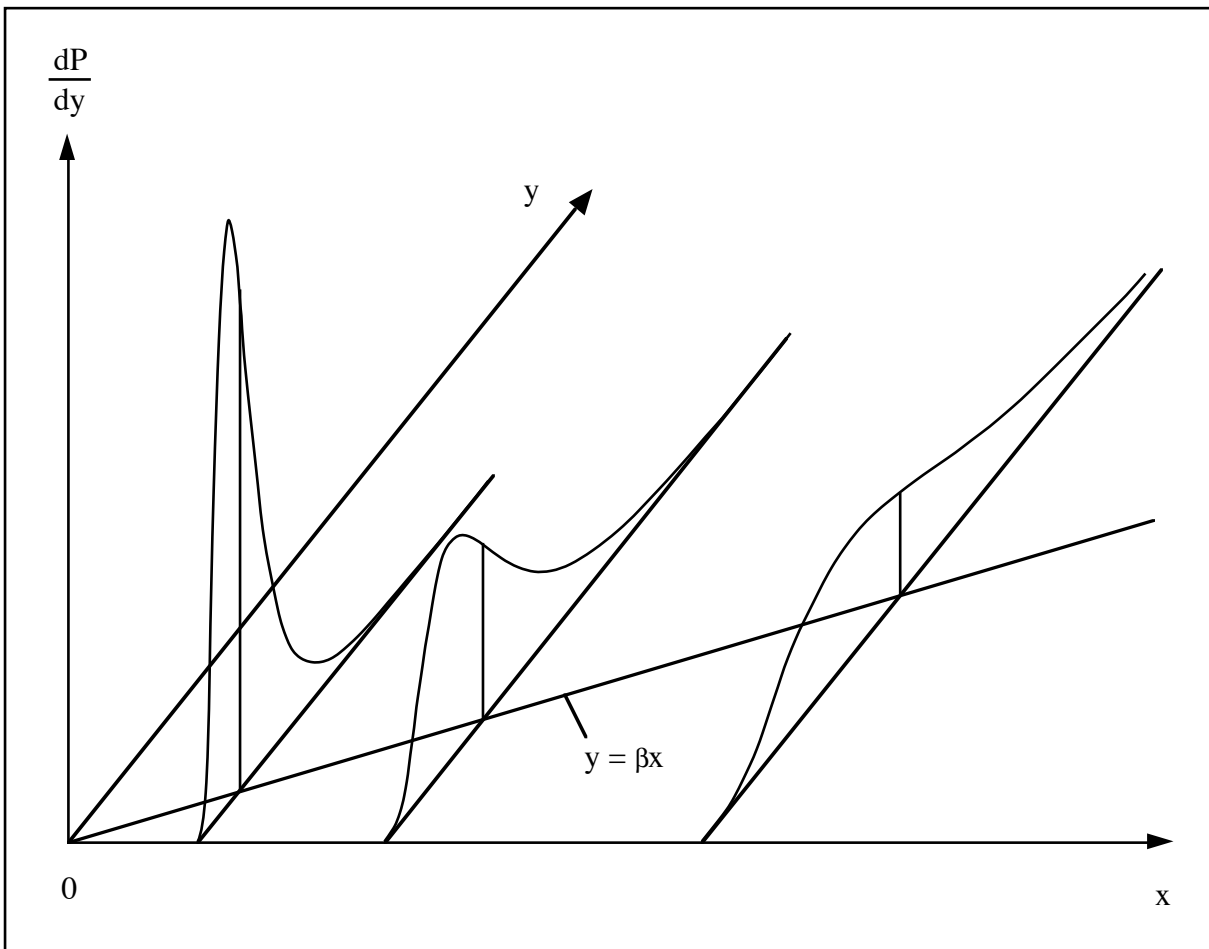


Fig. 8.1 Proportionaler Zusammenhang, Lognormalverteilung

Das Modell ist vor allem geeignet, wenn

- die Messwerte nur positiv sein können
- der relative (prozentuale) Messfehler der  $y$ -Werte unabhängig von  $x$  ist
- Mittelwert und Streuung der Messwerte  $y_i$  ähnlich gross sind ( $x$  fest). Dies ist z. B. in der Nähe der Nachweisgrenze der Fall.

Nur die beiden ersten Punkte sind als Bedingung zu verstehen. Der dritte Punkt wird hier nur aufgeführt, weil grosse relative Messfehler bei diesem Modell überhaupt nicht stören.

Zudem ist bemerkenswert, dass es bei diesem Modell keine Hebelpunkte geben kann, da die Punkte gewissermassen mit ihrem reziproken Hebelarm gewichtet sind.

Schätzung der Parameter:

Man setzt zur Abkürzung:

$$b_i^* = \ln\left(\frac{y_i}{x_i}\right) \quad (8.1)$$

Der Stern soll anzeigen, dass die Grösse logarithmisch ist. Das Logarithmieren dimensionsbehafteter Grössen ist problematisch. Man lässt die Einheit einfach weg und setzt sie beim späteren Entlogarithmieren wieder ein.

Gemäss Modell stammen die  $b_i^*$  aus einer normalverteilten Grundgesamtheit. Nennt man die beiden Parameter  $\beta^*$  und  $\sigma_b^*$ , erhält man für ihre Schätzungen mit den Gleichungen (4.2) ff.:

$$\bar{b}^* = \frac{1}{n} \sum_{i=1}^n b_i^* \quad n: \text{Anzahl Wertepaare } x_i, y_i \quad (8.2)$$

Standardabweichung der Einzelmessungen:

$$s_b^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (b_i^* - \bar{b}^*)^2} \quad (8.3)$$

Standardabweichung des Mittelwertes:

$$s_{\bar{b}}^* = \frac{s_b^*}{\sqrt{n}} \quad (8.4)$$

Für beide Standardabweichungen gilt:

$$v = n - 1 \quad (8.5)$$

Alle diese Grössen sind nicht sehr anschaulich. Erst beim Entlogarithmieren bekommen sie einen Sinn.

Schätzung für die Steigung  $\beta$ :

$$b = \exp\left[\bar{b}^*\right] \quad (8.6)$$

Die Fehlerschranken sind nicht symmetrisch bezüglich des Mittelwertes.

Vertrauensbereich  $VB_b$  der Steigung:

$$\begin{aligned} \text{untere Schranke:} & \quad b \cdot \exp\left[-t_v s_{\bar{b}}^*\right] \\ \text{obere Schranke:} & \quad b \cdot \exp\left[t_v s_{\bar{b}}^*\right] \end{aligned} \quad (8.7)$$

Schätzung  $\hat{y}$  für den Erwartungswert  $y$  eines Messwertes:

$$\hat{y} = bx \quad (\text{Ausgleichsgerade}) \quad (8.8)$$

Toleranzbereich  $TB_y$  eines einzelnen Messwertes:

$$\begin{aligned} \text{untere Schranke: } & \hat{y} \cdot \exp\left[-t_v s_b^*\right] \\ \text{obere Schranke: } & \hat{y} \cdot \exp\left[t_v s_b^*\right] \end{aligned} \quad (8.9)$$

Vertrauensbereich  $VB_{\hat{y}}$  des Mittelwertes  $\hat{y}$ :

$$\begin{aligned} \text{untere Schranke: } & \hat{y} \cdot \exp\left[-t_v s_{\bar{b}}^*\right] \\ \text{obere Schranke: } & \hat{y} \cdot \exp\left[t_v s_{\bar{b}}^*\right] \end{aligned} \quad (8.10)$$

Die Länge der Hebelarme ist in diesem Modell für alle Punkte gleich. In der Residuenanalyse trägt man die Residuen im logarithmierten Raum gegen beliebige Variablen auf.

In der analytischen Chemie will man häufig von einem Messwert  $y$  auf das zugehörige  $x$  schliessen. Beispiel: Man hat eine Kalibrationsgerade aus  $n$  Wertepaaren  $x_i, y_i$  berechnet. Jetzt misst man bei unbekanntem, aber festem  $x$  einige  $y$ -Werte. Wie gross ist der zugehörige  $x$ -Wert?

Schätzung für  $x$ :

$$\hat{x} = \frac{1}{b} \exp\left[\frac{1}{m} \sum_{j=1}^m \ln y_j\right] = \frac{1}{b} \left[\prod_{j=1}^m y_j\right]^{\frac{1}{m}} \quad (8.11)$$

$m$ : Anzahl  $y_j$ -Werte

Vertrauensbereich  $VB_{\hat{x}}$  für  $\hat{x}$ :

$$\begin{aligned} \text{untere Schranke: } & \hat{x} \cdot \exp\left[-t_v s_b^* \sqrt{\frac{1}{m} + \frac{1}{n}}\right] \\ \text{obere Schranke: } & \hat{x} \cdot \exp\left[t_v s_b^* \sqrt{\frac{1}{m} + \frac{1}{n}}\right] \end{aligned} \quad (8.12)$$

Anzahl Freiheitsgrade  $v = n - 1$

## Rechenbeispiel:

Die Konzentration von Blei in Regenwasser wird mit der Methode AAS/Graphitrohrküvette bestimmt. 50  $\mu\text{l}$  Regenwasser werden direkt ins Graphitrohr einpipettiert. Die Höhe des Absorptionssignals dient als Mass für die Konzentration. Es werden 4 Kalibrationslösungen mit bekannten Bleigehalten verwendet. Jede Lösung wird dreimal eingespritzt.

Das Ergebnis:

Konzentration $c_i$ $\mu\text{g/l}$	Signalhöhe $S_i$ mm
20	18,7
20	14,3
20	15,2
50	42,7
50	47,5
50	45,3
100	89,3
100	70,4
100	72,1
200	169,8
200	139,0
200	148,8
Regenwasser:	33,3
Regenwasser:	31,6
Regenwasser:	27,4
Bidest. Wasser:	< 0,5

Eine Blindprobe mit bidestilliertem Wasser zeigt, dass die Kalibrationskurve durch den Nullpunkt geht. Die Signalthöhen können prinzipiell nicht negativ werden. Zudem weiss man aus Erfahrung, dass die relativen Messfehler im Mittel etwa 10-20 % der Signalthöhe betragen. Die Wahl des Modells scheint also gerechtfertigt.

## Kalibration:

Konzentration $c_i$ $\mu\text{g/l}$	Signalhöhe $S_i$ mm	$b_i^* = \ln \frac{S_i}{c_i}$
20	18,7	-0,06721
20	14,3	-0,33547
20	15,2	-0,27444
50	42,7	-0,15782
50	47,5	-0,05129
50	45,3	-0,09872
100	89,3	-0,11317
100	70,4	-0,35098
100	72,1	-0,32712
200	169,8	-0,16370
200	139,0	-0,36384
200	148,8	-0,29571



$$\bar{b}^* = \frac{1}{12}(-2,59947) = -0,21662$$

$$s_b^* = \sqrt{\frac{0,15618}{11}} = 0,11916$$

$$s_{\bar{b}}^* = \frac{0,11916}{\sqrt{12}} = 0,034397$$

$$b = \exp(-0,21662) = 0,8052 \frac{\text{mm}}{\mu\text{g/l}}$$

$$\hat{S} = 0,8052 \frac{\text{mm}}{\mu\text{g/l}} \cdot c \quad (\text{Kalibrationsgerade}) \quad (\text{mit Gleichungen (8.1) bis (8.8)})$$

$$t_{11}(95\%) = 2,201 \quad (\text{aus einer t-Tabelle})$$

Toleranzbereich  $TB_S$  eines einzelnen Messwertes mit Gleichung (8.9):

$$\text{untere Schranke: } 0,8052 \frac{\text{mm}}{\mu\text{g/l}} \cdot c \cdot \exp(-2,201 \cdot 0,11916) = 0,6195 \frac{\text{mm}}{\mu\text{g/l}} \cdot c$$

$$\text{obere Schranke: } 1,0467 \frac{\text{mm}}{\mu\text{g/l}} \cdot c$$

Vertrauensbereich  $VB_{\hat{S}}$  des Mittelwertes  $\hat{S}$  mit Gleichung (8.10):

$$\text{untere Schranke: } 0,8052 \frac{\text{mm}}{\mu\text{g/l}} \cdot c \cdot \exp(-2,201 \cdot 0,034397) = 0,7465 \frac{\text{mm}}{\mu\text{g/l}} \cdot c$$

$$\text{obere Schranke: } 0,8686 \frac{\text{mm}}{\mu\text{g/l}} \cdot c$$

Die Hebelarme sind in diesem Modell alle gleich. Daher brauchen die Residuen nicht normiert zu werden. Für Residuenplots sind die Residuen  $b_i^* - \bar{b}^*$  im logarithmischen Raum zu verwenden:

Konzentration $c_i$ mg/l	Residuum
20	0,1494
20	-0,1189
20	-0,0578
50	0,0588
50	0,1653
50	0,1179
100	0,1035
100	-0,1344
100	-0,1105
200	0,0529
200	-0,1472
200	-0,0791

Schätzung der Bleikonzentration im Regenwasser mit Gleichung (8.11):  
(mittlere senkrechte Linie im Diagramm)

$$\hat{c} = \frac{1}{0,8052 \frac{\text{mm}}{\mu\text{g/l}}} \sqrt[3]{33,3 \text{ mm} \cdot 31,6 \text{ mm} \cdot 27,4 \text{ mm}} = 38,081 \mu\text{g/l}$$

Vertrauensbereich  $\text{VB}_{\hat{c}}$  von  $\hat{c}$  mit Gleichung (8.12):

$$\text{untere Schranke: } 38,081 \mu\text{g/l} \cdot \exp\left(-2,201 \cdot 0,11916 \sqrt{\frac{1}{3} + \frac{1}{12}}\right) = 32,150 \mu\text{g/l}$$

(linke senkrechte Linie im Diagramm)

$$\text{obere Schranke: } 45,105 \mu\text{g/l} \quad (\text{rechte senkrechte Linie im Diagramm})$$

Schlussresultat:  $\hat{c} = 38,1 \mu\text{g/l}$   
 obere Schranke:  $45,1 \mu\text{g/l}$   
 untere Schranke:  $32,1 \mu\text{g/l}$   
 statistische Sicherheit: 95 %

Graphische Darstellung der Messwerte:

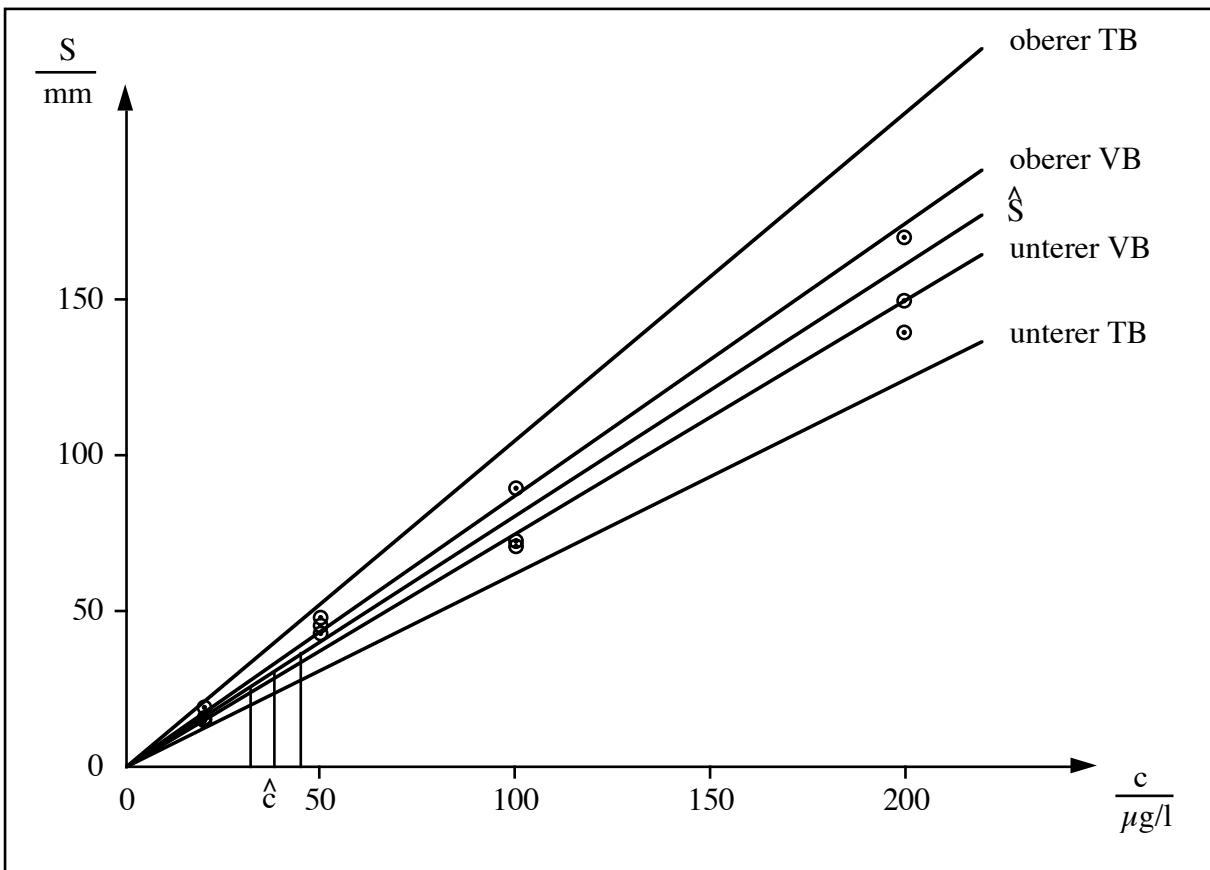


Fig. 8.2 Rechenbeispiel proportionaler Zusammenhang, Lognormalverteilung

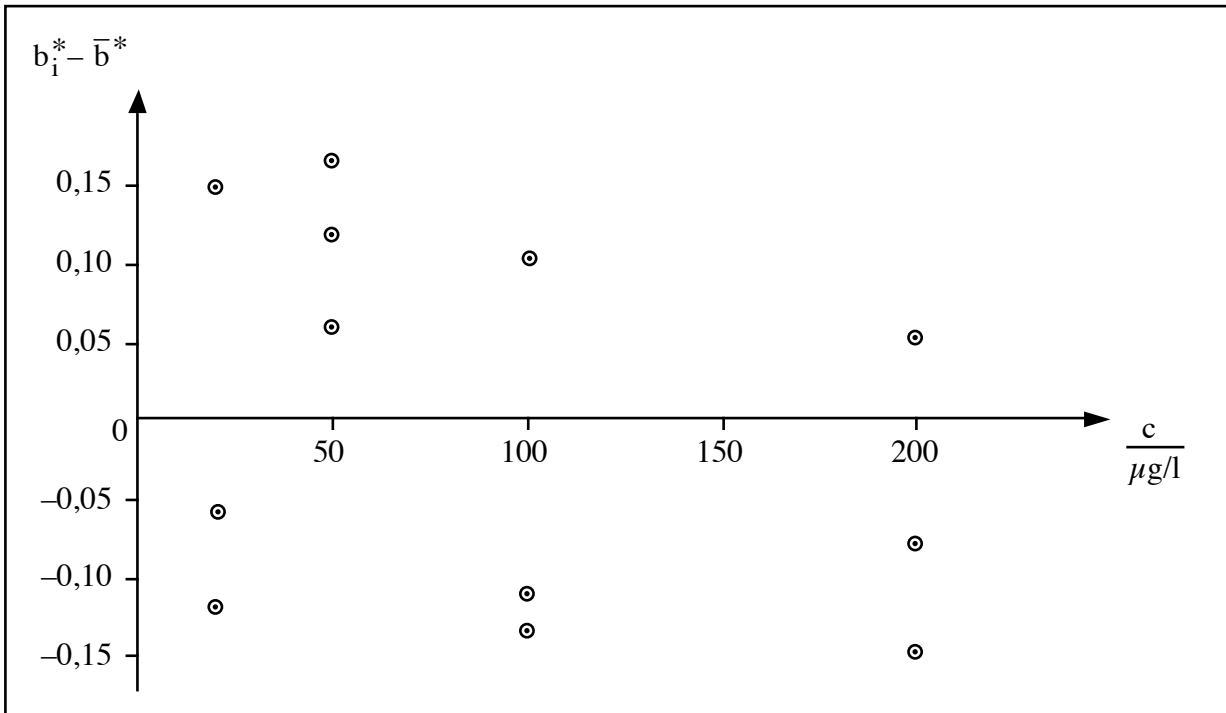


Fig. 8.3 Logarithmische Residuen versus Konzentration

## 8.2. Proportionaler Zusammenhang mit Normalverteilung

Annahmen:

Es besteht der Zusammenhang:  $y = \beta x$ .

Die  $y$ -Werte sind fehlerbehaftet, die  $x$ -Werte können exakt eingestellt werden (bei der Kalibration) oder sind unbekannt (bei der nachfolgenden Messung).

$\beta$  ist ein Parameter, den man schätzen will. Die Messfehler von  $y$  sind unabhängig von  $x$ , normalverteilt mit Mittelwert null und ohne serielle Korrelation.

Die Streuungen zeigen folgendes Bild:

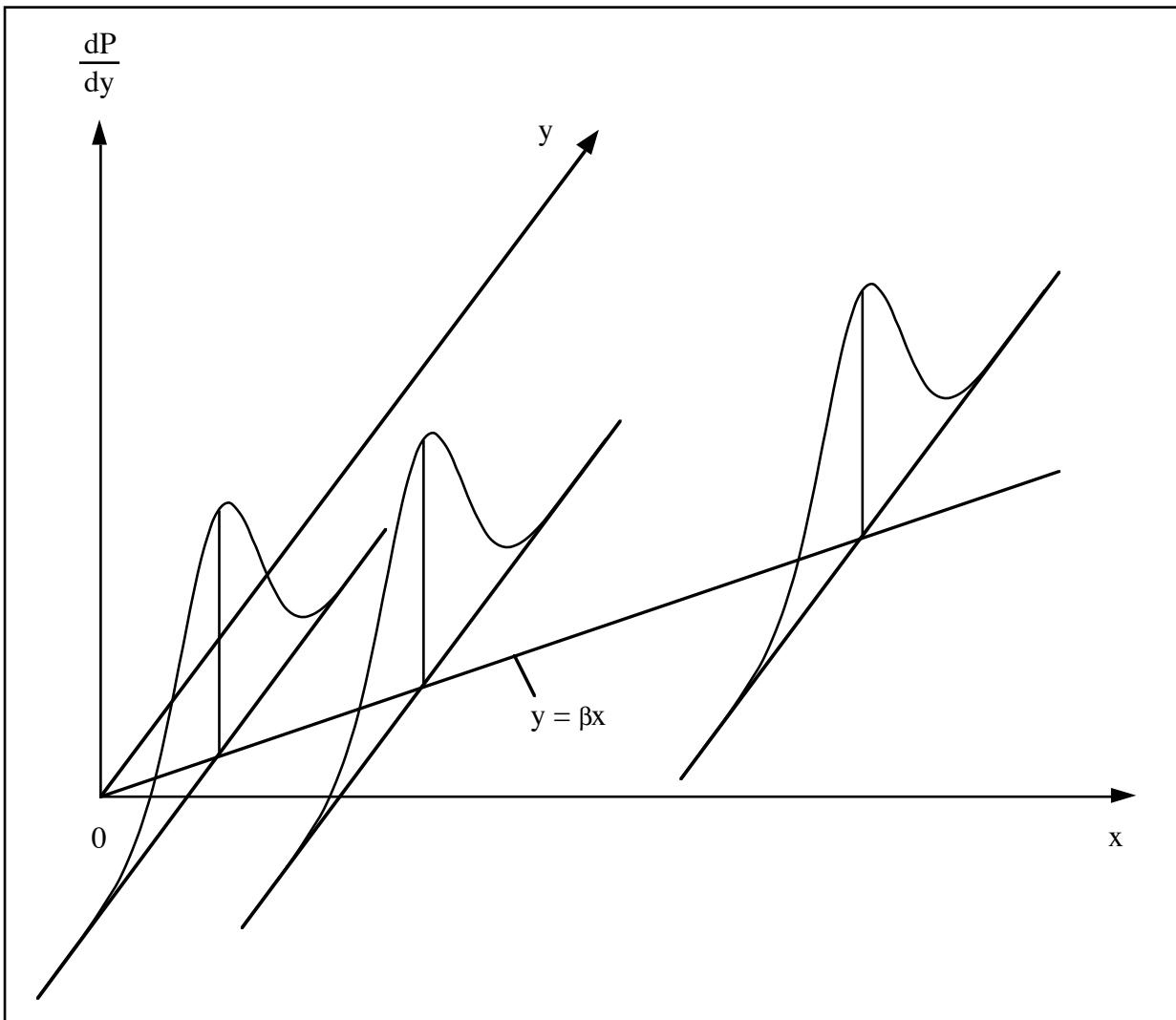


Fig. 8.4 Proportionaler Zusammenhang, Normalverteilung

Bei diesem Modell kann es vorkommen, dass Messwerte negativ sind, obwohl man weiss, dass der zugehörige Erwartungswert positiv ist. Solche Werte dürfen auf keinen Fall korrigiert werden.

Schätzung der Parameter:

$$\text{Hilfsgrößen: } \overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2 \quad \overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i \quad (8.13)$$

Standardabweichung der Messfehler von y:

$$s = \sqrt{\frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 - \frac{n \bar{xy}^2}{\bar{x}^2} \right]} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - bx_i)^2} \quad (8.14)$$

n : Anzahl Wertepaare  $x_i, y_i$

Anzahl Freiheitsgrade  $v = n - 1$

Steigung:

$$b = \frac{\bar{xy}}{\bar{x}^2} \quad (8.15)$$

Vertrauensbereich  $VB_b$  der Steigung b:

$$s_b = \frac{s}{\sqrt{n \bar{x}^2}} \quad (8.16)$$

$$VB_b = b \pm t_v s_b \quad (8.17)$$

Schätzung für den Erwartungswert eines Messwertes:

$$\hat{y} = bx \quad (\text{Ausgleichsgerade}) \quad (8.18)$$

Toleranzbereich  $TB_y$  eines einzelnen Messwertes:

$$TB_y = \hat{y} \pm t_v \sqrt{s^2 + x^2 s_b^2} = \hat{y} \pm t_v s \sqrt{1 + \frac{x^2}{n \bar{x}^2}} \quad (8.19)$$

Vertrauensbereich  $VB_{\hat{y}}$  des Mittelwertes  $\hat{y}$ :

$$VB_{\hat{y}} = \hat{y} \pm t_v s_b x \quad (8.20)$$

Die Länge des Hebelarmes für den i-ten Messwert:

$$\frac{1}{\sqrt{1 - \frac{x_i^2}{n \bar{x}^2}}} \quad (8.21)$$

In der analytischen Chemie will man häufig von einem Messwert y auf das zugehörige x schliessen. Beispiel: Man hat eine Kalibrationsgerade aus n Wertepaaren  $x_i, y_i$  berechnet. Jetzt misst man bei unbekanntem, aber festem x einige y-Werte. Wie gross ist der zugehörige x-Wert?

Schätzung für  $x$  :

$$\hat{x} = \frac{1}{b} \frac{1}{m} \sum_{j=1}^m y_j \quad (8.22)$$

$m$ : Anzahl  $y_j$ -Werte

Vertrauensbereich für  $\hat{x}$ :

$$VB_{\hat{x}} = \hat{x} \pm t_v \frac{1}{b} \sqrt{\frac{s^2}{m} + \hat{x}^2 s_b^2} \quad (8.23)$$

Anzahl Freiheitsgrade  $v = n - 1$

Rechenbeispiel:

In einem Apfelsaft wird photometrisch der Eisengehalt bestimmt. Für die Kalibration werden sechs Standardlösungen genau so aufgearbeitet wie die Probe. Ergebnis:

Konzentration $c_i$ mg/l	Extinktion $E_i$
3,27	0,331
5,84	0,376
9,71	0,701
12,36	0,973
15,08	1,255
17,26	1,383
Apfelsaft:	0,619
	0,526

Kalibration:

Hilfsgrößen:  $\overline{c^2} = \frac{1}{6} \sum_{i=1}^6 c_i^2 = \frac{1}{6} 817,17 \text{ (mg/l)}^2 = 136,19$

$$\overline{cE} = \frac{1}{6} \sum_{i=1}^6 c_i E_i = \frac{1}{6} 64,907 \text{ mg/l} = 10,818 \quad \text{nach (8.13)}$$

Schätzung der Parameter:

$$b = \frac{\overline{cE}}{\overline{c^2}} = \frac{10,818 \text{ mg/l}}{136,19 \text{ (mg/l)}^2} = 0,07943 \text{ l/mg} \quad \text{nach (8.15)}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^6 (E_i - bc_i)^2} = 0,06516 \quad \text{nach (8.14)}$$

$$t_{5,95\%} = 2,571 \quad (\text{aus einer t-Tabelle})$$

$$s_b = \frac{0,06516}{\sqrt{817,17 \text{ (mg/l)}^2}} = 0,002279 \text{ l/mg} \quad \text{nach (8.16)}$$

$$VB_b = (7,94 \pm 0,59) \cdot 10^{-2} \text{ l/mg} \quad \text{nach (8.17)}$$

$$\hat{E} = 0,07943 \text{ l/mg} \cdot c \quad \text{Kalibrationsgerade, Gleichung (8.18)}$$

Toleranzbereich  $TB_E$  einer einzelnen Messung:

$$TB_E = \hat{E} \pm 0,1675 \sqrt{1 + \frac{c^2}{817,17 \text{ (mg/l)}^2}} \quad \text{nach (8.19)}$$

Vertrauensbereich  $VB_{\hat{E}}$  des Mittelwertes  $\hat{E}$ :

$$VB_{\hat{E}} = \hat{E} \pm 0,005861 \text{ l/mg} \cdot c \quad \text{nach (8.20)}$$

Die Hebelarme, Residuen und mit den Hebelarmen normierten Residuen:

$\frac{\text{Konzentration } c_i}{\text{mg/l}}$	Hebelarm	Residuum	Residuum normiert
3,27	1,0066	0,0713	0,0717
5,84	1,0215	-0,0879	-0,0898
9,71	1,0632	-0,0703	-0,0747
12,36	1,1090	-0,0087	-0,0097
15,08	1,1771	0,0572	0,0673
17,26	1,2545	0,0120	0,0151

Schätzung der Eisenkonzentration  $\hat{c}$  im Apfelsaft:  
(mittlere senkrechte Linie im Diagramm)

$$\hat{c} = \frac{1}{0,07943 \text{ l/mg} \cdot 2} (0,619 + 0,526) = 7,208 \text{ mg/l} \quad \text{nach (8.22)}$$

Vertrauensbereich  $VB_{\hat{c}}$  des Mittelwertes  $\hat{c}$ :  
(linke und rechte senkrechte Linien im Diagramm)

$$VB_{\hat{c}} = 7,208 \text{ mg/l} \pm 32,37 \text{ mg/l} \sqrt{0,002393} \quad \text{nach (8.23)}$$

**Schlussresultat:**  $VB_{\hat{c}} = (7,21 \pm 1,59) \text{ mg/l} \quad (95 \%)$

Graphische Darstellung der Messwerte:

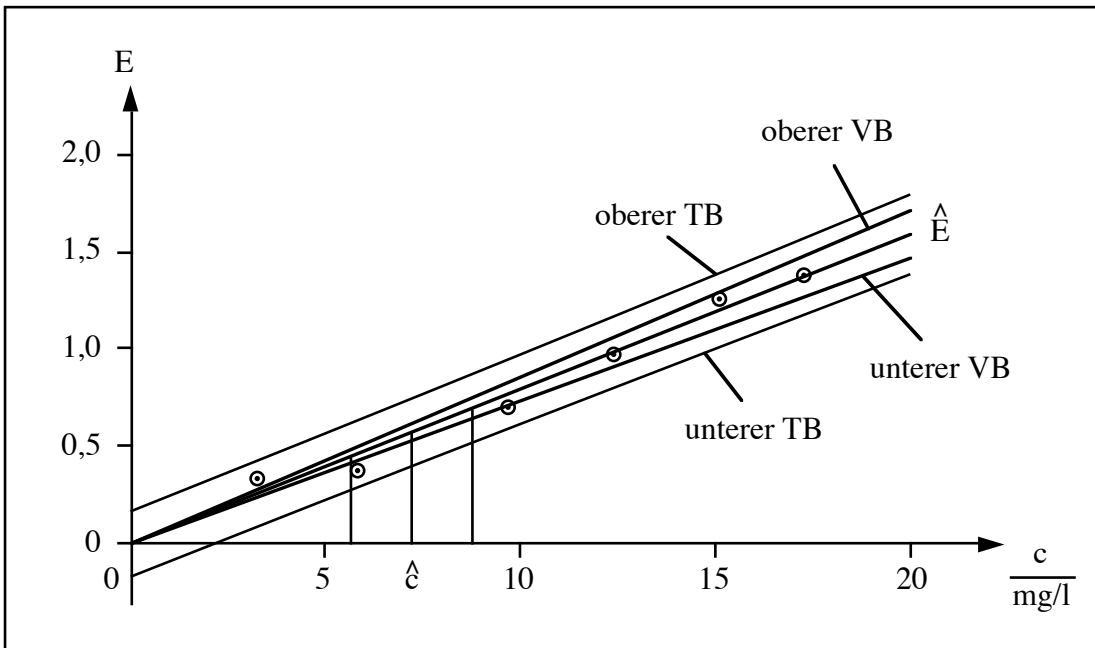


Fig. 8.5 Rechenbeispiel proportionaler Zusammenhang, Normalverteilung

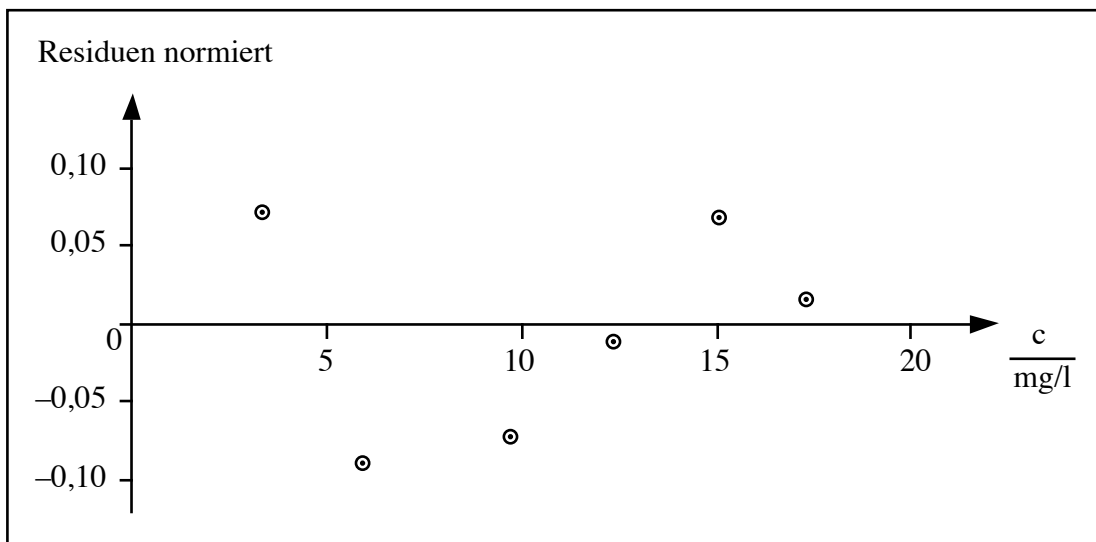


Fig. 8.6 Normierte Residuen versus Konzentration



### 8.3. Linearer Zusammenhang mit Normalverteilung

Dieses Modell heisst auch "Lineare Regression 1. Art". Vergleiche Kapitel 7.

Annahmen:

Es besteht der Zusammenhang  $y = \beta x + \alpha$

Die y-Werte sind fehlerbehaftet, die x-Werte können exakt eingestellt werden (bei der Kalibration) oder sind unbekannt (bei der nachfolgenden Messung).

$\alpha$  und  $\beta$  sind Parameter, die man schätzen will. Die Messfehler der y-Werte sind unabhängig von x, normalverteilt mit Mittelwert null und ohne serielle Korrelation.

Die Streuungen zeigen folgendes Bild:

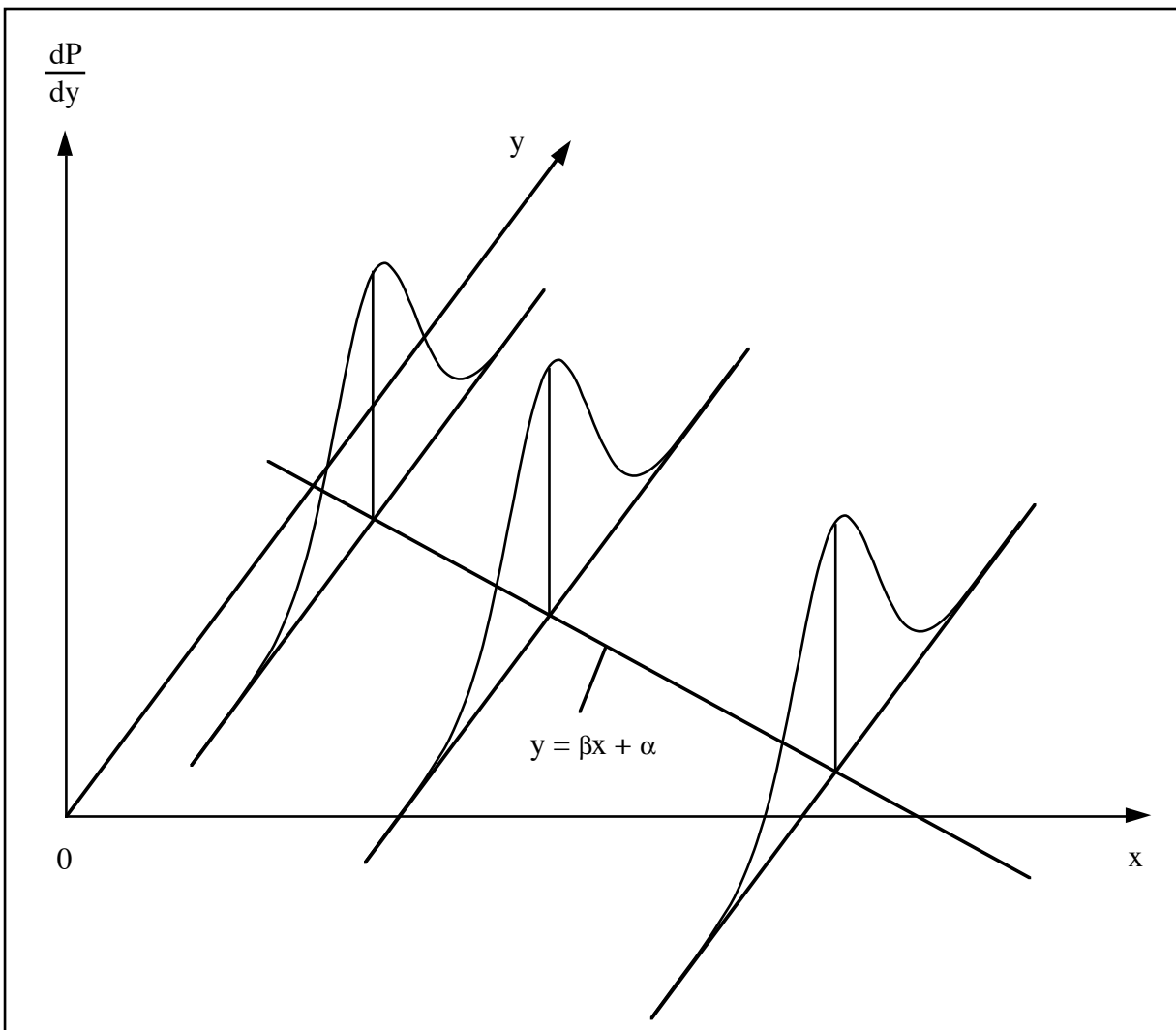


Fig. 8.7 Linearer Zusammenhang, Normalverteilung

Wichtige Anmerkung:

Man verwechsle dieses Modell nicht mit der linearen Regression 2. Art, bei der eine Normalverteilung für die x- und die y-Werte angenommen wird. Wir aber wollen x frei wählen und das zugehörige y messen. Das sind ganz andere Voraussetzungen. Zufälligerweise sind einige Formeln bei beiden Modellen gleich, was oft zu Verwechslungen führt. Namentlich hat ein Korrelationskoeffizient als Mass für den Zusammenhang zwischen x und y keine Bedeutung.

Schätzung der Parameter:

$$\begin{aligned} \text{Hilfsgrößen: } \quad \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i & \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i & \overline{x^2} &= \frac{1}{n} \sum_{i=1}^n x_i^2 \\ SS_x &= \sum_{i=1}^n (x_i - \bar{x})^2 & & & \text{SS steht für "Sum of squares"} \\ SS_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \end{aligned} \quad (8.24)$$

$$\text{Steigung: } \quad b = \frac{SS_{xy}}{SS_x} \quad (8.25)$$

$$\text{Achsenabschnitt: } \quad a = \bar{y} - b\bar{x} \quad (8.26)$$

Die Schätzungen  $a$  und  $b$  sind Zufallsvariablen. Da sie aus der gleichen Stichprobe geschätzt werden, sind sie meistens korreliert. Der Korrelationskoeffizient  $\rho_{a,b}$ , eine Zahl zwischen  $-1$  und  $1$ , ist ein Mass für den Zusammenhang der Schätzungen. Er ist exakt bekannt, da er nur von der unabhängigen Variablen abhängt. Für die Interpretation der Resultate ist er kaum von Bedeutung. Allenfalls wird er im Zusammenhang mit Fehlerfortpflanzung wichtig, wenn der Fehler einer Grösse berechnet werden soll, die eine Funktion der Parameter  $\alpha$  und  $\beta$  ist.

$$\rho_{a,b} = -\frac{\bar{x}}{\sqrt{\overline{x^2}}} \quad (8.27)$$

Standardabweichung der Messfehler von  $y$ :

$$s = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - a - bx_i)^2} \quad (8.28)$$

$$\text{Anzahl Freiheitsgrade: } \quad v = n - 2$$

Vertrauensbereich  $VB_b$  der Steigung  $b$ :

$$s_b = \frac{s}{\sqrt{SS_x}} \quad (8.29)$$

$$VB_b = b \pm t_v s_b \quad (8.30)$$

Vertrauensbereich  $VB_a$  des Achsenabschnittes  $a$ :

$$s_a = s \sqrt{\frac{\overline{x^2}}{SS_x}} \quad (8.31)$$

$$VB_a = a \pm t_v s_a \quad (8.32)$$

Schätzung für den Erwartungswert eines Messwertes:

$$\hat{y} = bx + a \quad (\text{Ausgleichsgerade}) \quad (8.33)$$

Toleranzbereich  $TB_y$  eines einzelnen Messwertes:

$$TB_y = \hat{y} \pm t_v s \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{SS_x}} \quad (8.34)$$

Vertrauensbereich  $VB_{\hat{y}}$  des Mittelwertes  $\hat{y}$ :

$$VB_{\hat{y}} = \hat{y} \pm t_v s \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{SS_x}} \quad (8.35)$$

Die Länge des Hebelarmes für den i-ten Messwert:

$$\frac{1}{\sqrt{1 - \frac{\bar{x}^2 - 2x_i\bar{x} + x_i^2}{SS_x}}} \quad (8.36)$$

In der analytischen Chemie will man häufig von einem Messwert  $y$  auf das zugehörige  $x$  schliessen. Beispiel: Man hat eine Kalibrationsgerade aus  $n$  Wertepaaren  $x_i, y_i$  berechnet. Jetzt misst man bei unbekanntem, aber festem  $x$  einige  $y$ -Werte. Wie gross ist der zugehörige  $x$ -Wert?

Schätzung für  $x$ :

$$\hat{x} = \left[ \frac{1}{b} \sum_{j=1}^m y_j \right] - \frac{a}{b} = \frac{1}{b} \left[ \frac{1}{m} \left( \sum_{j=1}^m y_j \right) - a \right] \quad (8.37)$$

$m$ : Anzahl  $y_j$ -Werte

Vertrauensbereich für  $\hat{x}$ :

$$VB_{\hat{x}} = \hat{x} \pm t_v s \sqrt{\frac{1}{b^2} \left[ \frac{1}{m} + \frac{1}{n} + \frac{(\hat{x} - \bar{x})^2}{SS_x} \right]} \quad (8.38)$$

Anzahl Freiheitsgrade  $v = n - 2$

Rechenbeispiel:

Die Daten der Eisenbestimmung (s. Kap. 7.2.) werden nun auch noch mit dem Modell der linearen Regression ausgewertet:

$$\begin{aligned} \text{Hilfsgrössen:} \quad \bar{c} &= 10,587 \text{ mg/l} & \bar{E} &= 0,83650 & \bar{c}^2 &= 136,19 \text{ (mg/l)}^2 \\ SS_c &= 144,70 \text{ (mg/l)}^2 & SS_{cE} &= 11,773 \text{ mg/l} & & \text{nach (8.24)} \end{aligned}$$

Schätzung der Parameter:

$$b = \frac{SS_{cE}}{SS_c} = 0,08136 \text{ l/mg} \quad \text{nach (8.25)}$$

$$a = \bar{E} - b\bar{c} = -0,02482 \quad \text{nach (8.26)}$$

$$s = \sqrt{\frac{1}{6-2} \cdot 0,02058} = 0,07172 \quad \text{nach (8.28)}$$

Anzahl Freiheitsgrade  $\nu = 4$

$$t_{4,95\%} = 2,776 \quad (\text{aus einer t-Tabelle})$$

$$s_b = \frac{0,07172}{\sqrt{144,70 \text{ (mg/l)}^2}} = 0,005962 \text{ l/mg} \quad \text{nach (8.29)}$$

$$VB_b = (0,08136 \pm 0,01655) \text{ l/mg} \quad \text{nach (8.30)}$$

$$s_a = 0,07172 \sqrt{\frac{136,19 \text{ (mg/l)}^2}{144,70 \text{ (mg/l)}^2}} = 0,06958 \quad \text{nach (8.31)}$$

$$VB_a = -0,02482 \pm 0,19316 \quad \text{nach (8.32)}$$

$$\hat{E} = 0,08136 \text{ l/mg} \cdot c - 0,02482 \quad \text{nach (8.33)}$$

$$TB_E = \hat{E} \pm 0,1991 \sqrt{\frac{7}{6} + \frac{(c - 10,587 \text{ mg/l})^2}{144,70 \text{ (mg/l)}^2}} \quad \text{nach (8.34)}$$

$$VB_{\hat{E}} = \hat{E} \pm 0,1991 \sqrt{\frac{1}{6} + \frac{(c - 10,587 \text{ mg/l})^2}{144,70 \text{ (mg/l)}^2}} \quad \text{nach (8.35)}$$

Die Hebelarme, Residuen und mit den Hebelarmen normierten Residuen:

Konzentration $c_i$ mg/l	Hebelarm	Residuum	Residuum normiert
3,27	1,4690	0,0898	0,1319
5,84	1,2148	-0,0743	-0,0903
9,71	1,0990	-0,0642	-0,0705
12,36	1,1100	-0,0078	-0,0086
15,08	1,2006	0,0529	0,0635
17,26	1,3794	0,0036	0,0049

Schätzung der Eisenkonzentration  $\hat{c}$  im Apfelsaft:

$$\hat{c} = \frac{1}{0,08136 \text{ l/mg}} \left( \frac{1}{2}(0,619+0,526) + 0,02482 \right) = 7,3418 \text{ mg/l} \quad \text{nach (8.37)}$$

$$VB_{\hat{c}} = 7,3418 \text{ mg/l} \pm 0,1991 \sqrt{\frac{\frac{1}{2} + \frac{1}{6} + 0,07277}{(0,08136 \text{ l/mg})^2}} \quad \text{nach (8.38)}$$

Schlussresultat:  $VB_{\hat{c}} = (7,34 \pm 2,11) \text{ mg/l} \quad (95 \%)$

Das Modell "Proportionaler Zusammenhang mit Normalverteilung" liefert für die gleichen Daten einen deutlich engeren  $VB_{\hat{c}}$ . Den Gewinn an Präzision verdankt man der zusätzlichen Annahme  $\alpha = 0$ .

Graphische Darstellung der Messresultate:

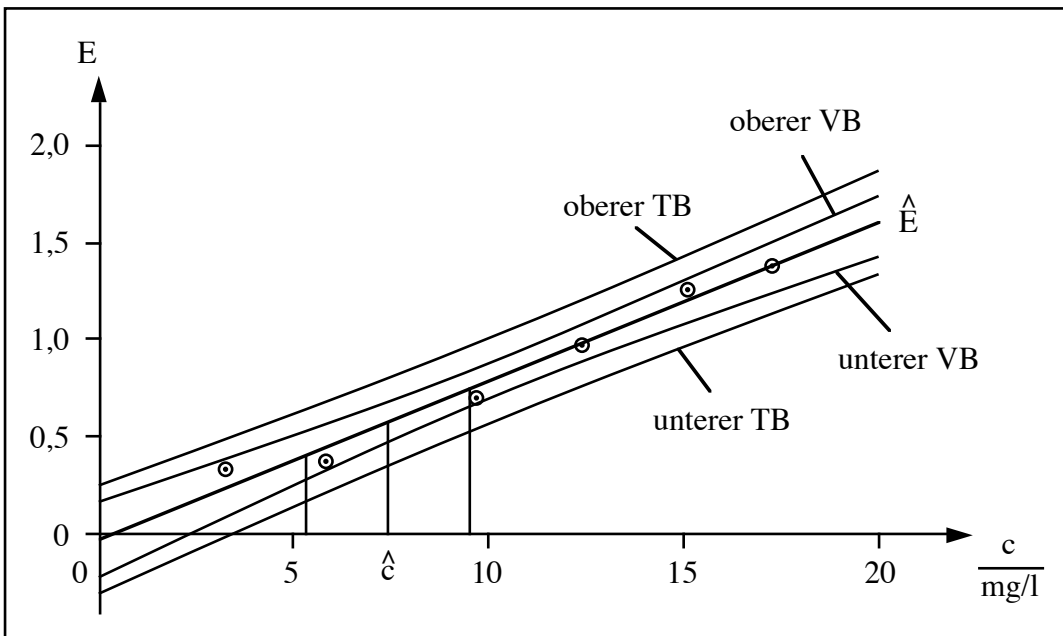


Fig. 8.8 Rechenbeispiel Linearer Zusammenhang, Normalverteilung

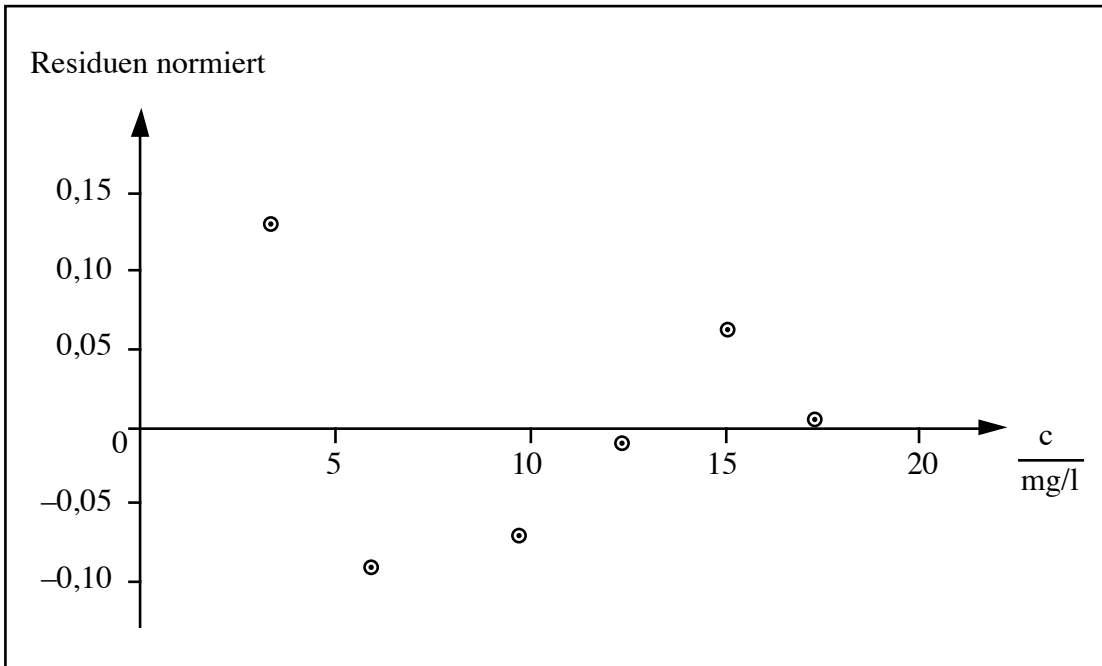


Fig. 8.9 Normierte Residuen versus Konzentration

#### 8.4. Standardaddition

Konzentrationsbestimmungen werden oft durch eine störende Matrix erschwert. Ein möglicher Ausweg ist das Herstellen von Kalibrationslösungen mit dem gleichen störenden Hintergrund. Einfacher ist es, dem Probenmaterial verschiedene bekannte Mengen der zu bestimmenden Substanz zuzusetzen und an den so behandelten Proben die Messungen durchzuführen. Diese Methode heisst Standardaddition (standard additions) oder Aufstocken.

Annahmen:

Es besteht ein proportionaler Zusammenhang zwischen der Konzentration  $c$  und dem gemessenen Signal  $S$ .

$$S = k c \quad (8.39)$$

Die Störungen durch die Matrix lassen sich durch ein multiplikatives Glied erfassen:

$$S = f_{\text{Stör}} k c \quad (8.40)$$

Die Matrix darf also nicht zu einem additiven Untergrund  $S_{\text{Stör}}$  führen. Für die Messgrösse  $S$  müssen zudem die Bedingungen für die lineare Regression erfüllt sein (s. Kapitel 8.3.).

Die unbekannte Konzentration  $c_0$  wird nach folgender Methode bestimmt:

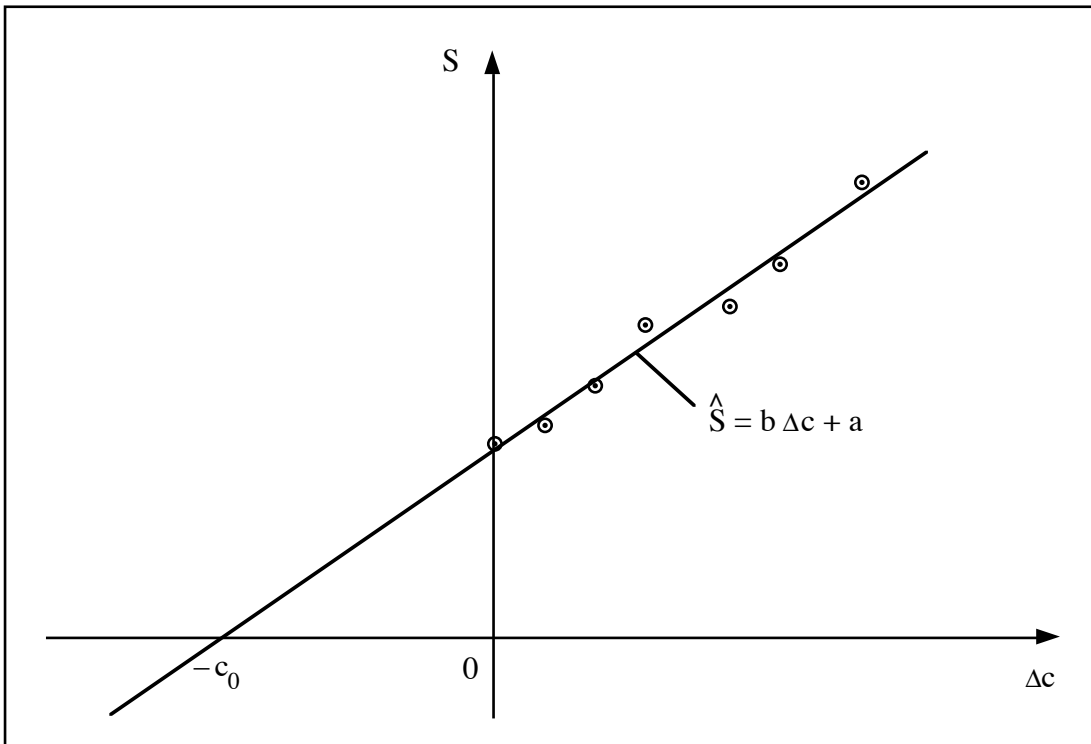


Fig. 8.10 Standardaddition

Man trägt die gemessenen Signale  $S_i$  gegen die Zusatzkonzentrationen  $\Delta c_i$  auf, passt eine Gerade  $\hat{S} = b \Delta c + a$  an und findet bei der Nullstelle die gesuchte Konzentration  $c_0 = \frac{a}{b}$ . Man berechnet  $a$  und  $b$  mit einer linearen Regression. Es gelten die Gleichungen (8.24) – (8.38). Man setzt  $\Delta c$  für  $x$  und  $S$  für  $y$ .

Für den Vertrauensbereich  $VB_{c_0}$  erhält man:

$$VB_{c_0} = \frac{a}{b} \pm t_v s \frac{1}{b} \sqrt{\frac{1}{n} + \frac{\bar{S}^2}{b^2 SS_{\Delta c}}} \quad (8.41)$$

$$v = n - 2$$

Nicht immer ist es bequem, die Zusatzkonzentrationen  $\Delta c_i$  für jeden Ansatz einzeln auszurechnen. In Spezialfällen geht es auch einfacher, z. B. wenn die Probe als Lösung vorliegt und die Aufstokkung durch Zugabe einer Kalibrationslösung geschieht.

Man bereitet eine Kalibrationslösung mit exakt bekannter Konzentration  $c_E$  zu, versetzt ein konstantes Volumen  $V$  Probelösung mit verschiedenen Volumina  $\Delta V_i$  Kalibrationslösung und misst von jedem Gemisch das Signal  $S_i$ . Nun passt man mit den Gleichungen (8.24) – (8.38) eine Gerade an. Dabei setzt man:

$$\begin{aligned} y_i &:= S_i (V + \Delta V_i) \\ x_i &:= \Delta V_i \end{aligned} \quad (8.42)$$

Diese Aufstockungsmethode ist etwas problematisch, da die einzelnen Ansätze nicht in gleicher Weise verdünnt werden und daher auch nicht den gleichen Hintergrund haben. Dieser Nachteil lässt sich mit einer anderen Methode beheben: Nach der Zugabe der Kalibrationslösung zur Probelösung bringt man das Volumen mit Lösungsmittel auf ein konstantes Endvolumen. In diesem Fall setzt man:

$$\begin{aligned} y_i &:= S_i \\ x_i &:= \Delta V_i \end{aligned} \quad (8.43)$$

In beiden Fällen erhält man für die gesuchte Konzentration  $c_0$ :

$$c_0 = \frac{a}{b} \frac{c_E}{V} \quad (8.44)$$

$$VB_{c_0} = \frac{a}{b} \frac{c_E}{V} \pm \frac{c_E}{V} t_v s \frac{1}{b} \sqrt{\frac{1}{n} + \frac{\bar{y}^2}{b^2 SS_x}} \quad (8.45)$$

$$v = n - 2$$

Rechenbeispiel:

Die Konzentration von Blei in Vollblut wird mit einer AAS-Graphitrohrmethode bestimmt. Das Blut wird mit  $\text{NH}_3 / \text{H}_2\text{O}_2$  nass verascht und dann direkt in die Graphitrohrküvette eingespritzt. Der Untergrund wird mit einer  $\text{D}_2$ -Lampe kompensiert. Die Aufstockung erfolgt mit einer wässrigen Lösung von  $\text{PbCl}_2$  mit einem Gehalt von  $1000 \mu\text{g/l Pb}^{2+}$ . Jeweils 1 ml Blut wird mit unterschiedlichen Mengen Standard versetzt. Um die Verdünnung konstant zu halten, werden die Ansätze mit Wasser auf ein totales Volumen von 1,3 ml gebracht. Jede aufbereitete Probe wird dreimal eingespritzt.

$\frac{\Delta V_i}{\mu\text{l}}$	$\frac{S_i}{\text{mm}}$
0	24,7
0	35,0
0	26,3
100	58,8
100	52,3
100	61,8
200	72,7
200	72,5
200	74,0
300	112,5
300	121,5
300	122,2

Man setzt:  $y_i := S_i$       $x_i := \Delta V_i$

$$\bar{x} = 150 \mu\text{l} \quad \bar{y} = 69,525 \text{ mm}$$

$$SS_x = 150000 \mu\text{l}^2 \quad SS_{xy} = 42845 \mu\text{l mm}$$

$$b = \frac{42845 \mu\text{l mm}}{150000 \mu\text{l}^2} = 0,28563 \frac{\text{mm}}{\mu\text{l}}$$



$$a = 69,525 \text{ mm} - 0,28563 \frac{\text{mm}}{\mu\text{l}} \cdot 150 \mu\text{l} = 26,68 \text{ mm}$$

$$s = \sqrt{\frac{1}{12-2} 664,96 \text{ mm}^2} = 8,1545 \text{ mm} \quad \text{nach (8.28)}$$

$$v = 10 \quad t_{10,95\%} = 2,228$$

$$c_0 = \frac{26,68 \text{ mm} \cdot 1000 \mu\text{g/l}}{0,28563 \text{ mm}/\mu\text{l} \cdot 1000 \mu\text{l}} = 93,41 \mu\text{g/l}$$

$$VB_{c_0} = c_0 \pm \frac{1000 \mu\text{g/l} \cdot 2,228 \cdot 8,154 \text{ mm}}{1000 \mu\text{l} \cdot 0,28563 \text{ mm}/\mu\text{l}} R = 63,607 \mu\text{g/l} \cdot R$$

$$\text{wobei } R = \sqrt{\frac{1}{12} + \frac{(69,525 \text{ mm})^2}{(0,28563 \text{ mm}/\mu\text{l})^2 \cdot 150'000 \mu\text{l}^2}} = 0,6916$$

$$VB_{c_0} = 93,41 \mu\text{g/l} \pm 43,99 \mu\text{g/l}$$

Schlussresultat:  $VB_{c_0} = (93 \pm 44) \mu\text{g/l} \quad (95 \%)$

## 9. Sammlung statistischer Tests

Die hier vorgestellten, häufig gebrauchten Tests setzen ausnahmslos die Normalverteilung der Messwerte voraus.

### 9.1. Vergleich zweier Varianzen, F-Test

Man hat aus zwei Grundgesamtheiten je eine Stichprobe genommen und möchte wissen, ob die Varianzen übereinstimmen, oder man möchte eine Varianz mit einem Sollwert vergleichen. Zu jeder Stichprobe gehört die Anzahl Freiheitsgrade  $\nu_i$ . Für Soll-Varianzen setzt man  $\nu_i = \infty$ .

$$\text{Nullhypothese } H_0 : \sigma_1^2 = \sigma_2^2$$

$$\text{Alternativhypothese } H_A : \sigma_1^2 \neq \sigma_2^2$$

Man wählt die Indizes 1 und 2 so, dass  $s_1^2 \geq s_2^2$ .

$$\text{Man berechnet } \hat{F} = \frac{s_1^2}{s_2^2} \quad (9.1)$$

und vergleicht mit dem Tabellenwert  $F_{\nu_1, \nu_2}$  für die gewünschte statistische Sicherheit. Übersteigt  $\hat{F}$  den Tabellenwert, muss mit einem Unterschied gerechnet werden.

Bemerkungen:

Der F-Test ist der mächtigste Varianztest, der in der analytischen Chemie sinnvoll ist. Trotzdem ist der Test nicht sehr mächtig. Er hat also Mühe, Unterschiede zu finden. Der progressive Entdecker wählt daher meist nur 90 % statistische Sicherheit. Auch der konservative Kritiker geht kaum über 95 %. Bei Abweichungen von der Normalverteilung reagiert der Test bezüglich der statistischen Sicherheit ziemlich robust. Wenn also die Varianzen wirklich gleich sind, stören die Abweichungen wenig. Die Macht leidet aber sehr stark. Unterschiede in den Varianzen kann der Test bereits bei kleinen Abweichungen von der Normalverteilung nur noch schlecht erkennen.

Rechenbeispiel:

Zu je 7 Proben "Tomate" und "Gurke" wurde eine konstante Menge KBr zugesetzt. Anschliessend wurde die Bromidkonzentration gemessen.

(J.A.Roughan, P.A.Roughan, J.P.G. Wilkins, Analyst **108**, 742 (1983) )

Tomate :	777	790	759	790	770	758	764	$\mu\text{g/g}$
Gurke:	782	773	778	765	789	797	782	$\mu\text{g/g}$

Sind die Schwankungen bei beiden Gemüsearten gleich? Die Indizes T und G stehen für Tomate und Gurke.

$$\text{Tomate: } \bar{c}_T = 772,6 \mu\text{g/g} \quad s_{c_T} = 13,6 \mu\text{g/g}$$

$$\text{Gurke: } \bar{c}_G = 780,9 \mu\text{g/g} \quad s_{c_G} = 10,4 \mu\text{g/g}$$

Zur Berechnung der Teststatistik setzt man die grössere Varianz in den Zähler:

$$\hat{F} = \frac{s_{c_T}^2}{s_{c_G}^2} = \frac{(13,6 \mu\text{g/g})^2}{(10,4 \mu\text{g/g})^2} = 1,70$$

Einer F-Tabelle entnimmt man für  $v_T = 6$  und  $v_G = 6$  :

$$F_{6,6} (80 \%) = 3,05$$

$$F_{6,6} (90 \%) = 4,28$$

$$F_{6,6} (95 \%) = 5,82$$

Es kann also kein Unterschied festgestellt werden.

## 9.2. Vergleich zweier Mittelwerte, t-Test

1. Fall: Man hat aus zwei Grundgesamtheiten je eine Stichprobe mit den Umfängen  $n_1$  und  $n_2$  genommen und möchte wissen, ob die Mittelwerte  $\mu_1$  und  $\mu_2$  der Grundgesamtheiten übereinstimmen. Gleichheit der Varianzen wird vorausgesetzt.

$$\text{Nullhypothese } H_0 : \quad \mu_1 = \mu_2$$

$$\text{Alternativhypothese } H_A : \quad \mu_1 \neq \mu_2$$

Man berechnet von den beiden Stichproben die Mittelwerte  $\bar{x}_1$  und  $\bar{x}_2$  sowie die Standardabweichungen  $s_1$  und  $s_2$ .

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{s_d} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \quad (9.2)$$

$$\text{wobei} \quad s_d = \sqrt{\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}}$$

$$\text{und} \quad v = n_1 + n_2 - 2$$

2. Fall: Man möchte einen Mittelwert ( $\bar{x}_1$ ) mit einem Sollwert ( $\mu_2$ ) vergleichen.

$$\text{Nullhypothese } H_0 : \quad \mu_1 = \mu_2$$

$$\text{Alternativhypothese } H_A : \quad \mu_1 \neq \mu_2$$

$$\hat{t} = \frac{|\bar{x}_1 - \mu_2|}{s_{\bar{x}_1}} = \frac{|\bar{x}_1 - \mu_2|}{s_{x_1}} \sqrt{n_1} \quad (9.3)$$

$$v = n_1 - 1$$

Man vergleicht  $\hat{t}$  mit dem Tabellenwert  $t_v$  für die gewünschte statistische Sicherheit. Übersteigt  $\hat{t}$  den Tabellenwert, so muss mit einem Unterschied gerechnet werden.

Bemerkungen:

Der Kritiker wählt als statistische Sicherheit etwa 99 %, der Entdecker etwa 95 %. Der t-Test ist recht robust gegen Abweichungen von der Varianzgleichheit und der Normalverteilung. Dies gilt besonders bei grossen Stichprobenumfängen.

Rechenbeispiel für den 1. Fall:

Das gleiche Beispiel wie beim F-Test wird verwendet. Sind die Mittelwerte verschieden?

$$\text{Tomate: } \bar{c}_T = 772,6 \mu\text{g/g} \quad s_{c_T} = 13,6 \mu\text{g/g}$$

$$\text{Gurke: } \bar{c}_G = 780,9 \mu\text{g/g} \quad s_{c_G} = 10,4 \mu\text{g/g}$$

sowie  $n_1 = n_2 = 7$

Einen Unterschied in den Varianzen konnte nicht nachweisen werden. Die Bedingung der Varianzgleichheit dürfte also erfüllt sein.

Mit Gleichung (6.3.):

$$s_d = \sqrt{\frac{(7-1)(13,6 \mu\text{g/g})^2 + (7-1)(10,4 \mu\text{g/g})^2}{7+7-2}} = 12,1 \mu\text{g/g}$$

$$v = 7 + 7 - 2 = 12$$

Mit Gleichung (6.4.):

$$\hat{t} = \frac{|772,6 \mu\text{g/g} - 780,9 \mu\text{g/g}|}{12,1 \mu\text{g/g}} \sqrt{\frac{7 \cdot 7}{7+7}} = 1,282$$

Der Vergleich mit dem Tabellenwert  $t_{12,95\%} = 2,179$  lässt keinen Unterschied erkennen.

### 9.3. Vergleich zusammengehörender Datenpaare

Es kommt häufig vor, dass man zwei verschiedene Messmethoden vergleichen möchte. Besonders wichtig ist die Frage nach systematischen Unterschieden in den Mittelwerten, also z. B. die Frage, ob Methode 1 grundsätzlich höhere Werte liefert als Methode 2. Dies lässt sich besonders effizient prüfen, wenn man beide Methoden auf das gleiche Probenmaterial anwendet und nur die Differenzen betrachtet. Man vergleicht dabei den Mittelwert der Differenzen mit dem hypothetischen Wert null, mit Hilfe des t-Tests.

Rechenbeispiel:

Eine neue AAS-Methode zur Bestimmung von Antimon in Luft wird mit der herkömmlichen kolorimetrischen Methode verglichen. Beide Methoden wurden auf sechs verschiedene Luftproben angewandt:

Probe	Gefundene Menge Antimon mg/m <sup>3</sup>	
	neue Methode	Standardmethode
1	22,2	25,0
2	19,2	19,5
3	15,7	16,6
4	20,4	21,3
5	19,6	20,7
6	15,7	16,8

(J.R. Catillo, J. Lanaja, Ma. C. Martinez, J. Aznarez, Analyst **107**, 1488 (1982) )

Unterscheiden sich die beiden Methoden?

Zuerst berechnet man für jede Probe die Differenz zwischen den beiden Messwerten:

$$d_i = x_{1i} - x_{2i}$$

Dabei beziehen sich die Indizes 1 und 2 auf die Methoden 1 und 2. Man erhält:

$$\begin{array}{lll} d_1 = -2,8 \text{ mg/m}^3 & d_2 = -0,3 \text{ mg/m}^3 & d_3 = -0,9 \text{ mg/m}^3 \\ d_4 = -0,9 \text{ mg/m}^3 & d_5 = -1,1 \text{ mg/m}^3 & d_6 = -1,1 \text{ mg/m}^3 \end{array}$$

Wenn es keinen Unterschied zwischen den beiden Methoden gibt, streuen die Differenzen  $d_i$  um den Wert null. Man testet also den Mittelwert der Differenzen gegen den hypothetischen Mittelwert null mit einem t-Test. In der Gleichung (9.3) setzt man  $\mu_2 = 0$  und erhält:

$$\hat{t} = \frac{|\bar{d}|}{s_{\bar{d}}} = \frac{|\bar{d}|}{s_d} \sqrt{n} \quad \text{und} \quad v = n - 1 \quad (9.4)$$

Im Beispiel:

$$\bar{d} = -1,183 \text{ mg/m}^3 \quad s_d = 0,845 \text{ mg/m}^3$$

$$\hat{t} = \frac{|-1,183 \text{ mg/m}^3|}{0,845 \text{ mg/m}^3} \sqrt{6} = 3,431$$

Aus der t-Tabelle entnimmt man für  $v = 5$ :  $t_5 (95 \%) = 2,571$   
 $t_5 (99 \%) = 4,032$

Da es gefährlich ist, systematische Unterschiede zwischen zwei Methoden nicht zu entdecken, nimmt man hier den Standpunkt des Entdeckers ein und testet auf dem 95%-Niveau. Der Unterschied ist dann also signifikant.