

Jede Menge Daten, was nun?
oder

Statistik für Chemiker
Ein „Kochbuch“

Alexander Schatten

21. März 1999

Inhaltsverzeichnis

1. Einleitung	11
2. Prolog: Information	17
2.1. Einleitung	17
2.2. Was ist Information?	17
2.3. Entropie	19
2.4. Statistik und die „Gewinnung“ von Information	21
3. * Statistische Tests	23
3.1. Einleitung	23
3.2. Hypothesen	23
3.3. Vertrauens- und Irrtumswahrscheinlichkeit	24
3.4. Fehler, Power eines Tests	24
4. Angabe eines Meßergebnisses mit Vertrauensbereich	27
4.1. Einleitung	27
4.2. Beispiele	29
4.3. Arten von Fehlern	29
4.4. Grundgesamtheit, Stichprobe	30
4.5. Verteilung von Daten — Normalverteilung	31
4.5.1. Einleitung	31
4.5.2. Histogramme	32
4.5.3. Stamm-Blatt-Diagramm	34
4.5.4. Normalverteilung	36
4.6. Mittelwert	38
4.7. Standardabweichung	39
4.8. Variationskoeffizient	42
4.9. Varianz	42
4.10. Vertrauensbereich	42
4.11. Nachweis- und Erfassungsgrenze	44
4.12. Abweichung der Meßwerte von der Normalverteilung	45
4.13. Robuste Größen: Median, Quartile	46
4.13.1. Median	46
4.13.2. p-te Perzentile, Quartile, Interquartiler Abstand	47
4.14. Signifikante Stellen	47
4.15. Korrekte Angabe der Ergebnisse der Beispiele	49

4.15.1. Beispiel 1: „Bromid“	49
4.15.2. Beispiel 2: „Blei in Wasser“	50
4.15.3. Beispiel 3: „GC-ECD“	50
4.15.4. Beispiel 4: „Nitrat im Trinkwasser“	50
5. Ausreißertests	53
5.1. Einleitung	53
5.2. 4- σ -Bereich	53
5.3. Test nach Dean und Dixon	54
5.4. Beispiel 3: „GC-ECD“	55
6. Vergleich von Messungen	57
6.1. Einleitung	57
6.2. Vergleich von Mittelwerten	58
6.3. Vergleich eines Mittelwertes mit einem erwarteten Wert	59
6.4. Vergleich von Standardabweichungen	60
7. Kausaler Zusammenhang zwischen Variablen	63
7.1. Einleitung	63
7.2. Beispiele	63
7.3. Scatter-Plot	64
7.4. Korrelation	65
7.5. Ergebnis	67
7.6. Korrelation = Kausalität ?	68
7.6.1. Einleitung	68
7.6.2. Korrelation zwischen Zeitreihen	69
7.6.3. Formale Korrelation	69
7.6.4. Korrelation durch Inhomogenität	69
7.6.5. Korrelation aufgrund einer gemeinsamen Basis	70
7.6.6. Kausalität?	71
7.7. Zusammenfassung	71
8. Bestimmen einer Eichgerade	73
8.1. Einleitung	73
8.2. Interpolation, Extrapolation	73
8.3. Beispiel	74
8.4. Scatter Plot	75
8.5. Regressionsrechnung	76
8.6. Analyse der Residuen	79
8.7. $A = kB + d$ und $B = kA + d$	82
8.8. Ergebnis/Zusammenfassung	84
A. Tabellenanhang	87

Abbildungsverzeichnis

4.1. Struktogramm: Angabe von Meßergebnissen	28
4.2. Systematischer und Zufälliger Fehler	30
4.3. Histogramm	33
4.4. „Weg“ von der diskreten zur kontinuierlichen Verteilung.	34
4.5. Normalverteilung	37
7.1. Scatterplot: Schadstoff/Krebs	65
7.2. Zeitreihe: Vergleich Scatter/Lineplot	66
7.3. Inhomogenitätskorrelation, Beispiel	70
8.1. Lineare Approximation der Sinusfunktion.	74
8.2. Scatterplot Eichgerade	76
8.3. Residuen der linearen Regression	76
8.4. Analyse der Residuen: linear, nicht-linear	80
8.5. Analyse der Residuen: Inhomogenitäten, Ausreißer	81
8.6. Regression: Auswirkung von Ausreißern	82
8.7. Beide Regressionsgeraden im Scatterplot (Visualisierung der Korrelation)	83

Tabellenverzeichnis

3.1. Teststatistik: Fehler 1. und 2.Art	25
4.1. Signifikante Stellen: Multiplikation	48
7.1. Korrelation: Schadstoff-Krebs	64
8.1. Konzentration und Peakflächen der Eichproben	75
A.1. Student t Verteilung	88
A.2. Q-Werte für Dean und Dixon Ausreißertest.	89
A.3. F-Test: 0,90 I	90
A.4. F-Test: 0,90 II	91
A.5. F-Test: 0,90 III	92
A.6. F-Test: 0,95 I	93
A.7. F-Test: 0,95 II	94
A.8. F-Test: 0,95 III	95
A.9. F-Test: 0,99 I	96
A.10.F-Test: 0,99 II	97
A.11.F-Test: 0,99 III	98

Symbolverzeichnis

Variablen

\bar{x}	Mittelwert der Stichprobe
\bar{x}_B	mittlerer Blindwert
μ	Mittelwert der Grundgesamtheit (sprich: mü)
σ	Standardabweichung der Grundgesamtheit (sprich: Sigma)
ξ	Erwarteter Mittelwert (sprich: xi)
d	Offset der Ausgleichsgerade.
e_i	i -tes Residuum.
F	Prüfwert für F-Test, tabelliert.
f	manchmal: Anzahl der Freiheitsgrade.
FG	Anzahl der Freiheitsgrade.
g	Gramm
I	Information
k	Steigung der Ausgleichsgerade.
N	in Summen: Anzahl der Meßwerte.
n	in Summen: laufende Variable, „Zähler“; ansonsten Anzahl der Meßwerte
$P(B)$	Wahrscheinlichkeit, daß Bedingung B eintritt.
p_i	Wahrscheinlichkeit des i -ten Zustandes eines Systems
Q	Prüfparameter für Dean und Dixon Ausreißertest.
r	Korrelationskoeffizient.
r^2	Bestimmtheitsmaß
S	Entropie
s	Standardabweichung der Stichprobe
s_B	Standardabweichung des Blindwertes
t	Tabellierter Wert der Student Verteilung; für Vertrauensbereich und t -Test.
V	Variationskoeffizient
v	Varianz
x	ein Meßwert
x_i	der i -te Meßwert einer Reihe von Meßwerten.
x_{links}	linke Grenze eines Intervalls
x_{rechts}	rechte Grenze eines Intervalls
z	Tabellierter Parameter für Vertrauensbereich.
A,B	Daten für lineare Regression; Variable A, B.
Abszisse	x - Achse eines Scatterplots; horizontale Achse;
ECD	Electron Capture Detector, Elektronen-Einfang-Detektor
EG	analytische Erfassungsgrenze
GC	Gaschromatograph
HPLC	Hochdruck-Flüssigkeits-Chromatographie; High Pressure Liquid Chromatography, manchmal auch High <i>Performance</i> ...
K	Kelvin

LS	Least Sum of Squared Residuals; manchmal auch nur Least Squares genannt.
m	Index. <i>m</i> -ter Wert einer Menge.
mg	Milligram, 10^{-3} g
n	Index. <i>n</i> -ter Wert einer Menge.
NG	analytische Nachweisgrenze
Ordinate	y - Achse eines Scatterplots; vertikale Achse
ppb	Parts per Billion, also ein Teilchen auf eine Milliarde Teilchen, $1:10^9$
ppm	Parts per Million, also ein Teilchen auf eine Million Teilchen, $1:10^6$

1. Einleitung

Klarheit ist ein intellektueller Wert an sich; Genauigkeit und Präzision aber sind es nicht. Absolute Präzision ist unerreichbar; und es ist zwecklos, genauer sein zu wollen, als es unsere Problemsituation verlangt.
Karl Popper [19]

Statistik ist als Grundlage für die naturwissenschaftliche Ausbildung nicht wegzudenken, trotzdem ist es oft ein ungeliebtes Fach, da der direkte Zusammenhang mit dem eigenen Fach nicht immer leicht erkenntlich ist, und der Stoff ziemlich Mathematik-lastig zu sein scheint. Dieses Skriptum soll einen praxisbezogenen Zugang für (analytische) Chemiker und verwandte Fächer liefern. Großer Wert wurde auf eine verständliche an Beispiel orientierten Einführung in die Anwendung ausgewählter Verfahren auf in der Praxis wichtige Anwendungen.

Es handelt sich nicht um eine umfassende Einführung in die Statistik, auch wird auf Herleitungen und umfangreiche mathematische Definitionen weitgehend verzichtet. Mit Hilfe dieses Manuskriptes soll der Praktiker in möglichst kurzer Zeit in der Lage sein, die hier beschriebenen Verfahren möglichst korrekt anzuwenden. Das vorliegende Manuskript ist noch nicht vollständig, und wird in unregelmäßigen Abständen erweitert werden. Falls Sie (per email) bei signifikanten Änderungen benachrichtigt werden möchten, teilen Sie mir dies bitte mit.

Sollten sich noch Fehler, oder die eine oder andere Ungereimtheit in dieser ersten Version finden, bitte ich dies zu entschuldigen, und mir diese Fehler oder Unklarheiten mitzuteilen. Nur durch einen aktiven Dialog zwischen Lesern (Studenten) und dem Autor ist eine gute Weiterentwicklung möglich. Selbstverständlich erhalten alle, die mir Fehler mitteilen bzw. mir Skripten mit Anmerkungen retournieren, eine aktualisierte Version zurück, sobald die Fehler bereinigt sind.

Wie Sie mit mir in Kontakt treten können, erfahren Sie in der Vorlesung bzw. am Ende dieses Skriptums. Weiters würde es mich freuen, wenn mir Leser Daten aus ihrer Meßpraxis zur Verfügung stellen, um diese gegebenenfalls in eine Beispielsammlung einbauen zu können. **Jeweils aktuelle Versionen finden Sie auf meiner Homepage (siehe Seite 15)**

Rückmeldungen

Zum didaktischen Aufbau: Dieses Skriptum ist „problemorientiert“ aufgebaut, d.h. es wird versucht anhand praktischer Problemfälle Lösungen zu entwickeln und die dafür notwendigen statistischen Methoden zu erklären. In manchen Fällen kommt es vor, daß zum Verständnis verschiedener Kapitel weitere Voraussetzungen notwendig sind, die sich nicht sinnvoll in ein Kapitel integrieren lassen. Diese Kapitel sind mit einem Stern * gekennzeichnet und **sollten nicht überblättert werden**.

An dieser Stelle sollte erwähnt werden, daß die gängigen Tabellenkalkulationen

Software

wie *Quattro Pro*, *Excel*, *1-2-3* und selbstverständlich alle mir bekannten Statistikpakete wie *SPSS*, *S+*, *Statistica*, ... und viele spezialisierte Programme zur Visualisierung von Daten wie *Origin* ebenso wie einige der modernen wissenschaftlichen Taschenrechner die meisten der Methoden, die hier vorgestellt werden, abdecken.

Gnuplot Weiters gibt es auch verschiedene Shareware und Freeware Produkte wie beispielsweise *Gnuplot* als Visualisierungs-Tool. Mit *Gnuplot* ist es möglich alle gängigen zwei- und dreidimensionalen Diagramme zu zeichnen. Leider unterstützt *Gnuplot* keine statistischen Plots wie Histogramme oder Box-Plots. Ein weitaus mächtigeres Paket ist *R*. Dabei handelt es sich um eine Freeware Implementierung der Statistiksprache *S*, die auch kommerziell unter *SPlus* (teuer) erworben werden kann. *SPlus* ist allerdings in vielen Funktionen deutlich umfangreicher als *S* oder *R*. Allerdings muß man sowohl zu *Gnuplot* als auch zu *R* oder *S* sagen, daß es sich eher um Skriptsprachen als um eine visuell bedienbare Software a la *Excel* handelt. Der Einstieg ist daher unter Umständen recht aufwendig, umso flexibler empfinden viele dann die weitere Anwendung. Daher würde ich bei gelegentlicher Anwendung statistischer Methoden, bzw. wenn die Probleme die in diesem Skriptum erwähnte Schwierigkeit nicht übersteigen eher von der Verwendung derartiger Programme abraten. Vor allem *S*, bzw. *R* oder *SPlus* unterscheiden sich sehr stark vom gängigen *Windows* bzw. *Apple Standard* und der aufwendige Einstieg lohnt eigentlich nur bei oftmaliger Verwendung bzw. komplexen Problemen. Vor allem die Verwaltung der Daten(tabellen) und Graphiken sind äußerst gewöhnungsbedürftig. Erst ein tieferer Einstieg in die Programmiersprache *S* offenbart hier die Details.

SPSS Sollten gängige Tabellenkalkulationen bzw. wissenschaftliche Taschenrechner für Analysen nicht mehr ausreichen, so würde ich persönlich am ehesten zu *SPSS* raten. Zwar kann auch dieses Paket seine *Mainframe/Workstation* Herkunft nicht verleugnen, die neuesten *Windows* Versionen verhalten sich jedoch recht *Windows* konform und stellen auch den Einsteiger nicht vor unlösbare Probleme. Ausserdem existieren an der *TU* Campuslizenzen, die für Institute die Kosten in Grenzen halten, was bei den ziemlich hohen Lizenzkosten für die genannten Pakete ein nicht unerhebliches Argument ist.

Statistica Für *Statistica* spricht unter anderem die Tatsache, daß eine in der Funktionalität etwas reduzierten Version als recht günstige Studentenlizenz zu erwerben ist. Dieses Programm ist vermutlich am besten in die *Windows* Umgebung integriert, verwirrt aber meiner Meinung nach manchmal durch ziemlich überladene Dialoge. Die Graphikfunktionalität von *Statistica* ist sicherlich eine Stärke dieser Software.

Grundsätzlich ist es (schon aus Kostengründen) empfehlenswert vorhandene Software zum Berechnen statistischer Parameter wie Mittelwert, Standardabweichung oder auch Median, Korrelationskoeffizient und Ausgleichsgerade zu verwenden. Dennoch ist es notwendig vor einer „blinden“ Anwendung irgend einer Methode, die eines der Softwarepakete anbietet, das entsprechende Kapitel dieses Skriptums durchzuarbeiten. Falls sich eine Methode hier noch nicht findet, kann möglicherweise eine der im Anhang empfohlenen Literaturstellen weiterhelfen.

Leider verleiten manche Softwareprodukte gerade dazu Methoden ohne entsprechende Kenntnisse zu verwenden, sei es durch komplexe Dialogboxen, die eine große Anzahl von Parametern zur Verfügung stellen, oder umfangreiche Ausgaben, die dem Einsteiger nicht klar sind. Oftmals sind dann zu allem Überfluß die Hilfetexte wenig

aussagekräftig. Leider ist in solchen Fällen vielfach zu beobachten, daß Anwender dann einfach die Grundeinstellungen ohne genauere Kenntnis der Bedeutung verwenden.

Manchmal sind aber gerade diese *Grundeinstellungen* das *Grundfalsche*, siehe z.B. die Bemerkungen zu Scatter- und Lineplot in Abschnitt 7.3 auf Seite 64 und Abb. 7.2. Gerade die ubiquitäre Verfügbarkeit sogar komplexer statistischer Methoden verführt zu unbedachter Vorgangsweise. Wenige würden sich wohl die Mühe machen „von Hand“ eine Ausgleichsgerade oder einen Korrelationskoeffizienten zu berechnen, wenn sie sich nicht einigermaßen im klaren darüber wären, was sie damit erreichen wollen. Da der Aufwand aber mittlerweile auf das Klicken eines Buttons reduziert ist, findet man derartiges heute recht häufig. Darum ist es — wie schon erwähnt — für den Einsteiger auch nur bedingt zu empfehlen die „großen“ Statistikpakete zu verwenden, da sich diese vorwiegend an Statistiker bzw. komplexere Problemstellungen richten.

In diesem Zusammenhang soll nicht unerwähnt bleiben, daß es mittlerweile Geräte gibt, die eher als *Palmtop Computer* als als Taschenrechner zu bezeichnen sind. Einige dieser Geräte beherrschen nicht nur eine Vielzahl mathematischer und statistischer Verfahren, manche verfügen auch über graphische Fähigkeiten zum Plotten von Funktionen, Scatterplots, Residual Plots oder auch Histogrammen und Boxplots. Als Beispiel sei hier der TI-83 der Firma Texas Instruments genannt. Vergleichbare Funktionalität bietet auch bspw. Sharp EL-9600 oder Casio CFX-9850Plus. Für manche Geräte gibt es auch portierte Versionen gängiger Mathematikpakete wie Derive, als Beispiel sei hier HP-95LX und Nachfolger genannt.

Taschenrechner

Für fast alle in der chemisch/analytischen Praxis auftretenden Probleme, ist die Funktionalität moderner Tabellenkalkulationen und Taschenrechner weit ausreichend und weniger verwirrend.

chemische Praxis

Die Anwendung statistischer Methoden die man nicht hinreichend versteht, bzw. über deren Einschränkungen und Randbedingung man nicht genau informiert ist, kann zu groben Verfälschungen des Ergebnisses führen und ist daher unbedingt zu vermeiden!

Aus diesem Grund ist auch ein Hauptaugenmerk in diesem Skriptum auf die *korrekte* Anwendung der vorgestellten Methoden gelegt worden, weniger auf eine vollständige Beschreibung.

Weiters finden sich im Anhang statistische Tabellen und weiterführende Referenzen. Im Buch *Statistik* aus dem Springer Verlag [9] findet man eine Einführung in die wichtigsten univariaten Methoden¹ und auch einen Einstieg in multivariater² Statistik. Eine sehr gute englischsprachige Einführung findet sich in *Statistical Methods*, Academic Press [10]. Mehr ins Detail geht das umfangreichere Lehrbuch von Lothar Sachs: *Angewandte Statistik*, Springer Verlag [21] bzw. von Joachim Hartung: *Statistik* aus dem Oldenbourg Verlag [12]. Diese sind allerdings ziemlich umfangreich und

Weiterführende Literatur

¹Univariate Methoden sind solche, die Zusammenhänge zwischen maximal zwei Variablen betreffen, bzw. alle grundlegenden deskriptive Verfahren wie Mittelwert, Standardabweichung, Ausreißer, ...

²Multivariate Statistik beschäftigt sich mit dem Aufdecken und Beschreiben von Zusammenhängen zwischen *mehreren* Variablen.

daher eher an den fortgeschrittenen Leser gerichtet. Eine weitere (kostenlose) Quelle findet sich als Vorlesungsskriptum der Vorlesung *Einführung in die Wahrscheinlichkeitsrechnung und Statistik* von Prof. Dutter an der TU-Wien [6, 7].

Dieses Skriptum ist parallel zur Arbeit an einem Multimedia Teachware Produkt zum Thema Datenanalyse/Statistik/Chemometrie entstanden. Für weitere Informationen siehe auch <http://www.vias.com>.

Allen Lesern empfehle ich *So lügt man mit Statistik* aus dem Campus Verlag [16]. Nach Lektüre dieses recht günstigen Taschenbuches wird man einige Aussagen, denen man im täglichen Leben z.B. in den Medien begegnet und die sich vordergründig auf Statistik oder exakte Zahlen stützen, mit anderen Augen gegenüberstehen. In eine recht ähnliche „Kerbe“ schlägt das Buch von Guardian Kettler *Zwei Nullen sind keine Acht — Falsche Zahlen in der Tagespresse* [13]. Kettler beschreibt hier in erster Linie den oftmals sorglosen Umgang mit Zahlenmaterial und Quellen in durchaus renommierten Medien und zeigt ähnlich wie auch Krämer die psychologische Wirkung **scheinbar** präziser Zahlenangaben auf.

Danksagungen

Vielen Dank an alle Kollegen und Freunde, die als „Testleser“ oder in anderer Form zur Verbesserung dieses Skriptums beigetragen haben (in alphabetischer Reihenfolge): Wolfgang Auer, David Bolius, Dr. Christina Gröllert, Petra Gruber, Hannes Kirschbaum, Prof. Lohninger und Marian Schedenig.

Weiters bin ich Prof. Paditz (HTW Dresden) zu Dank verpflichtet. An diesem Beispiel zeigt sich, welche interessanten Kontakte alleine durch das Medium Internet möglich sind. Prof. Paditz hat dieses Skriptum wohl eher zufällig auf meiner Homepage entdeckt und durch mehrere emails deutlich zur Verbesserung des Manuskriptes beigetragen. V.a. einige unklare bzw. falsche Formulierungen und Abbildungen wurden auf seine Anregungen hin korrigiert bzw. ersetzt. Weiters möchte ich mich noch einer für den Chemiker interessanten Literaturempfehlungen seinerseits anschließen: *Chemometrie* von Matthias Otto [18]. Dieses Lehrbuch gibt einen umfassenden Überblick über moderne chemometrische Methodik, leider sind nicht alle Verfahren für den Einsteiger optimal aufbereitet.

DI ALEXANDER SCHATTEN

Gallitzinstr. 7-13/7/7

1160 Wien

Tel.: 914 29 84

email: AlexanderSchatten@gmx.atURL: <http://www.bigfoot.com/~AlexanderSchatten>

2. Prolog: Information

*Nachher, vor allen anderen Sachen,
müßt Ihr Euch an die Metaphysik machen,
Da steht, daß Ihr tiefsinnig faßt,
was in des Menschen Hirn nicht paßt.*

Mephistopheles, Faust I

2.1. Einleitung

Dieser Abschnitt ist für das Verständnis der weiteren Kapitel nicht unmittelbar erforderlich. Alle, die es sehr eilig haben zum praktischen Kern vorzustößen, mögen also diesen kurzen Prolog überspringen. Alle anderen könnten einen kleinen Denkanstoß und sollten einen kleinen Einblick in die Ideen der Informationstheorie bekommen. Insofern, als *Information* und die Veränderung derselben die Basis jeder statistischen Analyse ist, ist dieser Teil des Skriptums andererseits doch wieder recht zentral. Jedoch muß einem unbedarften Lenker heutzutage auch nicht klar sein, wie ein Motor funktioniert, um erfolgreich ein Auto bedienen zu können, der Leser entscheide also selbst.

Es sollte noch angemerkt werden, daß der Inhalt dieses Kapitels im Gegensatz zu den anderen Kapiteln dieses Skriptums, vermutlich einigermaßen subjektiv ist, und sich durch meine persönlichen Anschauung geprägt darstellt.

2.2. Was ist Information?

Das problematische des Begriffes *Information* ist, daß er mindestens in zwei verschiedenen Bedeutungen Verwendung findet. In technischen, statistischen oder sogenannten *informationstheoretischen* Publikation ist Information üblicherweise als terminus technicus im Zusammenhang mit *Entropie* verwendet. Daneben findet das Wort Information auch noch in landläufiger Bedeutung Verwendung. In manchen Fällen werden diese beiden Bedeutungen zu allem Überfluß auch noch vermengt.

In eine vielbeachteten Buch [22] führten Shannon und Weaver die Begriffe *Informationstheorie*, *Information*, *Kommunikation*, *Entropie* als technische/mathematische Begriffe zur Beschreibung nachrichtenübermittelnder oder informatischer Probleme ein. Zu diesem Zweck wurden der an sich aus der Thermody-

namik stammenden Begriffe *Entropie* auch für die Anwendung in der Informationstheorie redefiniert.

Leider wurde und wird Information als terminus technicus auch relativ unkritisch auf alle möglichen und unmöglichen Probleme angewandt und die Verwirrung ist letztlich perfekt [8, 14]. Wollen wir also zunächst mal die technische — in weiterer Folge auch *informationstheoretische* Bedeutung von Information näher ansehen.

Definition Betrachten wir zunächst die mathematische Definition des Informationsbegriffes:

$$I_i = - {}_2 \log(p_i) \quad (2.1)$$

wobei

$$\sum_i p_i = 1 \quad (2.2)$$

Mit anderen Worten: Ein System besteht aus i Zuständen, von denen jeder mit einer gewissen Wahrscheinlichkeit auftritt. Die Wahrscheinlichkeit des i -ten Zustands wird als p_i bezeichnet. Die Summe der Wahrscheinlichkeiten muß natürlich gleich 1 sein, da ja irgendein Zustand immer auftreten muß, ansonsten hat man Zustände vergessen. Die Information des i -ten Zustandes wird nun als der Logarithmus zur Basis 2 der Wahrscheinlichkeit dieses Zustandes definiert.

Was hat das für eine Konsequenz? Der Informationsgehalt eines Zustandes ist nach dieser Definition umso *höher*, je *niedriger* seine Wahrscheinlichkeit ist. Dies erscheint auch intuitiv durchaus logisch zu sein. Betrachten wir einen Extremfall: Wird die Wahrscheinlichkeit eines Zustandes gleich 1, besteht das System also nur aus diesem einen Zustand, so ist der Informationsgehalt dieses Zustandes ziemlich gering. In einer Welt, in der alle Autos grün wären, wäre die Farbe als Entscheidungskriterium recht bedeutungslos.

Ist die Auftrittswahrscheinlichkeit eines Zustandes jedoch recht gering, so steigt der Informationsgehalt dieses Zustandes stark an. Um beim Auto Beispiel zu bleiben: Angenommen alle Autos wären grün, nur ein einziges Auto sei rot, scheint es einleuchtend zu sein, daß der Zustand: „Auto ist grün“ so gut wie keine Information beinhaltet, während der Zustand „Auto ist rot“ von ausgesprochen hoher Aussagekraft, also Information ist. Wir haben dann nämlich dieses eine Auto eindeutig identifiziert.

„reale“ Bedeutung?

Man darf bei dieser Definition jedoch niemals vergessen — und das ist die eigentliche Problematik — daß diese Definition der Information eine rein mathematische ist, und keinerlei Rückschluß auf die tatsächliche *Bedeutung* eines Ereignisses zuläßt. Diese Definition gibt sozusagen die theoretisch maximal erreichbare Information an. Versuchen wir die Problematik anhand eines anderen Beispiels zu erläutern:

Angenommen man untersucht verschiedene Kunststoffe. Die Proben, die zur Verfügung stehen, unterscheiden sich alle in der Farbe, wobei der Einfachheit halber angenommen wird, daß alle Farben gleich häufig vorkommen. Das zweite Kriterium dieser Substanzen sei die Zugfestigkeit. Hier stellt man im wesentlichen aber nur zwei Gruppen fest. Informationstheoretisch betrachtet würde diese Annahme bedeuten, daß die einzelne Farbe über einen hohen Informationsgehalt verfügt, die Zugfestigkeit aber nicht. Aus den obigen Erklärungen ist dies auch verständlich, denn eine

dieser Proben läßt sich anhand der Farbe eindeutig erkennen, anhand der Zugfestigkeit aber nicht.

Bleiben wir bei unserem Beispiel: Angenommen wir benötigen diesen Kunststoff als Rohstoff für irgendein neues technisches Produkt, und die einzige Eigenschaft an der wir *in diesem Zusammenhang* interessiert sind, ist die Zugfestigkeit — so wird sich plötzlich die Bedeutung der Information relativiert. Man sollte sich also davor in acht nehmen, eine aus der Formel errechneten Information mit realer Bedeutung gleichzusetzen!

2.3. Entropie

Wollen wir kurz wiederholen, was unter dem Begriff der *Entropie* (aus thermodynamischer Sicht) zu verstehen ist: Beobachtet man physikalische Systeme wie zum Beispiel Gase, so stellt man fest, daß immer wieder *gerichtete spontane* Vorgänge zu beobachten sind. Im Falle des Gases könnte man die Expansion betrachten. Jedes Gas *expandiert* in den größten zur Verfügung stehenden Raum. Dies erfolgt spontan. Die Umkehrung des Vorganges hingegen, also die *Kompression*, findet niemals spontan statt [1, 20]. Das Verbrennen von Kohle zu Kohlendioxid und Kohlenmonoxid läuft nur in diese Richtung ab. Das spontane Entstehen von Diamant oder Kohle aus Kohlendioxid hingegen wurde noch nie beobachtet.

Man kann einwenden, daß es natürlich bekannte chemische und physikalische Verfahren gibt, die diese Vorgänge umkehren, nur erfolgt dies eben niemals spontan, sondern nur unter Anwendung von äußerer Kraft.

Diese Beobachtungen sind interessant, da sie offensichtlich nicht von der *Gesamtenergie* des Systemes bestimmt ist, da diese konstant bleibt. Es ist also ein anderer Faktor zu suchen, der als Erklärung für diese Phänomene verstanden werden kann. Dieser „Faktor“ wurde in der Entropie gefunden, die als ein Maß für Ordnung bzw. Unordnung angesehen wird. Während der Ausgangszustand der oben erwähnten Systeme von hoher Ordnung ist, so sind die Endprodukte der Reaktionen von niedriger Ordnung. Systeme scheinen sich also spontan von höherer zu niedrigerer Ordnung zu bewegen.

Ersetzt man das Wort „Entropie“ durch ein der Alltagssprache geläufigeres, nämlich die „Wahrscheinlichkeit“, so wird der Zusammenhang verständlicher: Die *Wahrscheinlichkeit*, daß die Anordnung der Kohlenstoffatome im Diamant entsteht, ist zweifellos wesentlich geringer, als die *Wahrscheinlichkeit*, daß Kohlenstoff in Form eines Gases auftritt, wo wesentlich mehr mögliche Zustände als im Diamanten vorliegen.

Ein anderes Beispiel: Stellen wir uns einen Behälter mit weißem und schwarzem Sand vor. Zunächst ist weißer und schwarzer Sand sauber voneinander getrennt. Schütteln wir nun den Behälter so wird sich der Sand vermischen, jedoch werden wir niemals bemerken, daß sich der Sand spontan entmischt, und wieder den ursprünglichen Zustand annimmt.

Betrachten wir dieses System nun aus dem Blickwinkel der Wahrscheinlichkeit, so können wir leicht feststellen, daß es wesentlich mehr ungeordnete Zustände als geordnete Zustände (getrennter Sand) gibt. Folglich ist die Wahrscheinlichkeit (durchmischter Sand), daß sich der Sand spontan entmischt auch wesentlich geringer als

Entropie &
Thermodynamik

Entropie als
Wahrscheinlichkeit

daß er sich durchmischt. Setzen wir nun wieder das Wort Entropie an die Stelle des Wortes Wahrscheinlichkeit so sehen wir, daß sich die Systeme spontan von niedriger Entropie zu hoher Entropie bewegen. Man könnte dies auch in einem knappen Satz zusammenfassen: „**Unordnung entsteht, Ordnung muß gemacht werden**“.

Entropie &
Informationstheorie

Wie findet nun der Begriff der Entropie in der Informationstheorie Anwendung? Diese definiert die Entropie als

$$S = \sum_i -p_i \cdot \log(p_i) \quad (2.3)$$

wobei auch hier gilt

$$\sum_i p_i = 1 \quad (2.4)$$

Was sagt nun dieser informationstheoretische Begriff der Entropie aus? Betrachten wir als einfaches Beispiel ein System, das aus nur zwei Zuständen besteht. Die Wahrscheinlichkeit, daß Zustand 1 auftritt sei p_1 , daß Zustand 2 auftritt p_2 , die Summe muß natürlich wieder 1 ergeben.

Nehmen wir nun an, beide Zustände sind gleich wahrscheinlich. In diesem Fall wäre die Entropie gleich 1 und ein Maximum. Für alle anderen Fälle ist die Entropie des Systems niedriger. Was können wir daraus ablesen?

Bedeutung

Im Fall, daß ein Zustand viel häufiger auftritt als der andere, kann man dieses System als geordneter verstehen als für den „Extremfall“, daß alle möglichen Zustände gleichverteilt sind. Versuchen wir dies an einem praktischen Beispiel zu illustrieren: Verwahren wir unsere Autoschlüssel praktisch immer nur an zwei Orten, z.B. in der Tasche der Jacke und am Schlüsselbrett, alle anderen immerhin möglichen Orte in der Wohnung kommen praktisch nie vor, so ist dieses System wohl eher geordnet als ein System, indem die Wahrscheinlichkeit den Schlüssel zu finden in der ganzen Wohnung gleich groß ist, d.h. alle Zustände praktisch gleich häufig sind.

Ein weiteres Beispiel aus [8]: Die deutsche Sprache besteht aus 30 Symbolen (Buchstaben). Angenommen, alle 30 Buchstaben würden in unserer Sprache gleich häufig vorkommen, so nimmt die Entropie den Wert 4,9 an.

Nun sagt uns aber die Praxis, daß einerseits manche Buchstaben viel häufiger vorkommen als andere („e“ ist z.B. der häufigste Buchstabe), andererseits auch die Anordnung der Buchstaben nicht völlig gleich-wahrscheinlich ist. Man kann bspw. annehmen, daß die Wahrscheinlichkeit, daß auf ein „m“ ein „n“ folgt viel niedriger ist, als daß auf ein „m“ ein „o“ folgt.

Beachten wir nur den ersten Fall, also das unterschiedlich wahrscheinliche Auftreten der Buchstaben, und errechnen danach die Entropie, so verringert sich der Wert auf 4,1.

Was können wir daraus schließen? Ganz klar ergibt sich aus dieser Analyse, daß bei Verwendung einer geschickten Strategie das Erraten deutscher Worte erfolgreicher sein wird, als wenn die Wahl der Buchstaben nach reinem Zufallsprinzip erfolgt.

2.4. Statistik und die „Gewinnung“ von Information

„Eine statistische Analyse ist eine Art von Abbildung: Ausgehend von den Rohdaten versuchen wir durch Anwendung einer solchen Abbildung *Information* über das betrachtete System zu gewinnen.“ So, oder so ähnlich kann man sich den Vorgang, der einer statistischen Analyse innewohnt vorstellen. Schließlich liegt am Anfang oft ein unübersichtlicher Datenfriedhof, mit dem wir nicht viel anfangen können, am Ende der Analyse aber ein „klares“ Ergebnis, das für uns anschaulich, repräsentativ ist. Verwenden wir also das Wort *Information* in der Art und Weise, daß es *Zugänglichkeit* oder *Bedeutung* ausdrückt, so haben wir tatsächlich *Information* gewonnen.

Betrachten wir aber den Vorgang mit den Augen der Informationstheorie, so bleibt die Information im „besten“ Fall konstant, üblicherweise wird aber Information verloren werden. Wie kommt das? Wie schon im ersten Abschnitt angedeutet liegt zu Beginn oft eine unübersichtliche Menge von Daten vor, Rauschen überlagert diese Daten, ... Die statistische Analyse zielt ja oftmals gerade darauf hinaus, einen nicht unerheblichen Teil der in den Rohdaten vorhandenen Information herauszufiltern oder zu verallgemeinern um die Daten für uns besser zugänglich zu machen.

„Verallgemeinerungen“ wie das Berechnen von Mittelwerten oder das Entfernen von Rauschen reduzieren aber klarerweise den Informationsgehalt. Wie kann man nun diesen scheinbaren Widerspruch auflösen, daß durch Reduktion von *Information* im informationstheoretischen Sinne *Information* im subjektiv/menschlichem Sinne gewonnen werden kann?

Gerade das Reduzieren des (informationstheoretischen) Informationsgehaltes kann das Bild auf das „wesentliche“ freigeben. Man kann diesen Vorgang als eine Art der Fokussierung auf den Teil des Systemes verstehen, der uns im Moment konkret interessiert. (Dies muß nicht immer derselbe Teil sein. Aus einem Datensatz kann unter Umständen je nach Fragestellung verschiedene Information extrahiert werden.) Was ist nun aber wesentlich? Woraus resultiert der scheinbare Gewinn an Information oder Bedeutung nach der Analyse?

Klarerweise ist es in Wahrheit die *Verknüpfung* von Daten in unserem Gehirn. Dies läßt sich einfach anhand eines Beispiels illustrieren:

Für einen ungeschulten Betrachter, ist ein Chromatogramm einer Luftprobe nicht mehr als ein Blatt Papier mit einer zackigen-schwarzen Linie. Für den geübten Analytiker hingegen eröffnet sich sofort ein klares Bild, was diese Probe bedeuten könnte oder welche Fakten aufgrund dieser Graphik ausgeschlossen werden können. Warum? Einfach deshalb, weil sein Gehirn dieses eine „Muster“ automatisch mit hunderten oder gar tausenden anderen Mustern von Chromatogrammen in seinem Gedächtnis vergleicht, Verbindungen herstellt, assoziiert usw. Daher die Bedeutung für *ihn*. Das *Bild* des Baumes auf der Netzhaut wird für uns Erwachsene sofort zum *tatsächlichen* Baum, weil wir eben diesen oder ähnliche „Muster“, „Bilder“ schon unzählige Male gesehen haben. Das Baby muß dies erst mühsam erlernen.

Somit löst sich der scheinbare Widerspruch auf. Die Reduktion der Information schafft uns einen freien Blick auf die Teile der innewohnenden Information, die wir mit unserem Gehirn zu verarbeiten imstande sind und die wir assoziieren, verknüpfen können. Der Informationsgewinn kommt aus der durch Assoziation und Interpretation von *außen* eingebrachten Information, und steckt nicht im System selbst. Bei

Informationstheorie

Assoziationen

nur oberflächlicher Betrachtung (und man ist geneigt diese Quelle zu übersehen) erscheint es, als Entstände neue Information aus der Anwendung einer mathematischen Methodik. Folglich muß man sich im klaren sein, daß man durch Anwendung statistischer Verfahren nichts aus dem System „herausholen“ kann, was nicht an sich schon im System befindet. Man kann letztlich nur die Form der Darstellung verändern, bzw. für die jeweilige Problemstellung überflüssige Aspekte „ausblenden“.

Wir sind von unserer Evolution her nicht darauf getrimmt lange Zahlenkolonnen, Matrizen oder hochdimensionale Systeme intuitiv verstehen zu können. Daher auch der vielfältige Einsatz visualisierender Methoden und bildgebender Verfahren. Für den Rechner ist dies exakt umgekehrt. Die Leistungsfähigkeit und Nützlichkeit von Computersystemen eröffnet sich für uns ja gerade aus der Tasche heraus, daß sie speziell zu diesen Dingen befähigt sind, mit denen wir Schwierigkeiten haben. Das Speichern, Sortieren, Manipulieren von riesigen Datenbeständen ist für Computer ein leichtes, für uns ungeheuer schwierig. Dafür sind wir auf der anderen Seite nach wie vor selbst Großrechnern im Erkennen von Mustern weit überlegen.

Die Kunst der
Statistik

Der Weg der statistischen Analyse sollte also von der unanschaulichen Datenbank zu einer Form führen, mit der wir als Menschen (als Experten auf einem bestimmten Gebiet) etwas anzufangen wissen. So gesehen wird der Statistik Leben eingehaucht und sie verwandelt sich von der spröden, mathematischen Wissenschaft zum beinahe künstlerischen Werkzeug im Herauskrystallisieren verborgener Bedeutungen.

3. * Statistische Tests

3.1. Einleitung

In den nächsten Kapiteln werden wir uns zunächst mit der korrekten Angabe von Meßergebnissen, mit Ausreißertests und dann mit dem Vergleich von Messungen auseinandersetzen. Unter Ausreißertests versteht man die Fragestellung, ob einzelne, deutlich vom Rest der Daten abweichende Werte eben als Ausreißer bezeichnet werden können, und unter welchen Voraussetzungen sie vor weiteren Analysen entfernt werden dürfen. Ein Beispiel für den Vergleich von Messungen könnte die Fragestellung sein, ob sich die Mittelwerte zweier unabhängiger analytischer Meßreihen signifikant unterscheiden, oder nicht.

Für alle diese Kapitel ist es notwendig, die grundlegenden Ideen die hinter statistischen Tests stehen, bzw. sich die Begriffe Vertrauens- und Irrtumswahrscheinlichkeit zu verinnerlichen. Diese Prinzipien und die entsprechenden Termini können in den folgenden Abschnitten nur im Überblick beschrieben werden. Detailliertere Betrachtungen hierzu finden sich u.a. in [7, 12, 21]

3.2. Hypothesen

Die grundlegende Idee aller Testverfahren ist das Aufstellen von *Hypothesen*, die dann auf einem bestimmten Signifikanzniveau *getestet* werden. Der Test entscheidet schließlich, ob die gestellte *Hypothese* zugunsten einer *Alternativhypothese* verworfen werden kann, oder beibehalten werden muß.

Diese Aussage ist noch ein wenig abstrakt. Eine Hypothese könnte beispielsweise sein: „Die Mittelwerte sind gleich“. Diese Hypothese wird auch mit H_0 bezeichnet. Die Alternativhypothese H_1 würde in diesem Fall lauten: „Die Mittelwerte sind unterscheiden sich signifikant.“ Wichtig ist es zu verstehen, daß „gleich“ nicht ein „=“ im mathematischen Sinne bedeutet, denn jede analytische Meßserie ist ja nur eine Stichprobe aus einer (unendlich großen) Grundgesamtheit. Somit ist immer mit einer gewissen zufälligen Streuung zu rechnen. Dieses „gleich“ bedeutet vielmehr: *Unterscheiden sich die Parameter nur aufgrund der immer vorhandenen zufälligen Streuung der Meßwerte, oder ist ein signifikanter Unterschied vorhanden.*

Schon aus diesen Formulierungen wird klar, daß es hier keinen für jeden Fall eindeutigen Test und somit eindeutige Entscheidung geben *kann*. Es ist vielmehr nur möglich die Größe des Unterschiedes zu fassen und anzugeben, wie hoch letztlich die Wahrscheinlichkeit ist, daß ein Unterschied vorliegt. Doch dazu später mehr in den Kapiteln, die die statistischen Tests behandeln.

Beispiele

Ein anderes Beispiel wäre: Die Hypothese H_0 bezeichnet den Fall, daß der ausreißerverdächtige Wert x_7 *kein* Ausreißer ist. Die Alternativhypothese H_0 würde dann die andere Möglichkeit beschreiben, nämlich daß es sich bei x_7 um einen Ausreißer handelt.

Beweisen von Nullhypothesen

Ein statistischer Test soll also feststellen helfen, ob Unterschiede nur zufälliger Natur sind, oder ob sie als signifikant zu bezeichnen sind, und somit die Nullhypothese zu verwerfen ist. Sollte das Ergebnis eines Testes sein, daß die Nullhypothese *nicht* abgelehnt werden darf, so ist dies **kein Beweis** für die Richtigkeit der Nullhypothese! Mit anderen Worten: **Nullhypothesen können niemals bewiesen, sondern nur abgelehnt werden.**

Aus diesem Grund wird auch die Ablehnung der Nullhypothese als *starke Aussage*, die Beibehaltung hingegen als *schwache Aussage* bezeichnet.

3.3. Vertrauens- und Irrtumswahrscheinlichkeit

Vertrauensbereich

Den Begriffen *Vertrauens-* und *Irrtumswahrscheinlichkeit* begegnet man in der Statistik an verschiedenen Stellen. Beispielsweise setzen wir uns im Abschnitt 4.10 mit dem Vertrauensbereich auseinander. Die Problematik ist folgende: Es liegen eine gewisse Anzahl von Meßwerten (Stichprobe) vor, aus denen ein Parameter der Grundgesamtheit — also in diesem Fall der Mittelwert — abgeschätzt werden soll. Es wird festgestellt werden, daß die „Präzision“ der Schätzung von verschiedenen Faktoren abhängig ist.

Führen wir nun die Begriffe *Vertrauens-* und *Irrtumswahrscheinlichkeit* ein: Die Vertrauenswahrscheinlichkeit wäre die Wahrscheinlichkeit, daß der wahre Wert (also der Mittelwert der Grundgesamtheit) tatsächlich im angegebenen Intervall des Vertrauensbereiches liegt. Die Irrtumswahrscheinlichkeit ist komplementär definiert und bezeichnet die Wahrscheinlichkeit, daß der wahre Wert *außerhalb* der bezeichneten Grenzen liegt. Für die Irrtumswahrscheinlichkeit wird das Symbol α verwendet. Dieses α muß gewählt werden, und entscheidet über die Größe des Intervalles, denn: je größer das Intervall ist, desto kleiner die Wahrscheinlichkeit, daß der wahre Wert außerhalb dieser Grenzen zu liegen kommt.

Gängige Werte für die Irrtumswahrscheinlichkeit sind etwa: $\alpha = 0,1$; $\alpha = 0,05$ und $\alpha = 0,01$. Die Vertrauenswahrscheinlichkeit errechnet sich nach $P = (1 - \alpha)$ und beträgt daher in diesen Fällen 0,90; 0,95 und 0,99 (oder in Prozent ausgedrückt: 90 %, 95 % und 99 %).

3.4. Fehler, Power eines Tests

Tests

Bei den in den nächsten Kapiteln beschriebenen Tests beschreibt die Irrtumswahrscheinlichkeit α die Wahrscheinlichkeit, daß die Nullhypothese abgelehnt wird, obwohl tatsächlich die Nullhypothese wahr gewesen wäre. Im Falle des Vergleiches von Mittelwerten würde dies bedeuten, daß man feststellt, daß sich zwei Mittelwerte signifikant unterscheiden, obwohl dies „in Wahrheit“ nicht der Fall ist. Dieser Fehler wird auch Fehler 1.Art genannt (siehe Tab. 3.1).

Testergebnis	richtige Entscheidung wäre	
	H_0 ist wahr	H_1 ist wahr
H_0 wird beibehalten	richtige Entscheidung	Fehler 2.Art
H_0 wird abgelehnt	Fehler 1.Art	richtige Entscheidung

Tabelle 3.1.: Fehler 1. und 2.Art in der Teststatistik. Fehler 1.Art werden auch α Fehler, Fehler 2.Art β Fehler genannt.

Da die Irrtumswahrscheinlichkeit auch *Signifikanzniveau* genannt wird, formuliert man auch: *Die Alternativhypothese wurde auf Signifikanzniveau α angenommen.* Mit anderen Worten: Man irrt sich in $\alpha \cdot 100$ % der Fälle.

Von einem Fehler 2.Art spricht man unter der Voraussetzung, daß die *Alternativhypothese* H_1 wahr gewesen wäre, dennoch die *Nullhypothese* H_0 beibehalten wird. Fehler 2.Art, auch β Fehler genannt, hängen unmittelbar mit der Trennschärfe des Tests, *Power* genannt, zusammen. Tests mit geringer Power werden daher *konservativ* genannt, da sie, salopp formuliert, im Zweifelsfall die Nullhypothese beibehalten.

Ein anderer wichtiger Einflußfaktor ist die Anzahl der Daten, auf die ein Test angewandt wird. Grundsätzlich reduziert sich bei zunehmender Anzahl an Meßdaten die Wahrscheinlichkeit Fehler 2.Art zu begehen. Eine Entscheidung läßt sich oft „erzwingen“, indem man sehr viele Datenpunkte mißt. Leider sind hier in der Praxis üblicherweise (ökonomische) Grenzen gesetzt.

4. Angabe eines Meßergebnisses mit Vertrauensbereich

4.1. Einleitung

Eine der grundlegendsten „statistischen Aufgaben“, mit denen man in der *analytischen Praxis* konfrontiert ist, ist die Angabe eines Ergebnisses, das aus mehreren Messungen besteht. Zusätzlich ist oftmals eine Angabe der Genauigkeit des Ergebnisses gewünscht.

Aufgrund der Tatsache, daß die Angabe des Ergebnisses üblicherweise gleich dem *Berechnen des Mittelwertes* ist, und dieser wiederum jedem aus dem alltäglichen Leben geläufig ist, macht man sich keine großen Gedanken über die Verwendung dieses *statistischen Parameters*. Es wird vielleicht erstaunen, daß bereits die Anwendung dieses „einfachen“ und scheinbar allgemein-verständlichen Parameters an verschiedene Voraussetzungen gekoppelt ist. Die Nichtbeachtung dieser Voraussetzungen kann im schlimmsten Fall das Ergebnis einer Meßserie grob verfälschen!

Weiters ist ein Maß zu finden, mit dessen Hilfe es möglich ist, eine Genauigkeit oder Zuverlässigkeit des Ergebnisses anzugeben. Üblicherweise erfolgt dies durch Angabe des Mittelwertes **und** eines *Intervalles* in dem mit einer bestimmten Vertrauenswahrscheinlichkeit (z.B. 95 %) der wahre Wert liegt. Die Angabe dieses Intervalles ist oft mit einigen Irrtümern behaftet und wird daher auch anhand verschiedener Beispiele erläutert.

In manchen Fällen liegen Daten vor, bei denen die Angabe des Mittelwertes als Ergebnis nicht anzuraten ist, sondern vielmehr robuste Methoden wie *Median* und *Quartil*. Auch auf diese Problematik wird erfahrungsgemäß oft zu wenig Augenmerk gelegt.

Aus diesem Grund ist das Verständnis dieses Abschnittes vielleicht das wichtigste des gesamten Skriptums. Auch bei der Prüfung wird entsprechend viel Wert auf diese Grundlagen gelegt werden.

Einen Überblick über dieses Kapitel findet sich in Abb. 4.1. Anhand dieses Strukturgrammes kann man die Vorgangsweise für praktische Fälle leicht ermitteln und dann evt. in den entsprechenden Abschnitten nachlesen.

Mittelwert

Prüfung

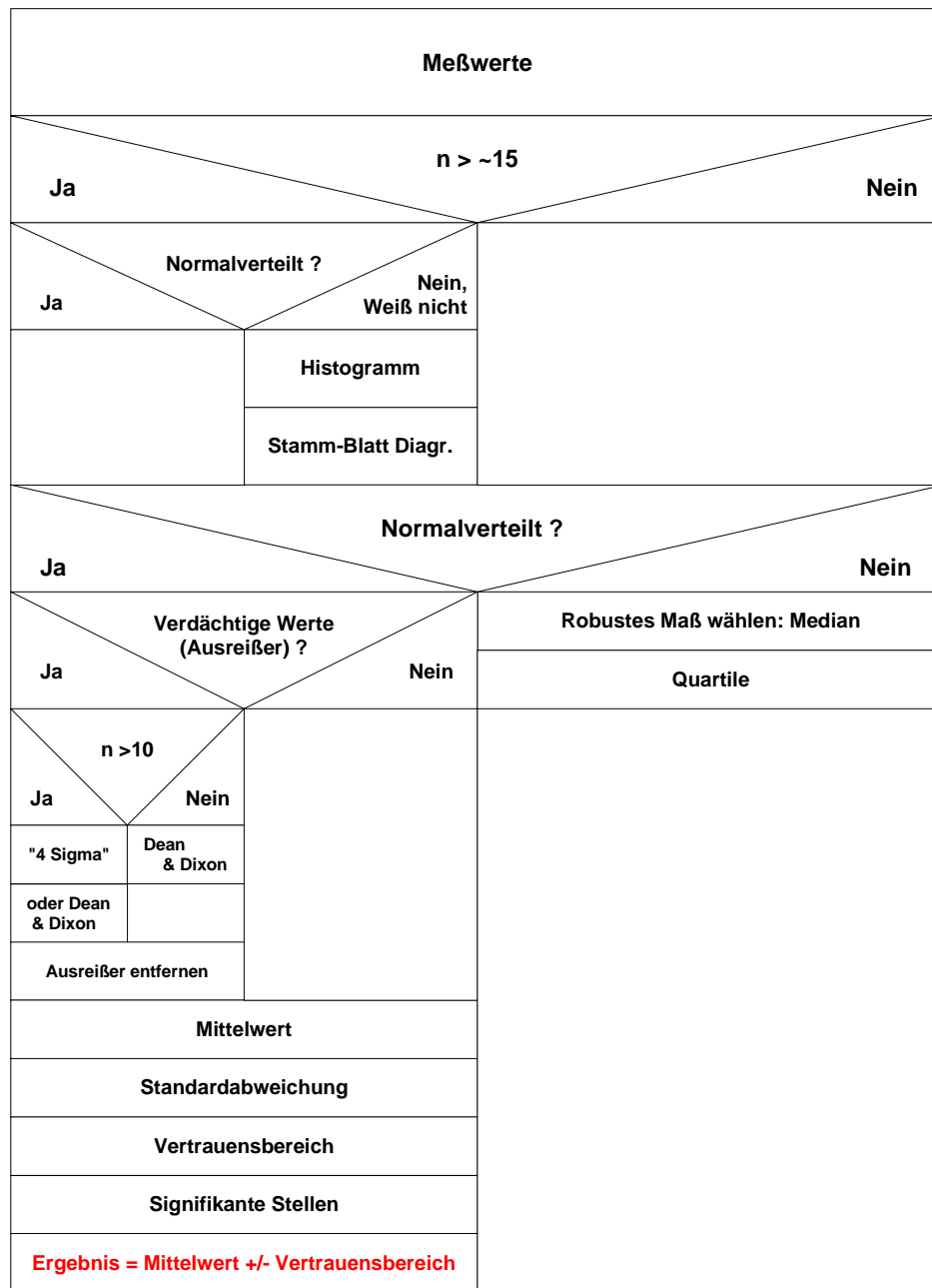


Abbildung 4.1.: Angabe von Meßergebnissen: Dieses Struktogramm dient als Übersicht wie man bei einem praktischen Problem vorgehen kann. Die einzelnen Schritte kann man dann in den entsprechenden Abschnitten nachlesen. Zur Erläuterung: Die Rechtecke geben Verfahren oder Zustände an, z.B. Histogramm bedeutet Zeichnen des Histogrammes. Entscheidungen/Fragen werden durch geteilte Rechtecke markiert: $n > 15$ ist also die „Frage“, ob die Anzahl der Daten größer als 15 ist, wenn ja wird der linke Weg (Ja), wenn nein, der rechte Weg beschritten.

4.2. Beispiele

Beispiel 1 Es wird gravimetrisch die Konzentration von Bromid (Br^-) aus wässriger Lösung durch Fällung mit AgNO_3 bestimmt. Es werden drei Bestimmungen durchgeführt, die gemessenen Konzentrationen sind: 5,31; 5,08 und 5,44 $\text{mg} \cdot \text{l}^{-1}$.

Beispiel 2 Bei der Bestimmung von Blei im Wasser werden sieben Meßwerte mittels *Atomabsorptionsspektroskopie* (AAS) gemessen: 3,81; 3,82; 3,91; 3,90; 4,01; 4,03 und 4,11 ppb.

Beispiel 3 Es werden mittels *Gaschromtographischer Trennung* (GC) und Detektion mittels *Elektronen-Einfang Detektor* (Electron Capture Detector-ECD) polychlorierte Biphenyle (PCBs) untersucht. Zunächst werden drei Bestimmungen vorgenommen. Die Meßwerte für ein bestimmtes PCB lauten: 0,38; 0,40; 0,19 ppb. Da der dritte Wert deutlich von den anderen beiden abweicht, beschließt der Analytiker noch zwei weitere Messungen vorzunehmen. Die Werte der weiteren zwei Messungen sind: 0,36; 0,42. Da es sich um ein Routineverfahren handelt, das seit Jahren in dieser Form ausgeführt wird, konnte eine Standardabweichung für die Methodik (also für die „Grundgesamtheit“) abgeschätzt werden¹, sie lautet: $\sigma = 0,018$.

Beispiel 4 Es liegen Meßwerte von Nitrat im Trinkwasser von 33 Brunnen einer Gemeinde vor: 7,02; 7,48; 7,64; 7,9; 8,03; 8,17; 8,27; 8,5; 8,66; 8,67; 8,8; 8,82; 7,8; 8,1; 8,89; 8,9; 8,9; 8,92; 8,94; 8,94; 8,96; 8,99; 9,13; 9,2; 9,2; 10; 9,39; 8; 9,5; 7,61; 7,23; 7,04 und 10

Aufgabe: Es sollen die Ergebnisse mit Vertrauensbereich (im Beispiel 4 ein mittlerer Wert) korrekt angegeben werden.

In den folgenden Abschnitten wird das statistische „Rüstzeug“ besprochen, das notwendig ist, um diese Beispiele korrekt zu lösen. Im letzten Abschnitt auf Seite 49 werden dann die Beispiele besprochen und die korrekten Lösungen angegeben.

4.3. Arten von Fehlern

Vorweg sollte ein kurzer Blick auf die möglichen Fehler einer analytischen Messung gelegt werden. Man unterscheidet üblicherweise zwischen zwei Arten von Fehlern:

Als sogenannte *Zufallsfehler* bezeichnet man die bei jeder Messung auftretenden Abweichungen einer Messung vom „wahren“ Wert. Diese Abweichungen können von unterschiedlichen Quellen kommen. Beispiele könnten elektronisches Rauschen, nicht exakt gleiches Volumen der einzelnen Tropfen beim Titrieren, leichte Unterschiede in den „Randbedingungen“ der Messung wie Temperatur, Druck, etc. sein².

Davon grundsätzlich verschieden sind die *systematischen Fehler*. Dabei handelt es sich um Fehler, die grundsätzlicher Natur sind. Diese Fehler beeinflussen *alle* Einzelmessungen. Fehler dieser Art sind beispielsweise: Falsche Einstellung der Titerlösung, falsche Kalibration oder Eichkurve eines Photometers, defekter Thermometer bzw. falsch gewählte Temperatur, ...

Unter *Reproduzierbarkeit* versteht man nun die Größe der Fehler der ersten Kate-

Zufallsfehler

Systematischer Fehler

Reproduzierbarkeit und Richtigkeit

¹Details dazu werden in den weiteren Abschnitten bzw. bei der Lösung des Beispiels gegeben.

²Diese Fehler finden sich grundsätzlich bei jeder Art von Messung. Sind keine Schwankungen festzustellen, ist die ein Indiz dafür, daß die Empfindlichkeit des Meßsystems nicht fein genug eingestellt ist.

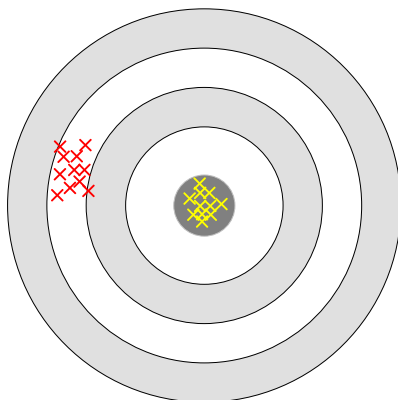


Abbildung 4.2.: Illustration des *zufälligen* und *systematischen* Fehlers anhand der Treffer auf einer Zielscheibe:

Man erkennt zwei „Gruppen“ von Treffern: Einerseits solche, die um die Mitte streuen. Dieser Schütze hat keinen *systematischen*, wohl aber einen gewissen *zufälligen* Fehler begangen. Andererseits die Gruppe, die deutlich außerhalb der Mitte streut. Der Schütze dieser Gruppe hat nicht nur einen *zufälligen*, sondern auch noch einen deutlichen *systematischen* Fehler begangen (möglicherweise ist der Lauf verzogen). Die Werte streuen eben nicht um das Zentrum (den „wahren“ Wert), sondern vielmehr um einen „falschen“ Wert ausserhalb.

gorie (also der Zufallsfehler), unter *Richtigkeit* des Ergebnisses versteht man folglich die (weitgehende) Abwesenheit von Fehlern der zweiten Kategorie (also der systematischen Fehler). Der grundsätzliche Unterschied zwischen diesen beiden Arten von Fehlern wird in Abb. 4.2 anhand eines Beispiels verdeutlicht.

4.4. Grundgesamtheit, Stichprobe

Für die Anwendung statistischer Methoden ist es wichtig zwischen *Grundgesamtheit* und *Stichprobe* zu unterscheiden. Diese beiden Begriffe lassen sich am einfachsten anhand eines Beispiels erklären:

Möchte man z.B. feststellen, wie es um die Gesundheit österreichischer Schüler bestellt ist, so kann man einerseits *alle* österreichischen Schüler untersuchen — also die Grundgesamtheit — oder man beschränkt sich (aus finanziellen Gründen beispielsweise) auf eine *Auswahl* aus dieser Grundgesamtheit — also auf eine Stichprobe.

Die Grundgesamtheit ist also die *gesamte Menge aller Elemente* eines untersuchten Systems, die Stichprobe hingegen nur eine *Auswahl*. Nun hat man es nicht immer mit einer *endlichen großen* Grundgesamtheit zu tun (wie im obigen Beispiel), oft ist die Grundgesamtheit auch *unendlich groß*. Für die Fälle, die in der (analytischen) Chemie auftreten, ist üblicherweise die Grundgesamtheit unendlich groß, da die Grundgesamtheit ja die Menge *aller möglichen Messungen* umfaßt. Diese Anzahl ist natürlich vom prinzipiellen Standpunkt her unendlich groß, es macht ja keinen Sinn zu sagen: „Die Grundgesamtheit der photometrischen Messungen einer Probe beträgt 500 Messungen.“ Warum nur 500 und nicht 501 oder 1000, etc.

Im Falle von analytischen Messungen ist die jeweilige Meßserie folglich eine Stichprobe der (unendlich großen) Grundgesamtheit aller möglichen Messungen. Die Grundgesamtheit ist daher in den uns interessierenden Fällen meist nicht direkt zugänglich. Aus der Unterscheidung zwischen *Stichprobe* und *Grundgesamtheit* ergeben sich schon intuitiv einige Konsequenzen:

unendlich große
Grundgesamtheit

- Statistische Parameter — also z.B. der Mittelwert — werden sich unterscheiden, je nachdem ob man sie aus der Grundgesamtheit oder aus einer Stichprobe ermittelt. Dies gilt umso mehr als die Stichprobe oftmals viel kleiner ist als die Grundgesamtheit.
- Zieht man unterschiedliche Stichproben aus einer Grundgesamtheit, werden sich die errechneten statistischen Parameter zwischen den unterschiedlichen Stichproben unterscheiden.
- Die „Genauigkeit“ der errechneten Parameter (also die Ähnlichkeit, mit dem Parameter der Grundgesamtheit) wird umso besser sein, je größer die Stichprobe ist.
- An dem „Schülerbeispiel“ läßt sich klar erkennen, daß der Wahl der Stichprobe entscheidende Bedeutung zukommen kann: Es führt sicherlich zu unterschiedlichen Ergebnissen, wenn in einem Fall nur Schüler aus ländlichen Regionen gewählt werden, im anderen Fall jedoch Schüler aus Großstädten.

Konsequenzen

4.5. Verteilung von Daten — Normalverteilung

4.5.1. Einleitung

1. Man wirft eine Münze 50 mal. Wie oft wird Kopf erscheinen? Wie hoch ist die Wahrscheinlichkeit, daß beim 51. Wurf Kopf erscheint?
2. Eine Versicherungsgesellschaft möchte das Risiko errechnen, daß eine Frau, die heute 65 Jahren alt ist, älter als 75 wird.
3. Mediziner testen eine neue Therapie. Die Versuchsleiter sind daran interessiert, ob die neue Behandlungsform erfolgreicher ist als herkömmliche Therapien.

Alle diese Fälle haben eines gemeinsam: Man ist weniger am Einzelfall interessiert, sondern vielmehr daran, aus einer großen Anzahl von Daten allgemeine Schlüsse zu ziehen. Auch der Faktor der Wahrscheinlichkeit spielt bei den Beispielen eine wichtige Rolle.

Zu **Beispiel 1**: Aus der Betrachtung der einzelnen Werte wird man nicht allzu viele allgemein gültige Schlüsse ziehen können, da ja jedes Einzelereignis von zufälligem Ausgang ist. (Das ist ja die Voraussetzung eines Glücksspiels.) Zählt man jedoch die Anzahl der aufgetretenen „Köpfe“ und „Adler“ zusammen, so wird man schnell feststellen, daß das Verhältnis, in dem beide auftreten, etwa 1:1 ist. Aus dieser Betrachtung heraus wird auch die Beantwortung der Frage einfach: Kopf wird etwa

Beispiele

25 mal erscheinen, und die Wahrscheinlichkeit, daß beim 51. Wurf Kopf erscheint ist 0,5 also 50 %.³

Zu **Beispiel 2**: Auch in diesem Beispiel wird es nicht möglich sein ein konkretes Risiko für *eine bestimmte* Frau auszurechnen. Es besteht allerdings die Möglichkeit, eine große Anzahl von Personen (bspw. alle Österreicherinnen über 60) über mehrere Jahre zu untersuchen und das durchschnittliche Risiko einer Person dieser Personen-Gruppe zu bestimmen. Das ist ja auch das Prinzip der Versicherung: über den Einzelfall kann natürlich keine konkrete Aussage getroffen werden, eine größere Anzahl von Personen hingegen wird kalkulierbar.

Zu **Beispiel 3**: Auch hier wird der Mediziner nicht in erster Linie am *Individuum* interessiert sein. Bei einzelnen Personen sind die Einflüsse einfach zu groß als daß man aus einem Einzelfall einen positiven oder negativen Schluß ziehen könnte. Mit Hilfe einer größeren Testgruppe hingegen kann die Frage unter Umständen beantwortet werden.

Ein wichtiges Hilfsmittel um eine solche „Abstraktion“ aus einer größeren Anzahl von Daten vorzunehmen ist das *Histogramm* oder das *Stamm-Blatt Diagramm*.

4.5.2. Histogramme

Ein Histogramm ist eine statistische Graphik mit deren Hilfe es möglich ist, Rückschlüsse über die Verteilung der Daten zu ziehen. Im Falle des ersten Beispiels wäre das Zeichnen des „Histogramms“⁴ noch recht einfach: Man könnte ein Balkendiagramm zeichnen, bei dem die Höhe eines Balkens für die Anzahl der „Köpfe“, die des anderen für die Anzahl der „Adler“ steht. Man kann sich leicht vorstellen, daß in diesem Beispiel beide Balken etwa gleich hoch sein werden.

Zeichnen von
Histogrammen

Um ein Histogramm für den zweiten Fall zu zeichnen ist schon etwas mehr Vorarbeit nötig: Angenommen die Anzahl der Personen, die in die angegebene Gruppe fällt, beträgt mehrere zehn- wenn nicht hunderttausend. Jeden einzelnen Fall darzustellen wäre weder sehr anschaulich, noch könnte man daraus allgemeine Schlüsse ziehen. Um ein Histogramm zu zeichnen wählt also folgenden (allgemein gültigen) Weg:

1. Zunächst wird die Spanne, über die sich die Daten erstrecken in Gruppen, sogenannte *Klassen*⁵, eingeteilt, wobei die Intervalle gleich groß sein müssen⁶ und außerdem keine Lücken entstehen dürfen. Mit anderen Worten: Für jeden

³In der Statistik arbeitet man gerne mit auf 1 normierten Wahrscheinlichkeiten(P). $P = 1$ bedeutet daher, daß nur das entsprechende Ereignis vorkommt (also Wahrscheinlichkeit gleich 100 %), $P = 0,5$ heißt Wahrscheinlichkeit gleich 50 %. $P = 0$ meint folglich, daß das Ereignis überhaupt nicht vorkommt (also Wahrscheinlichkeit gleich 0 %).

⁴In diesem Beispiel ist es nicht ganz korrekt von einem *Histogramm* zu sprechen, da Histogramme nur von metrisch skalierten, stetigen Daten berechnet/gezeichnet werden dürfen. Die Ergebnisse eines Münzwurfes entsprechen nicht diesen Voraussetzungen, daher handelt es sich zunächst nur um ein Balkendiagramm, daß die Häufigkeiten der beiden Fälle veranschaulicht.

⁵Diese Klassen sind üblicherweise *halboffene Intervalle* $[a, b)$. Eine Klasseneinteilung zum Alter könnte man folglich so definieren: $[20, 30)$, $[30, 40)$, $[40, 50)$ Da diese Intervalle *rechts-halboffen* sind, bedeutet dies bspw. für das erste Intervall, daß der Wert 20 noch im ersten Intervall aufscheint, der Wert 30 jedoch nicht mehr.

⁶Eventuell kann man hier verschiedene Intervall-Größen ausprobieren.

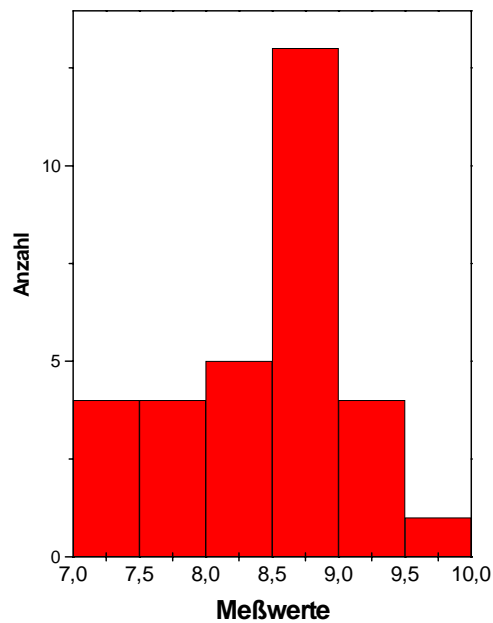


Abbildung 4.3.: Histogramm des Beispiels: *Nitrat im Trinkwasser*.

Wert der Datenmenge muß *genau eine* Gruppe existieren und jeder Datenwert darf nur in genau einer Gruppe aufscheinen.

In diesem Beispiel könnte man folgende Gruppen bilden: Anzahl der lebenden Personen zwischen 55–60, 60–65, 65–70, 70–75, 75–80, 80–85, 85–90, 90–95, 95–⁷ Jahren.

2. Dann werden die Anzahl der Personen in den jeweiligen Gruppen gezählt. Diese Größe wird auch *Klassenhäufigkeit* genannt.
3. Schließlich zeichnet man eine Balkendiagramm, in dem jede Gruppe einem Balken und die Höhe des Balkens der ermittelten Anzahl der Fälle in der Gruppe entspricht⁸.

Dies wird konkret am Beispiel des „Nitrat im Trinkwasser“ von Seite 29 durchgeführt. In Abb. 4.3 ist das Histogramm dieses Falles dargestellt. Für jede Gruppe wurde eine „Breite“ von 0,5 gewählt, der niedrigste Wert liegt bei 7. D.h. die Balken stehen für die Gruppen: 7,0–7,5; ... ; 9,5–10. Auf der x - Achse (= Abszisse) sind die Gruppen aufgetragen, die y - Achse (= Ordinate) repräsentiert die Anzahl der Werte, die in jeder Gruppe auftreten.

Beispiel

⁷Nun kann man natürlich kritisieren, daß die letzte Gruppe eigentlich nicht korrekt ist, da sie ja eindeutig größer ist, als die anderen. Allerdings ist die Anzahl der über 100-jährigen im Verhältnis zu der Anzahl in den anderen Gruppen vermutlich nahezu vernachlässigbar gering, und fällt somit kaum ins Gewicht. Wollte man ganz exakt vorgehen, könnte man natürlich noch mehr Gruppen definieren, die schließlich wirklich alle Bedingungen erfüllen, also vermutlich bis zu 105-110.

⁸Korrektweise muß man darauf hinweisen, daß nicht die *Höhe*, sondern vielmehr die *Fläche* der Balken die relative Klassenhäufigkeit repräsentiert. Da wir jedoch in der chemischen Praxis nahezu ausschließlich mit konstanten Klassenbreiten arbeiten, sind beide Größen natürlich equivalent.

An diesem Beispiel sollte man sich in dieser Stelle v.a. merken, daß es sich um eine *unsymmetrische* Verteilung handelt. Derartige Verteilungen werden *schiefe Verteilung* genannt, doch dazu später mehr.

diskrete
Verteilung

Die hier dargestellte Verteilung ist eine sogenannte *diskrete* Verteilung. Das liegt daran, daß nur eine endliche Anzahl von Werten zur Verfügung steht, und folgerichtig auch die Breite der Gruppen einen bestimmten (eben diskreten) Wert annimmt. Stellen wir uns nun vor, die Anzahl der Meßwerte würde ins unendliche anwachsen, die Breite der Gruppen hingegen unendlich schmal werden, so würde man schließlich eine sogenannte *kontinuierliche* Verteilung erhalten. Anschaulich bedeutet das, daß die Balken des Histogramms immer schmaler werden, solange, bis sie unendlich schmal sind und in eine kontinuierliche Verteilung übergehen (In Abb.4.4 ist dies veranschaulicht, wobei es hier nicht auf die *quantitativen* Zusammenhänge ankommt — die Zahlen sind beliebig — vielmehr soll es qualitativ das Prinzip dieses Überganges zeigen.). Die *Normalverteilung*, die im nächsten Abschnitt besprochen wird und für die praktische Arbeit des analytischen Chemikers die bedeutendste ist, ist eine solche kontinuierliche Verteilung.

kontinuierliche
Verteilung

4.5.3. Stamm-Blatt-Diagramm

Das sogenannte „Stamm-Blatt“ Diagramm soll nicht unerwähnt bleiben. Es handelt sich um eine einfache Möglichkeit, auch händisch (also z.B. im Labor) die Verteilung mittlerer Datenmengen zu untersuchen [9]. Die Vorgangsweise ist recht einfach:

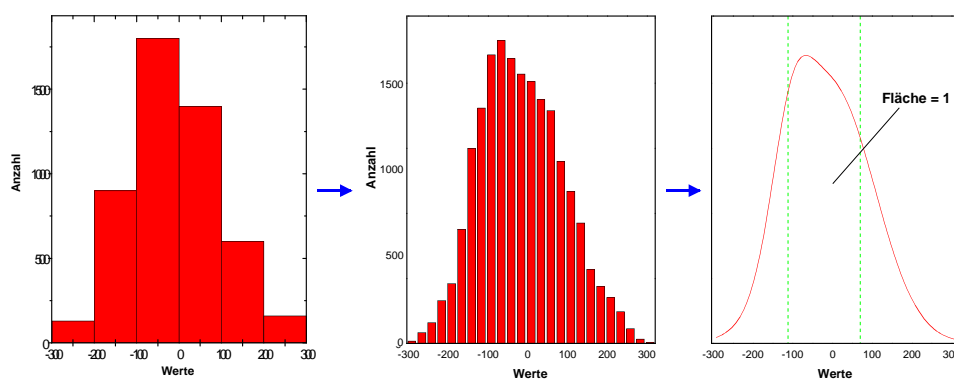


Abbildung 4.4.: Der „Weg“ von der diskreten zur kontinuierlichen Verteilung:

Wie im Text beschrieben, kann man sich vorstellen, daß die Anzahl der Meßwerte sukzessive zunimmt, die Intervalle des Histogrammes immer kleiner wird bis unendlich viele Werte vorliegen. Die linken beiden Graphiken stellen noch diskrete Verteilungen dar, die rechte Figur ist die bei unendlich vielen Werten entstehende kontinuierliche Verteilung, wobei dann die Fläche unter der Kurve als Wahrscheinlichkeit interpretiert wird, und die Gesamtfläche auf 1 normiert ist (= 100 %).

(Anmerkung: Es ist wichtig zu beachten, daß sich die Skalierung der y-Achse natürlich im letzten Schritt ändert!)

1. Sortieren der Daten (Dies ist nicht unbedingt erforderlich, erleichtert nur das Auffinden der Werte. Falls das Diagramm von Hand gezeichnet wird, sollte man diesen Schritt evt. weglassen.). Zeichnen von Stamm-Blatt-Diagrammen
2. Einteilen des Datenbereiches in Intervalle. Also z.B. Intervallgröße von 1; 0,5; 0,25 oder bei kleineren Werten auch kleiner⁹.
3. Die erste Ziffer des Intervalles wird als „Stamm“ untereinander aufgetragen.
4. Die Zahlen werden auf die nächste Ziffer gerundet, und diese jeweils neben den passenden Stamm geschrieben.

Die Vorgangsweise wird anhand eines Beispielen sofort klar. Es wird das Diagramm für dasselbe Beispiel wie im Histogramm-Abschnitt gezeichnet: **Beispiel**

Zunächst werden die Werte sortiert: 7,02; 7,04; 7,23; 7,48; 7,61; 7,64; 7,8; 7,9; 8; 8,03; 8,1; 8,17; 8,27; 8,5; 8,66; 8,67; 8,8; 8,82; 8,89; 8,9; 8,9; 8,92; 8,94; 8,94; 8,96; 8,99; 9,13; 9,2; 9,2; 9,39; 9,5; 10; 10. Es bietet sich an, ein Intervall der Größe 0,5 zu wählen. Der Stamm läßt sich also sofort zeichnen:

```

7 |
7 |
8 |
8 |
9 |
9 |
10 |
10 |

```

D.h. der erste 7er steht für den Wertebereich 7,0-7,5, der zweite für 7,5 bis 8,0 usw. Nun werden wie in Punkt 2 erläutert die Daten auf eine Stelle nach der Stamm-Stelle gerundet also: 7,0; 7,0; 7,2; 7,5; 7,6; 7,6; 7,8; 7,9; 8,0; 8,0; 8,1; 8,2; 8,3; 8,5; 8,7; 8,7; 8,8; 8,8; 8,9; 8,9; 8,9; 8,9; 8,9; 8,9; 9,0; 9,0; 9,1; 9,2; 9,2; 9,4; 9,5; 10,0; 10,0. Im letzten Schritt trägt man jetzt die gerundete zweite Ziffer neben den passenden Stamm. Für die ersten drei Zahlen wäre das der erste Siebener, für die nächsten fünf

⁹Falls sehr kleine Werte vorliegen, kann man diese evt. neu skalieren, das muß man sich eben von Problem zu Problem neu überlegen. Hat man bspw. alle Werte im Bereich von 0,0045 bis 0,0068 so kann man eine Multiplikation mit dem Faktor 1000 in Betracht ziehen — das ändert ja an der Verteilung, an der wir interessiert sind, nichts.

Zahlen also von 7,5–7,9 ist das der zweite Siebener usf. Komplette sieht das Diagramm dann so aus:

```

7 | 002
7 | 56689
8 | 00123
8 | 57788999999
9 | 001224
9 | 5
10 | 00
10 |

```

Entscheidend für die Beurteilung ist einerseits die *Anzahl* der Ziffern, also die Länge der Blätter, andererseits kann man aus den Ziffern auch gewisse Rückschlüsse ziehen. Nehmen wir als Beispiel den zweiten „achter“ Ast heraus: Man erkennt sofort, daß es sich um die Gruppe mit den meisten Werten handelt (11) und, daß diese Gruppe den Wertebereich zwischen 8,5 und 9 abdeckt. Interessant ist bei näherer Betrachtung unter Umständen auch die Tatsache, daß mehr als die Hälfte (6) der Werte bei 8,9 liegen.

Vergleicht man dieses Diagramm mit dem Histogramm in Abb. 4.3, so erkennt man, daß beide sehr ähnliche aussehen. Auch hier erkennt man sofort, daß eine links-schiefe Verteilung¹⁰ vorliegt.

Dennoch sind die Unterschiede interessant, und wurden beispielhaft belassen. Das Histogramm wurde mit einem bekannten wiss. Graphikprogramm gezeichnet, wobei die Intervalle nicht rechts-halboffen sondern links-halboffen gewählt werden. Es ist interessant zu sehen, daß bereits solche scheinbar marginalen Unterschiede in den Diagrammen deutlich zu erkennen sind. Es empfiehlt sich, beim Analysieren von Histogrammen die Parameter wie Intervallgrößen und -lage zu variieren und die Unterschiede zu beachten.

4.5.4. Normalverteilung

Die Normalverteilung ist die für die Auswertung analytischer Meßdaten wichtigste Verteilung¹¹ (siehe Abb. 4.5). Die Achsen haben die bei Histogrammen übliche Bedeutung (x -Achse steht für Wertebereich, y -Achse für Anzahl der Werte im jeweiligen Intervall). Aus dieser Graphik kann man sofort herauslesen, daß sich alle Werte symmetrisch um den Wert, der mit μ beschriftet ist verteilen. Weiters ist die Häufigkeit am Punkt μ am größten und nimmt in beide Richtungen (also zu größeren und kleineren Werten) gleichmäßig ab¹².

Fläche unter der
Kurve

Eine besondere Bedeutung erhält die Fläche unter der Kurve. Diese kann man

¹⁰Man spricht von *linksschiefen* oder *rechtssteilen* Verteilungen, wenn die Verteilung nicht symmetrisch ist und der Anstieg der Verteilungskurve an der rechten Seite steiler ist, als an der linken — so wie in diesem Beispiel; das Gegenteil davon ist *rechtsschief* und *linkssteil*.

¹¹Eine wichtige Ausnahme sind analytische Verfahren, deren Ergebnis durch „Auszählen“ von Ereignissen zustande kommen. Als Beispiele könnte man die Röntgenspektroskopie nennen. In diesem Fall liegt nämlich eine Poissonverteilung vor.

¹²Diese Kurve wird aufgrund ihrer Form auch *Gaußsche Glockenkurve* genannt.

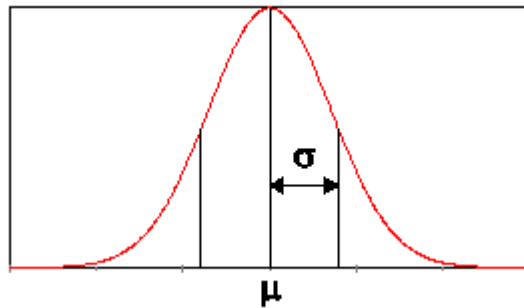


Abbildung 4.5.: Normalverteilung — eine kontinuierliche Verteilung — und der Zusammenhang zwischen der Verteilung, Mittelwert und Standardabweichung.

als *Wahrscheinlichkeit* auffassen. Die Gesamtfläche unter der Kurve ist gleich 1. Mit anderen Worten: Die Wahrscheinlichkeit einen Meßwert *irgendwo* unter der Kurve anzutreffen muß gleich 1, also 100 % sein¹³. Diese Zusammenhänge kann man am besten mathematisch formulieren:

Wenn X eine beliebige (zufällige) Meßgröße ist, so ist die Wahrscheinlichkeit (P), daß X zwischen den Werten a und b liegt durch die *Fläche unter der Kurve* zwischen diesen beiden Punkten zu verstehen. Oder als Gleichung geschrieben:

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad (4.1)$$

wobei $f(x)$ gleich der Funktion ist, die die Normalverteilung beschreibt.

Beispiel: Die Wahrscheinlichkeit, daß sich ein zufällig gewählter Wert in dem in Abb. 4.5 eingezeichneten Bereich $\mu \pm \sigma$ befindet, ist gleich der Fläche unter der Kurve zwischen $(\mu - \sigma)$ und $(\mu + \sigma)$:

$$\int_{\mu - \sigma}^{\mu + \sigma} f(x) dx \quad (4.2)$$

Man erhält dafür den Wert 0,6827 (= 68 %).

Die Normalverteilung erhält ihre für die analytische Praxis hohe Bedeutung v.a. aus der praktischen Erfahrung, daß sehr viele Meßergebnisse dieser Verteilung folgen. Wird z.B. die Konzentration einer organischen Komponente gravimetrisch bestimmt, so ist zu erwarten, daß (bei korrekten Laborbedingungen) die Verteilung mehrere aufeinanderfolgende Messungen eben der Normalverteilung folgt.

Statistisch wird dies auch durch den *zentralen Grenzwertsatz* ausgedrückt:

Beispiel

Zentraler
Grenzwertsatz

¹³Die Fläche unter der Kurve von Verteilungsfunktionen, egal welcher Form, ist normalerweise auf 1 normiert, mathematisch formuliert: $\int_{-\infty}^{\infty} f(x) = 1$

[...] eine Summe von vielen unabhängigen, beliebig verteilten Zufallsvariablen gleicher Größenordnung [ist] annähernd normalverteilt, und zwar umso besser angenähert, je größer ihre Anzahl ist.[...]

Unter einer Zufallsvariable versteht man — stark vereinfacht ausgedrückt — die Werte einer bestimmten Variable. Die Zufallsvariable X könnte also bspw. für die Temperatur stehen, d.h. alle möglichen Meßwerte des physikalischen Zustandes der als Temperatur verstanden wird beinhalten.

Für detailliertere Information siehe auch [21].

Anmerkung

Der Terminus „Normalverteilung“ darf nicht so mißinterpretiert werden, als wäre die Normalverteilung die in der Natur am häufigsten auftretende, sozusagen die „normale Verteilung“. Dies ist nicht der Fall. Davon abgesehen gilt die Normalverteilung auch aus der Definition des zentralen Grenzwertsatzes oftmals nur annähernd. Trotzdem ist sie eine in der *analytischen Praxis* häufig auftretende Verteilung.

4.6. Mittelwert

Der Mittelwert ist wohl einer der am meisten benutzten statistischen Parameter. Es ist kaum möglich eine Tageszeitung aufzuschlagen, ohne über gemittelte Werte zu stolpern. Die Berechnung ist auch denkbar einfach: Es werden alle Werte addiert und danach durch die Anzahl der Werte dividiert. Als Ergebnis ist ein mittlerer Wert erwartet:

$$\mu = \frac{\sum_{n=1}^N x_n}{N} \quad (4.3)$$

wobei μ für den Mittelwert steht, N ist die Anzahl der Meßwerte und x_1, x_2, \dots, x_N sind die einzelnen Meßwerte.

Sehr wichtig, aber weit weniger bekannt ist die Tatsache, daß der Mittelwert der *Erwartungswert*¹⁴ der Normalverteilung ist. Mit anderen Worten: der Mittelwert wird von der Normalverteilung abgeleitet und ist der zentrale Punkt, das Maximum der Gaußschen Glockenkurve (siehe Abb. 4.5).

Es ist vor allem dann wichtig, sich diese Tatsache vor Augen zu führen, wenn die Verteilung der Daten *nicht* der Normalverteilung folgt. Für diesen Fall muß genau geprüft werden, ob das Berechnen des Mittelwertes zum gewünschten Ergebnis führt, oder ob nicht bspw. ein robustes Maß wie der Median das geeignetere Maß wäre. Mehr dazu in den folgenden Abschnitten.

Grundgesamtheit,
Stichprobe: μ, \bar{x}

Weiters muß — wie in Abschnitt 4.4 auf Seite 30 erwähnt — zwischen Mittelwert einer *Grundgesamtheit* und einer *Stichprobe* unterschieden werden. Der Mittelwert

¹⁴Der Begriff *Erwartungswert* leitet sich aus der Wahrscheinlichkeitstheorie ab, und kann hier nicht näher erläutert werden. Im Rahmen der Mathematik Vorlesung wird hierauf aber hinreichend eingegangen werden, siehe auch [3,5].

einer Grundgesamtheit wird mit μ bezeichnet, der Mittelwert einer Stichprobe hingegen mit \bar{x} .

Diese Unterscheidung ist wichtig, zumal diese beiden Mittelwerte unterschiedlich sein können, und üblicherweise auch sind. Wobei sich \bar{x} umso mehr μ annähert, je größer die Stichprobe ist. Im Falle analytischer Meßwerte habe wir es praktisch ausschließlich mit Stichproben zu tun und verwenden daher \bar{x} .

Der Unterschied ist aber nicht etwa nur *formaler* Natur, sondern v.a. insofern von praktischer Bedeutung, als man beim Lesen statistischer Bücher, respektive beim **Verwenden statistischer Software** aufpassen muß, die richtige Formel bzw. das richtige Verfahren auszuwählen¹⁵!

Statistik
Software

Auch in der weiteren Folge wird diese Unterscheidung (sofern notwendig) sehr konsequent durchgeführt. Eine Verwendung des einen oder anderen Symbols ist also nicht willkürlich sondern von z.T. erheblicher Bedeutung!

4.7. Standardabweichung

Wir haben uns bereits in den vorangegangenen Abschnitten mit dem Problem der Streuung der Meßdaten auseinandergesetzt. In Abschnitt. 4.3 auf Seite 29 wurden verschiedenen Fehlerarten erklärt und mittels Histogramm oder Stamm-Blatt Diagramm schließlich die Verteilung (Streuung) der Daten betrachtet und erklärt.

Streuung

Bis jetzt haben wir allerdings erst einen statistischen Parameter kennengelernt, nämlich den Mittelwert. Nun ist der Mittelwert einer Meßserie *alleine* nicht immer sehr aussagekräftig. Schließlich macht es einen Unterschied, ob bei gleichen Mittelwerten die Streuung der Daten (also der zufällige Fehler) groß oder klein ist, aus dem Mittelwert läßt sich das ja nicht erkennen.

In diesem Abschnitt soll folglich ein weiterer sehr wichtiger statistischer Parameter eingeführt werden, der eben diese Streuung beschreibt, nämlich die *Standardabweichung*: Zunächst werden wir die Standardabweichung als formalen statistischen Parameter auffassen, der dann zunehmend Bedeutung für die Beurteilung des Ergebnisses einer Messung bekommen wird:

Die Standardabweichung ist für den Fall der Normalverteilung definiert als der *Abstand zwischen Mittelwert und Wendepunkt*¹⁶. Für die Standardabweichung der Grundgesamtheit wird das Symbol σ verwendet, für die Standardabweichung einer Stichprobe s . Auch hier gilt wieder: Die *Standardabweichung einer Stichprobe* nähert sich mit zunehmender Größe der Stichprobe der *Standardabweichung der Grundgesamtheit* an. Der Zusammenhang zwischen Normalverteilung und Standardabweichung ist in Abb. 4.5 auf Seite 37 ersichtlich.

Definition

¹⁵Erwähnenswert ist an dieser Stelle vielleicht, daß manche ältere Software wie bspw. DOS Versionen von Lotus 1-2-3 überhaupt nur die Standardabweichung der *Grundgesamtheit* als Formel anbietet. Dies ist insofern noch von Bedeutung, als manche Palmtop Computer noch mit derartiger Software arbeiten.

¹⁶Im Wendepunkt einer Funktion $f(x)$ ist die *zweite Ableitung* dieser Funktion gleich 0, also gilt $(d^2f(x)/dx^2) = 0$. Anschaulich kann man den Wendepunkt als den Punkt definieren, an dem sich die Kurve von der konvexen zu konkaven Form ändert (oder umgekehrt).

Die Standardabweichung errechnet man mit folgenden Formeln¹⁷ :

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \quad (4.4)$$

für den Fall, daß **Stichproben** vorliegen und

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{n}} \quad (4.5)$$

falls eine **Grundgesamtheit** Basis der Statistik ist. Wobei nochmals bemerkt werden soll, daß für die analytische Praxis eher nur die erste Formel von Bedeutung ist.

Wahrscheinlichkeiten
Aus den Eigenschaften der Normalverteilung ergeben sich weiters folgende Eigenschaften:

In $\mu \pm 1\sigma$ liegen 68,27 %,
in $\mu \pm 2\sigma$ 95,45 % und
in $\mu \pm 3\sigma$ 99,73 % aller Meßwerte.

oder anders gerechnet:

In $\mu \pm 1,96\sigma$ liegen 95 %,
in $\mu \pm 2,58\sigma$ 99 % und
in $\mu \pm 3,29\sigma$ 99,9 % aller Meßwerte.

Reproduzierbarkeit Nehmen wir also an, eine Meßserie folgt der Normalverteilung. In diesem Fall würde eine „breite“ Glockenkurve eine relativ *große Streuung* (großes σ) der einzelnen Werte bedeuten, eine „schmale“ Glockenkurve hingegen eine *geringe Streuung* (kleines σ) der Meßwert. Diese Streuung wurde in Abschnitt 4.3 auf Seite 29 auch als *Reproduzierbarkeit* bezeichnet. Wie schon erwähnt, ist für die Angabe eines Meßergebnisses nicht nur der *Mittelwert* der Meßserie, sondern auch die *Reproduzierbarkeit* dieses Wertes von Bedeutung. Dies soll an einem Beispiel erläutert werden:

Beispiel Ein Analytiker mißt die Konzentration einer Verunreinigung einer industriellen Probe. Der Mittelwert aus fünf Bestimmungen der Konzentration ist 4,53 ppm. Die Firma garantiert für ihr Produkt eine maximale Verunreinigung durch diese Substanz von 5 ppm. Der Mittelwert wäre mit 4,53 ppm noch deutlich *unter* dieser Grenze. Die Frage ist jedoch, mit welcher Reproduzierbarkeit die Messung erfolgt ist. Da die fünf Messungen ja nur eine Stichprobe darstellen, nähert sich der Mittelwert folglich nur mit einer gewissen Genauigkeit an den wahren Wert an. Bei großer Unsicherheit der Messung wäre es möglich, daß der wahre Wert trotzdem **über dem Grenzwert** zu liegen kommt!

¹⁷Der Ausdruck $(n - 1)$ wird auch als *Anzahl der Freiheitsgrade* bezeichnet. Dies kommt von folgender Idee: Nachdem der Mittelwert berechnet wurde, kann man von allen vorhandenen Einzelwerten n nur noch $(n - 1)$ „frei wählen“ — daher die Bezeichnung *Freiheitsgrad*.

Angenommen, der Fehler der Messung wird mit einer Irrtumswahrscheinlichkeit von $\alpha = 0,1$ berechnet und ergibt, daß mit einer Wahrscheinlichkeit von 90 % der wahre Wert im Bereich zwischen $4,53-0,4$ ppm und $4,53+0,4$ ppm zu liegen kommt. Andererseits bedeutet dies aber auch, daß immerhin mit einer Wahrscheinlichkeit von 10 % der wahre Wert *nicht* in diesem Intervall liegt. Also könnte der wahre Wert durchaus auch über 5 ppm liegen!

Aus diesem Beispiel ist ersichtlich, daß man es sich mit der Angabe solcher Ergebnisse und v.a. aus den Schlüssen, die man daraus zieht, nicht zu leicht machen darf. Der nächste Schritt wäre nun, die Angabe so zu Ergänzen, daß nicht mehr alleine der Mittelwert als Ergebnis angegeben wird, sondern der Mittelwert mit Angabe der Genauigkeit, also etwa in der Form:

$$\text{Konzentration der Substanz } x \text{ in Probe } y = \text{Mittelwert} \pm \text{Fehler} \quad (4.6)$$

Aus dem bisher gelernten *könnte* man nun den voreiligen Schluß ziehen das Ergebnis $\bar{x} \pm s$? in folgender Form anzugeben:

$$\text{Konzentration der Substanz } x \text{ in Probe } y = \bar{x} \pm s \quad (4.7)$$

Diese Angabe ist aus verschiedenen Gründen **falsch**:

Da aus einer Stichprobe ermittelt, sind \bar{x} und s ja nur *Schätzwerte* für die wahren Werte μ und σ . **Beide** Schätzwerte sind nun aber mit einem Fehler behaftet, der abhängig von der *Größe der Stichprobe* ist. Dieser Fehler geht aber in dieser Angabe überhaupt nicht ein!

Weiters muß man sich die Frage stellen, was der „Leser“ erwartet, wenn er als Ergebnis einer Analyse ein solches Intervall als Resultat vorfindet: Man kann annehmen, daß er es für die Angabe der *Zuverlässigkeit* des Ergebnisses hält und schließt, daß der wahre Wert — der ja aus der Stichprobe nicht direkt ermittelt werden kann — mit einer hohen Wahrscheinlichkeit in diesem Interfall liegt.

Genau das ist aber nicht der Fall. Tatsächlich finden sich in einem $\pm 1s$ Intervall um den Mittelwert nur etwa 68 % der Meßwerte der Stichprobe. Dies sagt noch nichts darüber aus, mit welcher Wahrscheinlichkeit sich der wahre Wert in diesem Intervall befindet¹⁸.

Mit anderen Worten: **Die obige Angabe (selbst wenn wir das Intervall auf $\bar{x} \pm 2s$ oder $\bar{x} \pm 3s$ vergrößern, gibt nur die Streuung der Meßdaten an, nicht aber, mit welcher Wahrscheinlichkeit sich der wahre Wert innerhalb dieses Intervalls befindet!**

Gerade dies wäre aber die interessante Aussage. Möchte man nur die Streuung der Daten angeben, so steht es ja frei die Zahlenwerte für Mittelwert und Standardabweichung anzugeben, nur die Angabe in dieser Form ist verwirrend und aus erwähnten Gründen falsch. Die Wahrscheinlichkeit den wahren Wert innerhalb dieses Intervalles zu finden ist bei relativ kleinen Stichproben, wie sie gerade in der Chemie oft vorkommen, noch deutlich geringer.

Fazit: Eine Angabe der Form $\bar{x} \pm s$ ist bei Meßwerten irreführend und **nicht** Fazit

¹⁸Davon abgesehen finden sich in $\bar{x} \pm s$ nur etwa 68 % der Meßwerte, anders gesagt: fast ein Drittel der Meßwerte befindet sich *außerhalb* dieses Bereiches!

zulässig. Ein anderes Maß für die Güte einer Messung muß gefunden werden, die auch den zusätzlichen Fehler berücksichtigt der daraus resultiert, daß nur (kleine) Stichproben vorliegen. Dieses Maß ist der **Vertrauensbereich** der in Abschnitt 4.10 beschrieben wird.

4.8. Variationskoeffizient

Ein weiteres, sich aus der Standardabweichung ableitendes Maß soll hier noch kurz erwähnt werden: zur Beschreibung der Güte eines Verfahrens wird oft ein relatives Maß für die Streuung bevorzugt, der *Variationskoeffizient*:

$$V = \frac{s}{\bar{x}} \quad (4.8)$$

Der Variationskoeffizient hat verglichen mit der Standardabweichung den Vorteil, daß er in Relation zum Mittelwert steht. Alleine aus der Standardabweichung hingegen kann man keine großen Schlüsse ziehen. Aus diesem Grund wird in manchen Fällen der Variationskoeffizient bevorzugt.

4.9. Varianz

Die Varianz ist eine mit der Standardabweichung vergleichbare Größe. Sie wird manchmal alternativ zur Standardabweichung verwendet. Die Varianz ist das Quadrat der Standardabweichung:

$$v = s^2 = \frac{\sum(x - \bar{x})^2}{n - 1} \quad (4.9)$$

Anmerkungen Bemerkt sollte noch werden, daß die Varianz etwa doppelt so genau anzugeben ist wie die Standardabweichung s . z.B.: $s = 1,3$ und $s^2 = v = 1,69$, da bei der Rückrechnung von s^2 zu s durch das Wurzelziehen sonst Genauigkeit verloren werden würde. Außerdem ist die Dimension der Varianz gleich dem Quadrat der Einheit: Wird beispielsweise Temperatur gemessen und die Einheit *Kelvin* K verwendet, so hat der *Mittelwert* und die *Standardabweichung* die Einheit K, die Varianz hingegen K^2 .

4.10. Vertrauensbereich

Unter dem Begriff *Vertrauensbereich* verstehen wir die Angabe eines Intervalls in dem sich der wahre Wert mit einer *bestimmten Wahrscheinlichkeit* befindet. Mathematisch ausgedrückt: $x_{links} \leq x \leq x_{rechts}$. Als Vertrauenswahrscheinlichkeit wird oft 0,95 oder 0,99 verwendet, was bedeutet, daß sich der wahre Wert im ersten Fall mit einer Wahrscheinlichkeit von 95 % im anderen Fall mit einer Wahrscheinlichkeit von 99 % zwischen x_{links} und x_{rechts} befindet.

Wie wir im Abschnitt 4.7 über die Standardabweichung schon festgestellt haben, ist die Angabe eines Vertrauensbereiches von verschiedenen Parametern abhängig

und darf nicht mit der Standardabweichung der Stichprobe verwechselt werde. Die Standardabweichung ist ja „nur“ als Maß für die Streuung der Meßdaten zu verstehen. Der Vertrauensbereich hingegen spannt ein Intervall auf, das mit hoher Wahrscheinlichkeit den wahren Wert einschließt. Dies ist ein feiner, aber wichtiger Unterschied. In der Praxis müssen zwei Fälle unterschieden werden:

σ bekannt?

1. Der Mittelwert wird aus den gemessenen Proben errechnet, die Standardabweichung der Grundgesamtheit ist aber bekannt.
2. Der Mittelwert **und** die Standardabweichung werden aus den gemessenen Proben errechnet.

Nun wird man sich vermutlich fragen, wie es im **Fall 1** sein kann, daß die Standardabweichung der Grundgesamtheit bekannt ist, zumal weiter oben festgestellt wurde, daß die Grundgesamtheit unendlich groß ist, und somit einer Messung nicht zugänglich. Die Lösung ist folgende: Selbstverständlich hat sich an diesen Prinzipien nichts geändert, die Standardabweichung der Grundgesamtheit kann natürlich nicht präzise bestimmt werden. Es kann jedoch vorkommen, daß eine bestimmte analytische Routinemethode in einem bestimmten Labor von einem bestimmten Laboranten über lange Zeit in der gleichen Art und Weise durchgeführt wird. Aus der großen Anzahl der gemachten Messungen, kann man nach einer gewissen Zeit die „Standardabweichung der Grundgesamtheit“ mit hinreichender Genauigkeit abschätzen. Daher kann man eben diese geschätzte Standardabweichung σ verwenden, anstatt für jede Meßserie eine neue Standardabweichung s zu errechnen.

Standardabweichung der Grundgesamtheit?

Jedoch sollte man sich im Falle einer solchen Vorgangsweise im klaren darüber sein, daß der Vertrauensbereich dann kleiner ist, da ja u.a. $\sigma \leq s$ gilt. **Diese Methodik sollte also nur unter klar definierten Randbedingungen, die man garantieren kann, angewandt werden.** Es handelt sich also eher nicht um den Regelfall sondern um die Ausnahme. Schon zwei verschiedene Analytiker werden an *demselben* Gerät mit *derselben* Methode unter Umständen unterschiedliche Reproduzierbarkeiten erreichen.

Normalerweise wird also eher **Fall 2** auftreten: Es wird die aus der Meßserie errechnete Standardabweichung der Stichprobe s zur Bestimmung des Vertrauensbereiches verwendet.

Standardabweichung der Stichprobe

Der zweiseitige Vertrauensbereich¹⁹ errechnet sich nach folgenden Formeln:

Fall 1: Standardabweichung der Grundgesamtheit σ bekannt	$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$
Fall 2: Standardabweichung der Stichprobe s berechnet	$\bar{x} \pm t \frac{s}{\sqrt{n}}$

¹⁹Es soll an dieser Stelle nicht auf die Unterschiede zwischen *einseitigen* und *zweiseitig* begrenzten Vertrauensbereichen eingegangen werden. In den hier erwähnten praktischen Beispielen wird der zweiseitige Vertrauensbereich angewendet (siehe auch [21]).

mit n als Anzahl der Meßwerte²⁰, $z = 1,96$ (95 %), $z = 2,58$ (99 %) und $z = 3,29$ (99,9 %), t ergibt sich aus der sogenannten *Student-Verteilung* und ist tabelliert (siehe Tab. A.1). In der Tabelle sind für drei Wahrscheinlichkeiten: 90 %, 95 % und 99 % die t Werte in Abhängigkeit der Freiheitsgrade (FG) tabelliert. Der Freiheitsgrad ist die um eins verminderte Anzahl der Meßwerte. Bei 5 Meßwerten liegen also 4 Freiheitsgrade vor.

Aus dem bisher gesagten ergibt sich eine klare Konsequenz:

Scharfe Aussagen sind unsicher, sichere Aussagen sind unscharf.

Schärfe des Intervalls Mit anderen Worten: eine scharfe Aussage, also eine mit niedriger gewählter Wahrscheinlichkeit und daher kleinerem Vertrauensbereich birgt ein höheres Risiko, daß der wahre Wert nicht getroffen wird, und sind daher eher unsicher. Vertrauensbereiche, die mit hoher gewählter Wahrscheinlichkeit bestimmt werden sind sicher, aber aufgrund des größeren Intervalls entsprechend unschärfer.

Die Entscheidung, welche Sicherheit gewählt wird, ist klarerweise problemabhängig, und hängt sicher mit den Konsequenzen zusammen, die ein falscher Wert haben kann.

4.11. Nachweis- und Erfassungsgrenze

Für den analytischen Chemiker wichtige Größen sind weiters die sogenannten *Nachweis-* und die *Erfassungsgrenze*. Für diese Größen finden sich in der Literatur manchmal leicht unterschiedlichen Definitionen. An dieser Stelle wird die Definition vorgestellt, die sich aus der Standardabweichung des Blindwertes ableitet, und somit relativ leicht quantifizierbar ist (siehe auch [17]).

absolute und relative Präzision Man unterscheidet zwischen *relativer* und *absoluter* Präzision eines Verfahrens [15]. Die *absolute Präzision* gibt den absoluten Fehler einer Messung in einem bestimmten Konzentrationsbereich an. Dieser kann in manchen Fällen über weite Bereiche konstant sein (bspw. der Tropfenfehler einer Bürette oder auch Wägefehler). Die *relative Präzision* ist eine prozentuelle Angabe, die sich aus Meßwert und absolutem Fehler ergibt, also: $s/\bar{x} * 100$. Bei konstantem absolutem Fehler ist natürlich der *relative Fehler* bei kleinen Konzentrationen größer, als bei hohen Konzentrationen!

Die *Erfassungs-* und *Nachweisgrenze* sind zwei wichtige Größen zur Beschreibung eines analytischen Prozesses. Diese Größen definieren sich wie folgt:

Blindwert Jedes Verfahren verfügt über einen *Blindwert*. Dieser ist das Ergebnis analytischer Messung ohne Vorhandensein der zu messenden Substanz. Sucht man z.B. den Blindwert der spektroskopischen Bestimmung von Blei, so stellt man eine Lösung her, die eben kein Blei enthält, und unterwirft diese der chemischen Probenaufbereitung und der spektroskopischen Messung. Das gemessene Signal wird *Blindwert*

²⁰Vorausgesetzt ist eine unendlich große Grundgesamtheit (oder eine endlich große mit Zurücklegen der Stichprobe). Die Division durch \sqrt{n} erklärt sich wie folgt: Für $\lim_{n \rightarrow \infty} 1/\sqrt{n} = 0$. Folglich ist der Vertrauensbereich bei unendlich großer Stichprobe auch 0. Mit anderen Worten, mit wachsendem n nähert sich der Mittelwert der Stichprobe \bar{x} dem Mittelwert der Grundgesamtheit μ immer näher an und der Vertrauensbereich geht gegen Null; wenn $n = \infty$ gilt dann $\bar{x} = \mu$. Falls die Grundgesamtheit nicht unendlich groß ist, darf diese Formel nicht verwendet werden!

genannt. Dieser Blindwert stellt folglich eine untere Schranke des Messbereiches dar. Führt man die genannte Bestimmung des Blindwertes mehrmals durch, so kann man mit den bereits bekannten Verfahren die *Standardabweichung des Blindwertes* er rechnen.

Unter der Voraussetzung, daß die Streuung des Blindwertes einer Normalverteilung folgt (was meist der Fall ist), so definiert man die *Nachweisgrenze*²¹ des betrachteten Verfahrens als *mittlerer Blindwert plus dreimal der Standardabweichung des Blindwertes*. Die *Erfassungsgrenze* ist als *mittlerer Blindwert plus sechsmal der Standardabweichung des Blindwertes* definiert. Ein Meßwert gilt also erst dann als sicher, wenn er die Erfassungsgrenze überschreitet. Bezeichnen wir den mittleren Blindwert als \bar{x}_B , die Standardabweichung des Blindwertes mit s_B , die Erfassungsgrenze EG und die Nachweisgrenze als NG so folgt:

$$NG = \bar{x}_B + 3 \cdot s_B \quad (4.10)$$

und

$$EG = \bar{x}_B + 6 \cdot s_B \quad (4.11)$$

Nachweis- und
Erfassungsgrenze

4.12. Abweichung der Meßwerte von der Normalverteilung

Auf Basis der Erkenntnis, daß die Ergebnisse einer Messung der Normalverteilung entsprechen, können die daraus abgeleiteten statistischen Parameter Mittelwert, Standardabweichung, Varianz und Vertrauensbereich berechnet und angewandt werden.

Falls die Ergebnisse einer Messung aber *nicht* (angenähert) der Normalverteilung folgen ist die Verwendung der erwähnten statistischen Parameter zumindest fragwürdig²². Es muß also vor Angabe eines Ergebnisses durch bspw. Mittelwert und Vertrauensbereich unbedingt geprüft werden, ob diese Voraussetzungen gegeben sind. Zwei wichtige Probleme und die Vorgangsweise zur Lösung solcher Aufgaben sind:

1. Die Daten sind normalverteilt, aber es liegen ein oder mehrere *Ausreißer* vor
Unter Ausreißern versteht man solche Werte, die sich signifikant von den anderen Werten einer Meßreihe unterscheiden, und aus diesem Grund als Meßfehler klassifiziert werden.

²¹Die Nachweisgrenze wird auch als obere Rausch- oder Störgrenze bezeichnet.

²²Oftmals sind zu wenig Messungen vorhanden, um aufgrund dieser wenigen Daten feststellen zu können, ob diese der Normalverteilung folgen. Werden bsp. vier photometrische Messungen durchgeführt ist aus diesen wenigen Werten eine solche Entscheidung nicht möglich. Nun ist es auch nicht erforderlich für *jede Stichprobe* diesen Nachweis zu erbringen. In Fällen, wo bspw. aus langjähriger Erfahrung klar ist, daß die Daten (bei korrekter Durchführung des Experiments) normalverteilt sind, kann dies also als gegeben angenommen werden. Hat man es allerdings mit Daten einer Messung zu tun, wo dies *nicht* durch längere Praxis oder andere Untersuchungen klar ist, so sollte man dies vorher klären.

2. Die Daten folgen nicht einer angenäherten Normalverteilung, sondern z.B. einer links- oder rechtsschiefen Verteilung.

Vorgangsweise

Für den ersten Fall sollte ein Ausreißertest wie in Abschnitt 5 dargelegt durchgeführt werden und gegebenenfalls die entsprechenden Werte entfernt werden. Für den zweiten Fall sollte man sich im klaren darüber sein, daß die Verwendung von statistischen Parametern, die von der Normalverteilung abgeleitet sind zu Fehlinterpretationen führen kann. Der Mittelwert einer solchen Messung kann also recht wenig aussagekräftig sein. Für diesen Fall sollte die Anwendung robuster Meßgrößen wie die des *Medians* in Erwägung gezogen werden. Eine kurze Einführung in die robuste Statistik wird in Abschnitt 4.13 gegeben.

Betrachten wir Beispiel 4 aus der Einleitung auf Seite 29, hier hatten wir mit Nitrat-Meßdaten aus 33 Trinkwasser-Brunnen zu tun. Für diese Daten wurde ein Histogramm gezeichnet (Abb. 4.3 auf Seite 33), aus dem man sofort herauslesen kann, daß keine Normalverteilung sondern vielmehr eine *schiefe* Verteilung vorliegt mit der höchsten „Dichte“ bei 8–9. Errechnet man nun den „einfachen“ Mittelwert, so ergibt sich: $\bar{x} = 8,53$. Merken wir uns diesen Wert und gehen wir über zum nächsten Kapitel den *robusten Größen*.

4.13. Robuste Größen: Median, Quartile

4.13.1. Median

Unter robuster Statistik versteht man Methoden, die wesentlich weniger anfällig auf Ausreißer und Abweichungen der Daten von der Normalverteilung sind. Der Median ist — wie der arithmetische Mittelwert, den wir schon berechnet haben — ein durchschnittlicher, mittlerer Wert der sich wie folgt bestimmen läßt:

1. Alle Werte werden der Größe nach sortiert.
2. Der Wert, der in der *Mitte* der sortierten Reihe steht ist der Median.
3. Für den Fall, daß die Anzahl der Meßwerte *gerade* ist, gibt es zwei Werte (a , b), die in der Mitte stehen. In diesem Fall ist der Median²³ $\frac{(a+b)}{2}$.

Beispiele Beispiel: Wie lautet der Median der Meßwerte: 3,6; 4,5; 2,7; 3,0; 3,3; 3,4;

1. **Sortieren:** 2,7; 3,0; 3,3; 3,4; 3,6; 4,5;
2. **Den oder die mittleren Wert(e) suchen:** 2,7; 3,0; **3,3; 3,4**; 3,6; 4,5;
3. **Zwei mittlere Werte liegen vor:** Median = $(3,3 + 3,4)/2 = 3,35$

Beispiel: Berechne den Median der Werte: 4,4; 5,1; 4,1; 6,2; 5,7; 5,6; 7,0. Zunächst die Werte sortieren: 4,1; 4,4; 5,1; **5,6**; 5,7; 6,2; 7,0. Der mittlere Wert, also der Median ist also 5,6.

Vergleich: Median-Mittelwert

Bestimmen wir nun den Median aus Beispiel 4, so ergibt sich ein Wert von **8,8**.

²³Dieser Wert wird manchmal auch *Pseudomedian* genannt, da es sich ja bei diesem Wert um keinen tatsächlich vorkommenden Meßwert handelt.

Verglichen mit dem Mittelwert bei **8,53** stellen wir fest, daß der Median näher beim Dichtemaximum (auch Modalwert genannt) liegt, also dort, wo wir einen mittleren Wert auch vermuten würden.

Ein anderes Beispiel wo das Anwenden des Mittelwertes zu problematischen Werten führen kann, ist die Berechnung des *Durchschnittseinkommens* einer Bevölkerung. Dieses ist nicht normal-, sondern vielmehr schief verteilt (Es gibt wesentlich mehr Personen die wenig verdienen als solche, die sehr viel verdienen.) Berechnet man nun den *Mittelwert* dieser Daten, so kommt der Mittelwert bei zu hohen Summen zu liegen. Auch hier ist der *Median* das geeignetere Maß.

Schon in diesen Beispielen läßt sich sehen, daß der **Median wesentlich weniger anfällig auf schiefe Verteilungen oder Ausreißer in den Meßdaten ist**. Der Mittelwert wird unter Umständen schon von recht wenigen Ausreißern oder eben nicht normalverteilten Daten stark beeinflusst. Dies ist der Grund, warum bei der Erklärung des Mittelwertes und der Standardabweichung so großer Wert auf den Zusammenhang mit der Normalverteilung gelegt wurde.

Für den Fall, daß genug Meßwerte vorliegen (mit 4 Meßwerten läßt sich kaum die Verteilung abschätzen), sollte also einerseits immer ein Histogramm gezeichnet werden und sicherheitshalber sowohl Mittelwert als auch Median berechnet werden. Erkennt man aus dem Histogramm deutliche Abweichungen von der Normalverteilung bzw. unterscheiden sich Mittelwert und Median stark voneinander (wie in unserem Beispiel), so sollte man eher den Median als den Mittelwert angeben (evt. auch beide, um die Problematik zu demonstrieren).

Jedenfall sollte man sich Gedanken über den verwendeten statistischen Parameter machen und nicht aus Routine oder Bequemlichkeit (mein Taschenrechner kann keinen Median ausrechnen ...) **automatisch den Mittelwert wählen**. Die Auswahl der *statistischen Methodik* sollte eine ebenso **rationale** sein, wie die Wahl der *analytischen Meßmethodik*.

Vorgangsweise in der Praxis

4.13.2. p-te Perzentile, Quartile, Interquartiler Abstand

Verwendet man den Median, so kann man natürlich nicht die Standardabweichung zur Angabe von Reproduzierbarkeit verwenden. Diese ist an den Mittelwert gebunden. Um die Streuung der Meßwerte um den *Median* zu beschreiben, gibt man die Quartile oder den interquartilen Abstand an. Um den Begriff Quartil zu verstehen, sollte zunächst der Terminus *Perzentil* geklärt werden: **Derjenige Wert, unter dem p % der Meßwerte liegen wird p-tes Perzentil genannt**. Wäre bspw. 23,5 das Perzentil mit p=64, dann liegen 64 % der Meßwerte unter 23,5.

Perzentil

Die Quartile sind die Perzentile mit p=25, 50 und 75. Das zweite Quartil ist folglich der Median.

Quartile

Der interquartile Abstand ist der Wert zwischen *erstem* und *drittem* Quartil und kann somit als Parameter für die Streuung der Meßwerte verstanden werden.

Interquartiler Abstand

4.14. Signifikante Stellen

Ein weiterer wichtiger Punkt, der oft nicht beachtet wird, ist das Angeben des Ergebnisses mit der korrekten Anzahl an Dezimalstellen — eben den signifikanten Dezi-

	Wert	U_{abs}	$U_{rel} = U_{abs}/\text{Wert}$
	$3,2 \cdot 10^{-2}$	10^{-3}	$10^{-3}/3,2 \cdot 10^{-2} = \mathbf{3,1\%}$
	$4,10 \cdot 10^{-3}$	10^{-5}	$10^{-5}/4,10 \cdot 10^{-3} = \mathbf{0,44\%}$
	18,02	10^{-2}	$10^{-2}/18,02 = \mathbf{0,055\%}$
Produkt	$2,364 \cdot 10^{-3}$	10^{-6}	$10^{-6}/2,364 \cdot 10^{-3} = \mathbf{0,042\%}$
	$2,4 \cdot 10^{-3}$	10^{-4}	$10^{-4}/2,4 \cdot 10^{-3} = \mathbf{4,1\%}$

Tabelle 4.1.: Beispiel: Signifikante Stellen nach einer Multiplikation.

malstellen [17]. Nur weil der Taschenrechner 16 Kommastellen „ausspuckt“, macht es in den allerwenigsten Fällen Sinn, alle anzugeben:

Diejenigen Dezimalstellen gelten als signifikant, die mit Sicherheit bekannt sind, plus der ersten unsicheren Stelle.

Beispielsweise ist die Mengenangabe einer Auswaage von 13,443 g gleichbedeutend mit $13,443 \pm 0,001$ g.

Addition,
Subtraktion

Daraus folgt unmittelbar, daß bei Anwendung der Addition oder Subtraktion²⁴ die Genauigkeit des Ergebnisses maximal so hoch sein kann, wie die Ungenauigkeit des Parameters mit den wenigsten signifikanten Stellen. (Eine Kette ist nur so stark wie ihr schwächstes Glied.)

Beispiel: $3,233 + 54 = 57$ und **nicht** $57,233!$

Multiplikation,
Division

Für den Fall der Multiplikation oder Division muß man anders vorgehen. Hier muß man sich auf sogenannte *relative Unbestimmtheiten* ($U_{rel} = U_{abs}/\text{Wert}$) beziehen. Diese müssen für alle Werte bestimmt werden. Das Ergebnis ist dann in der Größenordnung der größten *relativen* Unbestimmtheit anzusetzen. Wie das im Detail funktioniert wird an einem Beispiel in Tab. 4.1 demonstriert.

Beispiel

In diesem Beispiel liegen drei Meßwerte vor, die multipliziert werden. Der Wert U_{abs} gibt die *absolute Unbestimmtheit* an (also die letzte signifikante Stelle). Dann werden die *relativen Unbestimmtheiten* U_{rel} errechnet. Die letzten beiden Zeilen stellen das Ergebnis der Multiplikation exemplarisch in zwei Genauigkeiten dar. Für diese beiden Angaben werden dann wieder die relativen Unbestimmtheiten berechnet. Im konkreten Fall sieht man, daß das gerundete Ergebnis $2,4 \cdot 10^{-3}$ mit 4,1 % relativer Ungenauigkeit im Bereich der größten Ungenauigkeit der Messungen mit 3,1 % liegt. Folglich ist diese Rundung als korrekt anzusehen. Eine Angabe größerer „Präzision“ wäre nicht mehr signifikant, also irreführend.

²⁴Ein Spezialfall und eine besonders kritische Sache ist die Subtraktion einander sehr ähnlicher Größen, da gegebenenfalls **alle** ursprünglich signifikanten Stellen „verschwinden“, und im Ergebnis sogar die höchstwertige signifikante Stelle falsch sein kann. Beispiel: $3,221 - 3,218$, beide Zahlen mit Signifikanz $\pm 0,005$: Das numerische Ergebnis wäre: $0,003$; da diese Stelle jedoch mit der Unsicherheit $\pm 0,005$ behaftet ist, ergibt sich, daß im Ergebnis keine einzige Stelle mehr signifikant ist!

4.15. Korrekte Angabe der Ergebnisse der Beispiele

4.15.1. Beispiel 1: „Bromid“

In diesem Fall kann man (muß man) annehmen, daß die Grundgesamtheit normalverteilt ist — jedenfalls kann man dies aus drei Bestimmungen nicht herauslesen — und die Werte sehen auch recht plausibel aus: Also wird zunächst der Mittelwert errechnet:

$$\bar{x} = \frac{5,31 + 5,08 + 5,44}{3} = \underline{\underline{5,28}} \quad (4.12)$$

Dann errechnen wir die Standardabweichung und verwenden dafür die Formel für die **Stichprobe**:

$$s = \sqrt{\frac{(5,31 - 5,28)^2 + (5,08 - 5,28)^2 + (5,44 - 5,28)^2}{3 - 1}} = \underline{\underline{0,1823}} \quad (4.13)$$

An dieser Angabe könnte man einwenden, die vier Stellen, mit der die Standardabweichung angegeben ist, sind nicht signifikant. Dies ist an sich richtig, nur handelt es sich bei dieser Angabe noch nicht um das Endergebnis. Da wir mit diesem Wert noch weiterrechnen, ist es günstiger (sicherheitshalber) noch mit höherer Genauigkeit zu arbeiten²⁵. Beim Endergebnis wird dann selbstverständlich genau auf die Anzahl der signifikanten Stellen geachtet werden.

4 signifikante Stellen?

Als nächstes wird der Vertrauensbereich berechnet. Da die Standardabweichung der Grundgesamtheit nicht bekannt ist, verwendet wird die Formel, die mit der Standardabweichung der Stichprobe arbeitet. Für diese Formel ist es nötig, den Student t Wert aus der Tabelle A.1 auszulesen. Für die Tabelle sind zwei Angaben erforderlich: die Anzahl der Freiheitsgrade, also Anzahl der Werte weniger eins — in unserem Fall also 2 — und die gewünschte Präzision. Wir nehmen an, daß eine Irrtumswahrscheinlichkeit von $\alpha = 0,05$ akzeptabel ist. Sieht man in der Tabelle nach, so findet man den Wert: 4,303. Also errechnet sich der Vertrauensbereich nach:

$$\bar{x} \pm 4,303 \frac{0,1823}{\sqrt{3}} = 5,28 \pm 0,45 \quad (4.14)$$

oder anders ausgedrückt: mit einer Wahrscheinlichkeit von 95 % liegt der wahre Wert im Intervall: $4,83 \leq x \leq 5,73$. Würde man die Wahrscheinlichkeit auf 99 % erhöhen, so ergäbe sich ein Intervall von $4,24 \leq x \leq 6,32$. Bei Wahl dieses Intervalles, läge man nur mehr in 1 % der Fälle falsch.

Ein korrektes Ergebnis könnte also so angegeben werden:

²⁵Dies kann man gerne als Grundsatz auffassen. Während der Rechnung sollte man die Zwischenergebnisse ruhig mit höherer Genauigkeit angeben. Schließlich ist es nicht wünschenswert, wenn das Endergebnis durch ungünstiges Runden während der Rechnung unnötig ungenau wird. Das Endergebnis muß dann natürlich als solches gekennzeichnet werden und entsprechend der Signifikanz der Stellen angegeben werden, hier wäre eine zu hohe Genauigkeit ja — im Gegensatz zu Zwischenergebnissen bei der Rechnung — irreführend

Die Konzentration von Brom in wässriger Lösung wurde gravimetrisch durch Fällung mit $AgNO_3$ bestimmt und beträgt $5,28 \pm 0,45 \text{ mg} \cdot \text{L}^{-1}$ (bei $P = 95 \%$).

4.15.2. Beispiel 2: „Blei in Wasser“

Messung hat einen Gang Dieses Beispiel ist gewissermaßen eine „Fangfrage“ und nimmt eine Sonderstellung ein: Betrachtet man die Werte (und nimmt man an, sie stehen in der Reihenfolge, in der sie gemessen wurden) so stellt man sofort eines fest: Sie steigen kontinuierlich an. Man sagt dazu auch: *Sie haben einen Gang*.

Es ist nun recht unwahrscheinlich, daß bei immerhin sieben Werten zufällig diese kontinuierliche Zunahme eintritt. Ein solches Ergebnis ist oft ein Indiz dafür, daß bei der Messung irgendetwas schief gegangen ist: Beispiele (nicht nur für diesen konkreten Fall) könnten sein: die Temperatur ist nicht konstant geblieben, eine Lösung war nicht homogen, der Detektor driftet, ...

Man sollte in einem solchem Fall unbedingt die Ursache suchen und nicht einfach den Mittelwert berechnen und das Ergebnis angeben. Der resultierende Fehler kann sonst wesentlich größer sein, als man vermutet.

4.15.3. Beispiel 3: „GC-ECD“

Dieses Beispiel müssen wir an dieser Stelle erst mal überspringen. Aufgrund der Tatsache, daß ein Wert scheinbar von den anderen deutlich abweicht, erfordert weitere Untersuchungen. Weder ein einfaches miteinbeziehen des Wertes ist ratsam, noch ein unfundiertes „unter den Tisch fallen lassen“. Konkret sollte man mittels eines Ausreißertests überprüfen, ob das Abweichen signifikant ist, oder nicht. Daher wird dieses Beispiel an Abschnitt 5 weitergereicht.

4.15.4. Beispiel 4: „Nitrat im Trinkwasser“

Dieses Beispiel wurde schon zu einem großen Teil in den vorigen Abschnitten besprochen. Fassen wir die Ergebnisse zusammen:

1. Es liegen genug Werte vor, um ein Histogramm zu zeichnen, was auch gemacht wurde (Abb. 4.3 auf Seite 33).
2. Beim Betrachten des Histogrammes stellt man fest, daß die Daten doch deutlich von der Normalverteilung abweichen. Eine Angabe des Mittelwertes ist also wahrscheinlich nicht aussagekräftig.
3. Es wurden Mittelwert (8,53) und Median (8,8) errechnet und festgestellt, daß der Median deutlich näher am Dichtemaximum²⁶ liegt, wo man ihn auch erwartet hätte.

²⁶Unter dem Dichtemaximum verstehen wir das Maximum der Verteilungskurve, also im Falle der Normalverteilung den Mittelwert.

4. Wir beschließen als Ergebnis den Median und die Quartile anzugeben, ergänzend wird auch der Mittelwert, mit einer kurzen Bemerkung zur Problematik der Verteilung, hinzugefügt.

Das Ergebnis könnte also wie folgt angegeben werden:

Der durchschnittliche Nitratgehalt der Brunnen der Gemeinde xyz beträgt $8,8 \text{ mg} \cdot \text{L}^{-1}$.

(Berechnet wurde der Median, da die Daten von der Normalverteilung abweichen. Der arithmetische Mittelwert beträgt $8,53 \text{ mg} \cdot \text{L}^{-1}$, der interquartile Abstand $0,96 \text{ mg} \cdot \text{L}^{-1}$)

5. Ausreißertests

5.1. Einleitung

Wie schon im vorigen Abschnitt angedeutet, ist es manchmal notwendig zu überprüfen, ob einzelne Werte, die recht deutlich vom Rest der Meßwerte abweichen, als sogenannte *Ausreißer* zu bewerten sind. Natürlich kann ein statistisches Verfahren keine schlechte Meßmethodik korrigieren. **Eine statistische Methode kann nicht entscheiden, ob es gerechtfertigt ist, einen Wert zu entfernen**, sie kann nur behilflich sein festzustellen, ob die Abweichung eines Wertes von den anderen mit einer bestimmten Wahrscheinlichkeit als *signifikant* zu sehen ist, oder eben nicht. In diesem Abschnitt werden zwei Tests vorgestellt, die für diese Überprüfung herangezogen werden können.

Entfernen eines Meßwertes

Wird ein Wert als *signifikant abweichend* klassifiziert, so muß die Überlegung folgen, *warum* dieser Wert abweicht. Handelt es sich um analytische Messungen, und ist das Auftreten solcher Ausreißer selten, so kann man normalerweise diesen Wert als Meßfehler, Verunreinigung o.ä. entfernen. Sollten allerdings regelmäßig solche Werte auftauchen, so ist festzustellen, ob nicht irgend ein systematischer Fehler oder eine Schwäche in der Methodik vorliegt!

Werden ein oder auch mehrere Werte als Ausreißer nach Anwendung einer der Tests entfernt¹, so ist dies im Ergebnis anzumerken!

Bezogen auf Kapitel 3 auf Seite 23 ist die Nullhypothese daß der verdächtige Wert *kein* Ausreißer ist, die Alternativhypothese, daß es sich bei dem Wert auf Signifikanzniveau α um einen Ausreißer handelt.

5.2. 4- σ -Bereich

Diese Faustregel kann verwendet werden wenn *mindestens* 10 Werte, besser aber mehr als 25 Meßwerte vorliegen. Ein Ausreißer wird dann so erkannt, daß die Standardabweichung (s_o) und der Mittelwert (\bar{x}_o) *ohne* den „verdächtigsten“ Wert berechnet werden. Dann überprüft man, ob der mögliche Ausreißer sich außerhalb des Bereiches von $\bar{x}_o \pm 4s_o$ befindet² [10].

¹Umgekehrt können Ausreißertests auch dazu verwendet werden, Meßpunkte (Beobachtungen), die extremere Werte annehmen, aber bedeutungsvoll sein könnten, aufzudecken [21].

²Der 4-Sigma Bereich umfaßt bei großen Stichprobenumfängen und Normalverteilung 99,99 % der Werte. Selbst bei geringeren Stichprobenumfängen werden noch etwa 94 % der Werte abgedeckt.

5.3. Test nach Dean und Dixon

Der Dean und Dixon Ausreißertest wird nach folgendem Schema angewandt:

1. Die Meßwerte werden der Größe nach sortiert.
2. Der Q Testwert wird berechnet (Formel 5.1).
3. Der Q -Wert wird mit der Tabelle A.2 auf Seite 89 im Anhang verglichen.
4. Aus dem Vergleich mit dem Wert in der Tabelle kann geklärt werden, mit welcher Wahrscheinlichkeit der Wert ein Ausreißer ist.

Q wird nach folgender Formel errechnet:

$$Q = \frac{|x_1 - x_2|}{|x_1 - x_n|} \quad (5.1)$$

wobei x_1 der „verdächtige“ Wert, x_2 der dem verdächtigen Wert nächste und x_n der äußerste Werte auf der anderen Seite der sortierten Daten ist. (Der als Ausreißer verdächtige Wert muß ja entweder der höchste oder niedrigste Wert der sortierten Daten sein.) Ist x_1 also der niedrigste Wert der sortierten Meßdaten, so ist x_2 der nächst größere und x_n der größte Wert. Der umgekehrte Fall wird analog behandelt — das folgende Beispiel behandelt einen solchen potentiellen Ausreißer:

Beispiel Liegen in den Meßdaten 4,6; 4,7; 4,3; **5,7**; 5,0; 4,6; 4,6 Ausreißer vor? „Verdächtig“ ist klarerweise zunächst der Wert 5,7. Wir gehen wie oben beschrieben vor:

1. Sortieren: 4,3; 4,6; 4,6; 4,6; 4,7; 5,0; 5,7
2. $Q = (5,7 - 5,0)/(5,7 - 4,3) = (0,7/1,4) = \mathbf{0,5}$
3. Vergleichen mit der Tab. A.2: In der Reihe für $n = 7$ finden wir, daß unser Wert zwischen dem Wert für $Q_{0,90}$ und $Q_{0,95}$.
4. Ergebnis: $Q_{0,9} < 0,5 < Q_{0,95}$ also: Mit einer Irrtumswahrscheinlichkeit von $\alpha = 0,1$ ist die Nullhypothese abzulehnen und die Alternativhypothese zu akzeptieren (es handelt sich um einen Ausreißer), jedoch nicht mehr bei einer Irrtumswahrscheinlichkeit von 0,05.

Der Schluß aus dieser Analyse muß sachbezogen gefällt werden. Im konkreten Fall könnte man schließen, daß der besagte Wert *nicht* mit hoher Wahrscheinlichkeit Ausreißer ist, der Verdacht dennoch groß ist.

5.4. Beispiel 3: „GC-ECD“

Um die Analyse dieses Beispiels haben wir uns im vorigen Abschnitt gedrückt, da das Hilfsmittel des *Ausreißertests* noch nicht bekannt war. Nun soll die Beantwortung dieser Frage nachgeholt werden:

Zur Wiederholung: zunächst wurden drei Werte gemessen, da der dritte Wert recht deutlich von den anderen abweicht, wurden weitere zwei Messungen vorgenommen. Die Werte lauten: 0,38; 0,40; **0,19**; 0,36; 0,42 ppb.

Nun soll überprüft werden, ob es sich beim Wert 0,19 um einen *signifikanten Ausreißer* handelt. Da nur fünf Meßwerte vorliegen, scheidet das „4-Sigma-Verfahren“ von vornherein aus, also halten wir uns an den Ausreißertest nach Dean und Dixon und gehen nach dem beschriebenen Schema vor:

Dean und Dixon

1. Sortieren: **0,19**; 0,36; 0,38; 0,40; 0,42
2. Q berechnen: $Q = |0,19 - 0,36| / |0,19 - 0,42| = (0,17/0,23) = \mathbf{0,74}$
3. Vergleich mit Tab. A.2: Der Q-Wert liegt zwischen den tabellierten Werten für $Q_{0,95} = 0,64$ und $Q_{0,99} = 0,76$, jedoch viel näher beim $Q_{0,99}$ Wert.
4. Ergebnis: Auf einem Signifikanzniveau von fast 1 % ist die Alternativhypothese anzunehmen, und der Wert als Ausreißer zu identifizieren.

Als Folge dieser statistischen Analyse beschließen wir diesen Wert als Meßfehler zu qualifizieren und zu entfernen. Im Ergebnis wird dies jedoch vermerkt werden. Wir nehmen an, daß die Ergebnisse der AAS-Messungen normalverteilt sind und berechnen den Mittelwert

Mittelwert,
 σ bekannt

$$\bar{x} = \frac{0,36 + 0,38 + 0,40 + 0,42}{4} = \underline{\underline{0,39}} \quad (5.2)$$

und den Vertrauensbereich (da die Standardabweichung der Methodik mit $\sigma = 0,018$ bekannt ist, braucht sie hier nicht berechnet werden). Also errechnen wir den Vertrauensbereich nach der Formel:

Vertrauens-
bereich

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} \quad (5.3)$$

da die Vertrauenswahrscheinlichkeit 99 % gewählt wird, ergibt sich für $z = 2,58$ und das Ergebnis lautet:

$$0,39 \pm 2,58 \frac{0,018}{\sqrt{4}} \longrightarrow 0,39 \pm 0,02 \quad (5.4)$$

Das Endergebnis könnte dann etwa so angegeben werden:

Der Bleigehalt der Probe xyz beträgt
 $0,39 \pm 0,02$ ppb. ($\alpha = 0,01$)

*(Es wurde einer von fünf Meßwerten als Ausreißer
identifiziert und entfernt.)*

6. Vergleich von Messungen

6.1. Einleitung

In manchen Fällen tritt das Problem auf, daß man Ergebnisse verschiedener Untersuchungen vergleichen möchte. Folgende Fragen könnten auftreten:

Fragestellungen

1. Sind die Ergebnisse zweier verschiedener Meßserien gleich oder unterscheiden sie sich (mit einer bestimmten Wahrscheinlichkeit) signifikant voneinander? Bspw. kann sich die Frage stellen, ob die Resultate, die zwei verschiedene Labors angeben nur unsignifikant unterschiedlich sind, oder ob es sich um signifikant verschiedene Ergebnisse handelt.
2. Weicht das Ergebnis einer Meßserie von einem erwarteten (gewünschten) Wert ab? Z.B. ist der Gehalt einer Substanz in einer Probe gleich dem gewünschten Wert, oder ist die Abweichung signifikant?
3. Ist die Reproduzierbarkeit zweier verschiedener Meßserien vergleichbar oder unterscheiden sie sich — etwa aus der Frage: ist der eine Analytiker oder die eine analytische Methode dem/der anderen überlegen oder liefern sie vergleichbare Ergebnisse.

Natürlich geht es nicht um die triviale Frage, ob die Ergebnisse zweier Messungen gleich im Sinne von $\bar{x}_1 = \bar{x}_2$ sind. Vielmehr werden Ergebnisse verschiedener Labors oder verschiedener Laboranten sich aufgrund der zufälligen Streuung der Meßwerte fast immer um einen gewissen Betrag unterscheiden. Aber gerade hier stellt sich ja die Frage: *Um wieviel dürfen sich die Werte unterscheiden, daß die Abweichungen noch im Bereich der „normalen“ Streuung liegen.*

Um subjektive Beeinflussung (auch unbewußte) auszuschließen wendet man auch hier statistische Tests an, um diese (und ähnliche) Fragen zu beantworten. Diese Tests helfen dem Analytiker mit einer gewissen (gewählten) Irrtumswahrscheinlichkeit eine Hypothese wie „Diese beiden Mittelwerte sind gleich“ zu überprüfen. Eine kurze allgemeine Einführung in die Testtheorie findet sich in Kapitel 3 auf Seite 23

statistische Tests

Auch in diesem Abschnitt ist zu beachten, daß die hier beschriebenen Tests von der Normalverteilung¹ abgeleitet sind. Es gilt also ähnliches zu beachten wie im Ab-

Anmerkung

¹Wie schon in Abschnitt 4.6 angekündigt, werden auch hier die Symbole s , \bar{x} und μ konsequent verwendet.

schnitt 4 beschrieben. Wenn diese Tests korrekte Ergebnisse liefern sollen, so müssen **die Meßdaten annähernd normalverteilt sein²!**

6.2. Vergleich von Mittelwerten

Um Beispiel 1 zu lösen muß man testen, ob sich zwei Mittelwerte \bar{x}_1 und \bar{x}_2 resultierend aus Meßserien mit n_1 und n_2 Werten statistisch signifikant unterscheiden oder nicht. Um also zu testen, ob die Unterschiede signifikant oder nur innerhalb der „normalen“ Streuung liegen, wendet man den *t*-Test an. Wir gehen von der Nullhypothese „Die Mittelwerte unterscheiden sich nicht“ und der Alternativhypothese „Die Mittelwerte unterscheiden sich signifikant“ aus.

Man errechnet einen Prüfwert nach folgender Formel:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left[\frac{n_1+n_2}{n_1 n_2}\right] \cdot \left[\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right]}} \quad (6.1)$$

Für den Fall, daß die Anzahl der Meßwerte, die den Mittelwerten zugrunde liegen *gleich* sind, also $n_1 = n_2$, kann man eine einfachere Formel verwenden:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} \quad (6.2)$$

Die Freiheitsgrade ergeben sich nach

$$FG = f = n_1 + n_2 - 2 \quad (6.3)$$

Der Unterschied zwischen zwei Mittelwerten gilt als signifikant mit Vertrauenswahrscheinlichkeit P wenn $t_{\text{berechnet}} > t(P, f)_{\text{Tabelle}}$, nachzuschlagen in Tab. A.1.

Beispiel Diese etwas komplex wirkenden Formeln werden hoffentlich anhand eines Beispiels klarer:

Es liegen Mittelwerte von Meßdaten zweier Labors vor:

	Mittelwert \bar{x}	Standardabweichung s	Anzahl Meßwerte
Labor 1	3,67	0,31	5
Labor 2	3,95	0,14	8

Wir setzen in die Formel ein. Da die Anzahl der Meßwerte der beiden Labors unterschiedlich sind, ist Formel 6.2 anzuwenden:

²Laut Sachs [21] ist der *t*-Test für nicht zu kleine Stichproben und nicht zu unterschiedliche Stichprobenumfänge recht stabil gegenüber *Abweichungen von der Normalverteilung*. Trotzdem sollte die Verteilung nicht ganz außer acht gelassen werden!

$$t = \frac{|3,67 - 3,95|}{\sqrt{\left[\frac{5+8}{5 \cdot 8}\right] \cdot \left[\frac{(5-1)0,31^2 + (8-1)0,14^2}{5+8-2}\right]}} = \frac{0,28}{\sqrt{0,325 \cdot 0,04742}} = \underline{\underline{2,26}} \quad (6.4)$$

Nun errechnen wir noch die Freiheitsgrade nach Formel 6.3:

$$FG = 5 + 8 - 2 = \underline{\underline{11}} \quad (6.5)$$

Schlagen wir in der Tabelle nach finden wir für folgende Werte:

$$\begin{aligned} t(0,90;11) &= 1,796 \\ t(0,95;11) &= 2,201 \\ t_{\text{berechnet}} &= 2,26 \\ t(0,99;11) &= 3,106 \end{aligned}$$

Wir erkennen also, daß der von uns berechnete t Wert *zwischen* den tabellierten Werten für 95 % und 99 % liegt, allerdings sehr nahe bei 95 %. D.h. wir müßten die Nullhypothese auf 5 % Signifikanzniveau ablehnen, bei 1 % allerdings beibehalten. Die Frage, ob ein Signifikanzniveau von 5 % ausreichend ist, muß je nach Problemstellung entschieden werden. In kritischen Fällen kann das Risiko hier schon zu hoch sein.

Nullhypothese ablehnen?

Im Falle analytischer Messungen wird man wahrscheinlich zum Schluß kommen, daß die Mittelwerte (also die Meßergebnisse der beiden Labors) mit hoher Wahrscheinlichkeit unterschiedlich sind.

6.3. Vergleich eines Mittelwertes mit einem erwarteten Wert

Es kann vorkommen, daß ein gemessener Mittelwert mit einem bekannten Wert verglichen werden soll. Als Beispiele könnte man anführen: Es wird eine Probe mit bekanntem Inhalt hergestellt um eine analytische Methode oder ein Labor zu testen.

Auch hier wird der t -Test angewandt, allerdings mit etwas modifizierter Formel: t-Test

$$t = \frac{|\bar{x} - \xi|}{s} \sqrt{n} \quad (6.6)$$

Dieser berechnete Wert wird dann — wie im vorigen Abschnitt schon beschrieben — mit den tabellierten Werten in Tab. A.1 verglichen.

Verglichen mit dem t -Test für zwei Mittelwerte ist die Anwendung dieses Tests einigermaßen simpel. Nehmen wir an, ein neues Analysegerät soll getestet werden. Es wird also ein Standard mit einer bestimmten Konzentration der zu messenden Substanz hergestellt. Dann werden 10 Messungen dieses Standards mit dem neuen

Beispiel

Gerät durchgeführt. Die Ergebnisse sind: Der Standard wird mit der Konzentration von 0,40 ppb hergestellt, der Mittelwert der 10 Messungen beträgt 0,37 ppb, die Standardabweichung 0,08 ppb. Wenden wir nun Formel 6.6 an:

$$t = \frac{|0,37 - 0,40|}{0,08} \sqrt{10} = \underline{\underline{1,186}} \quad (6.7)$$

Die Freiheitsgrade sind wieder $n-1$, sehen wir nun in Tab. A.1 nach so finden wir die Werte:

$$\begin{aligned} t_{\text{berechnet}} &= 1,186 \\ t(0,90;9) &= 1,833 \\ t(0,95;9) &= 2,262 \\ t(0,99;9) &= 3,250 \end{aligned}$$

Folglich müssen wir die Nullhypothese, daß sich der Mittelwert vom erwarteten Wert unterscheidet beibehalten. Mit anderen Worten: Das Ergebnis, das das Meßgerät ermittelt, ist im Bereich der „normalen“ Streuung. Eine andere Frage ist allerdings, ob man mit der Größe der Standardabweichung (= zufälliger Fehler) zufrieden ist. Ob der zufällige Fehler in der Norm liegt, oder zu groß ist, muß man auf andere Art und Weise ermitteln (Herstellerangaben, Literatur, Vergleich mit anderen Geräten³). Für den Fall, daß die Standardabweichung im normalen Bereich liegt, ist anzunehmen, daß **das Gerät entsprechend der Spezifikation arbeitet**.

6.4. Vergleich von Standardabweichungen

Zum Vergleichen zweier Standardabweichungen s_1 und s_2 , die bspw. aus Meßserien zweier verschiedener Laboranten stammen, verwendet man den *F-Test*⁴:
Zunächst berechnet man den F-Wert nach folgender Formel:

$$F = \frac{s_1^2}{s_2^2} \quad (6.8)$$

Der Bruch ist so zu wählen, daß $F > 1$ ist! Also für den Fall, daß ein F -Wert kleiner 1 herauskommt, sind Zähler und Nenner zu vertauschen. Diesen Wert vergleicht man nun mit den tabellierten Werten (siehe Tabellen A.3 bis A.11 ab Seite 90). Um in der Tabelle korrekt nachschlagen zu können, muß man drei Parameter

³Hier könnten wir das Beispiel natürlich weiterführen: Angenommen, man ist im Besitz eines zweiten Gerätes dieser Art bzw. kennt jemanden der mit einem solchen Gerät arbeitet: In diesem Fall könnten wir dieselbe Probe mit dem anderen Gerät messen und dann mittels F -Test die Standardabweichung prüfen. So könnten wir feststellen, ob unser Gerät mit einem vergleichbaren Fehler arbeitet oder nicht.

⁴Die Mittelwerte der beiden Meßserien müssen nicht unbedingt gleich sein. Will man bspw. nur die Zuverlässigkeit eines Gerätes prüfen, so ist der Mittelwert weniger interessant, sondern vielmehr interessiert man sich für die Standardabweichung, die das Gerät bei verschiedenen Messungen „produziert“. Diese kann man dann auch unabhängig vom eigentlichen Meßwert mit einem anderen Gerät oder einer anderen Methodik vergleichen.

wissen: Freiheitsgrad für s_1 und s_2 , diese werden mit f_1 und f_2 bezeichnet, und die Vertrauenswahrscheinlichkeit nach der getestet werden soll (P), also z.B. $P = 0,95$. (Zur Wiederholung: Der Freiheitsgrad ist die Anzahl der Meßwerte - 1.) Die Standardabweichungen gelten mit der Wahrscheinlichkeit P als unterschiedlich, wenn $F_{berechnet} > F(P, f_1, f_2)_{Tabelle}$.

Zwei Laboranten messen dieselbe Probe eines industriell gefertigten Rohstoffes. Es soll getestet werden, ob die Ergebnisse beider Laboranten eine vergleichbare Reproduzierbarkeit aufweisen, oder ob einer von beiden signifikant schlechter arbeitet:

Beispiele

	Standardabweichung	Anzahl Meßwerte	f
Laborant 1	$s_1 = 0,034$	6	5
Laborant 2	$s_2 = 0,051$	7	6

F errechnet sich also nach Formel 6.8: $F = 0,034^2 / 0,051^2 = 0,4$. Dies ist kleiner als eins, folglich „drehen wir den Bruch um“: $F = 0,051^2 / 0,034^2 = 2,25$. Schlägt man in der Tabelle A.6 nach, so findet man für $F(0,95; 6; 5) = 4,95$. Da 2,25 eindeutig *nicht größer* als der tabellierte Wert 4,95 ist, **kann ein signifikanter Unterschied nicht festgestellt werden**, und die Alternativhypothese muß zugunsten der Nullhypothese H_0 abgelehnt werden.

Standardabweichungen von Meßserien zweier verschiedener Meßmethodiken sollen getestet werden. Es stellt sich die Frage, ob die Präzision der einen Methode signifikant höher ist, als die der anderen:

	Standardabweichung	Anzahl Meßwerte	f
Anal. Methode A	$s_1 = 0,087$	9	8
Anal. Methode B	$s_2 = 0,22$	9	8

Zunächst errechnen wir wieder den F Wert: $F = 0,22^2 / 0,087^2 = 6,39$. Der tabellierte F -Wert ist: $F(0,95; 8; 8) = 3,44$ (Tab.A.6) und $F(0,99; 8; 8) = 6,03$ (Tab. A.9). Wir sehen sofort, daß $F_{berechnet} > F_{Tabelle}$ gilt, sogar für die Wahrscheinlichkeit von 99%. Somit können wir mit statistisch signifikant schließen, daß sich die **beiden Methodiken in ihrer Standardabweichungen unterscheiden** (Irrtumswahrscheinlichkeit unter 1%).

Schon aus diesen beiden Beispielen zeigt sich, daß für nur wenige Meßwerte erst ein recht großer Unterschied in den Standardabweichungen als signifikant unterschiedlich gilt. Bspw. darf bei $f_1 = f_2 = 3$ ein Unterschied erst dann angenommen werden, wenn eine Standardabweichung etwa dreimal so groß ist wie die andere.

Die Konsequenz ist, daß zum Testen eine **große Anzahl von Meßwerten** anzustreben ist. Will man zwei ähnliche Methoden — wie im zweiten Beispiel — untersuchen, sollten eher *mehr als 10 Meßwerte* vorliegen. Auch das erste Beispiel ist vermutlich letztlich nicht schlüssig, da bei so wenig Meßwerten eben die Unterschiede recht groß sein müssten.

Anmerkung:
Stichprobengröße

7. Kausaler Zusammenhang zwischen Variablen

7.1. Einleitung

Oft steht man vor der Aufgabe, Zusammenhänge zwischen Variablen ermitteln zu wollen bzw. quantitative Aussagen darüber zu treffen. Beispielsweise könnte man sich die Frage stellen, ob der Alkoholgehalt im Blut mit der Unfallhäufigkeit zusammenhängt; oder man ist im Besitz von Messungen verschiedener Schadstoffe an unterschiedlichen Standorten und versucht festzustellen, ob die Konzentration irgendeines Schadstoffes mit dem Auftreten einer bestimmten Krankheit in Zusammenhang steht.

Die Korrelationsrechnung kann einen Anhaltspunkt geben, ob zwischen zwei Variablen¹ ein statistischer Zusammenhang besteht. Es soll schon an dieser Stelle auf einen häufigen Fehler aufmerksam gemacht werden:

Aus statistischer Korrelation darf nicht auf Kausalität geschlossen werden. Statistische Korrelation kann einen HINWEIS auf einen kausalen Zusammenhang geben, bewiesen muß dieser aber mit anderen Methoden werden!

Beachtet man diesen Hinweis nicht, folgen daraus Beweise für Zusammenhänge zwischen Schuhgröße und Einkommen, man kann nachweisen, daß Kinder vom Storch gebracht werden oder muß zur Kenntnis nehmen, daß praktisch ausschließlich Ausländer für Kriminalität verantwortlich sind. Daß dies absurd und nicht haltbar ist, leuchtet wohl zumindest dem Naturwissenschaftler sofort ein, leider offensichtlich nicht jedem Politiker und Demagogen.

7.2. Beispiele

Beispiel 1: Es liegen Meßwerte eines Schadstoffes an verschiedenen Orten vor. Zusätzlich hat man das Auftreten einer bestimmten Krebsart an diesen Orten erhoben.

¹Prinzipiell ist man nicht auf zwei Variable beschränkt. Es ist auch möglich Zusammenhänge zwischen mehr als einer *AusgangsvARIABLE* und einer *Zielvariable* zu ermitteln — in diesem Fall spricht man von *multivariater Statistik*. Dies geht allerdings deutlich über den Rahmen dieses Skriptums hinaus und es muß auf weiterführende Literatur verwiesen werden [2, 10].

Es soll ermittelt werden, ob ein statistischer Zusammenhang zwischen diesen beiden Variablen zu erkennen ist. Die Daten sind in Tab. 7.1 zusammengefaßt.

Beispiel 2: Es werden Luftschadstoffe (z.B. Nitrat, Sulfat, pH, Chlorid, Ammonium, verschiedene organische Ionen, Calcium, Magnesium, etc.) an verschiedenen Stationen über einen längeren Zeitraum gemessen. Mit Hilfe der Korrelationsrechnung soll ermittelt werden, ob Zusammenhänge zwischen den Konzentrationen verschiedener Substanzen auftreten.

7.3. Scatter-Plot

Als ersten Schritt einer Korrelations- oder Regressionsanalyse sollte man immer einen sogenannten *Scatterplot* zeichnen. Das macht einerseits das Datenmaterial anschaulich und kann andererseits grobe Interpretationsfehler verhindern.

Ein Scatterplot ist eine zwei-dimensionale Graphik, bei der paarweise Meßdaten dargestellt werden. Auf der horizontalen x - Achse (auch *Abszisse* genannt) wird die eine Variable aufgetragen, auf der vertikalen y - Achse (auch *Ordinate* genannt) die andere.

Beispiel Zeichnen wir nun den Scatterplot von Beispiel 1. In Abb. 7.1 ist dieser Plot dargestellt. Exemplarisch ist der erste Punkt (Ort 1) herausgehoben. Alle anderen Orte wurden in derselben Art und Weise eingetragen.

Wir erkennen schon anhand dieser Graphik daß ein linearer Zusammenhang nahelegt. Nun wollen wir diesen Zusammenhang quantifizieren und zu diesem Zweck den *Korrelationskoeffizienten* errechnen.

Anmerkung Es ist sehr wichtig zwischen *Scatterplots* und einfachen *Lineplots* zu unterscheiden! Im speziellen gilt das dann, wenn man zum Erstellen der Graphik eine gängige Tabellenkalkulation wie z.B. Excel verwendet. Der Unterschied ist der: bei Scatterplots werden zwei Variable gegeneinander aufgetragen, wobei beide Achsen entspre-

Ort	Konzentration Schadstoff	Anzahl Krebsfälle
1	6,1	68
2	3,85	50
3	5,7	68
4	9,74	110
5	9,12	99
6	7,31	79
7	8,33	98
8	9,6	114
9	6,15	79
10	11	119
11	3,25	46
12	6,6	81
13	3,4	51
14	4,3	61

Tabelle 7.1.: Beispiel: Konzentration eines Schadstoffes und Auftreten von Krebs.

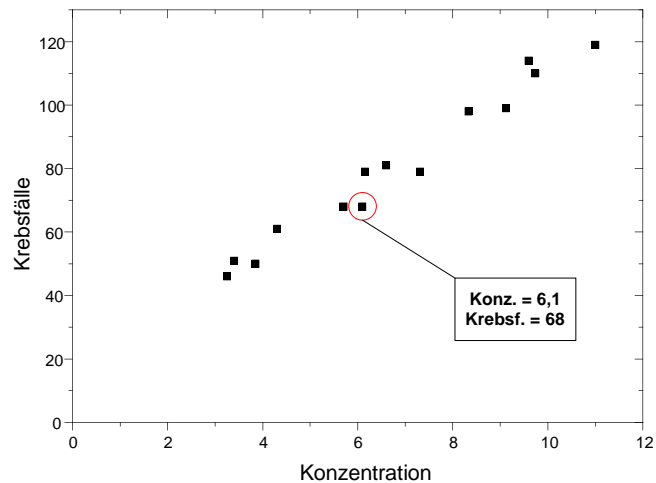


Abbildung 7.1.: Scatterplot des Datensatzes „Konzentration eines Schadstoffes und Auftreten einer bestimmten Krebsart.“

chend der Werte der Variablen korrekt skaliert werden. Bei Lineplots wird einfach eine Variable Wert für Wert aufgetragen, wobei die Abstände zwischen den Werten an der x - Achse gleich sind (also equidistant, normalerweise mit Abstand = 1).

Lineplots dienen also nur zum Darstellen von Wertpaaren, bei denen die Abstände an der x - Achse entweder wirklich **in jedem Fall** equidistant sind, oder wo dieser Abstand nicht von Bedeutung ist. Dies kommt im Falle analytisch-chemischer Daten aber nur in den seltensten Fällen vor! Sind z.B. die Werte einer Messung zeitabhängig, also werden bspw. alle paar Tage Messungen durchgeführt, aber nicht immer in regelmäßigen Abständen, so führt die Verwendung eines einfachen Lineplots zu einer Verzerrung der Abbildung. Ich weise darum speziell auf diese Gefahr hin, da Lineplots oft die Standardeinstellung des Graphiktyps sind, und man so leicht in diese Fallgrube stolpert. Der Unterschied wird in Abb. 7.2 anhand einer Zeitreihe illustriert.

Eine Zeitreihe ist eine Meßserie, bei der irgendein Wert abhängig von der Zeit gemessen wird. Selbst für den Fall, daß die Zeitintervalle an und für sich equidistant sind, muß man auf diese Problematik achten, da es kaum eine Meßserie gibt, bei der nicht der eine oder andere Wert fehlt. Für den Chemiker ist in der Praxis meist der Scatterplot angezeigt, der Lineplot findet nur in Ausnahmefällen Verwendung.

Zeitreihen

7.4. Korrelation

Der Korrelationskoeffizient ist ein statistisches Maß für die *lineare* Abhängigkeit zweier Variabler voneinander. Der Korrelationskoeffizient wird üblicherweise mit r bezeichnet. r kann Werte zwischen -1 und 1 annehmen:

mögliche Werte für r

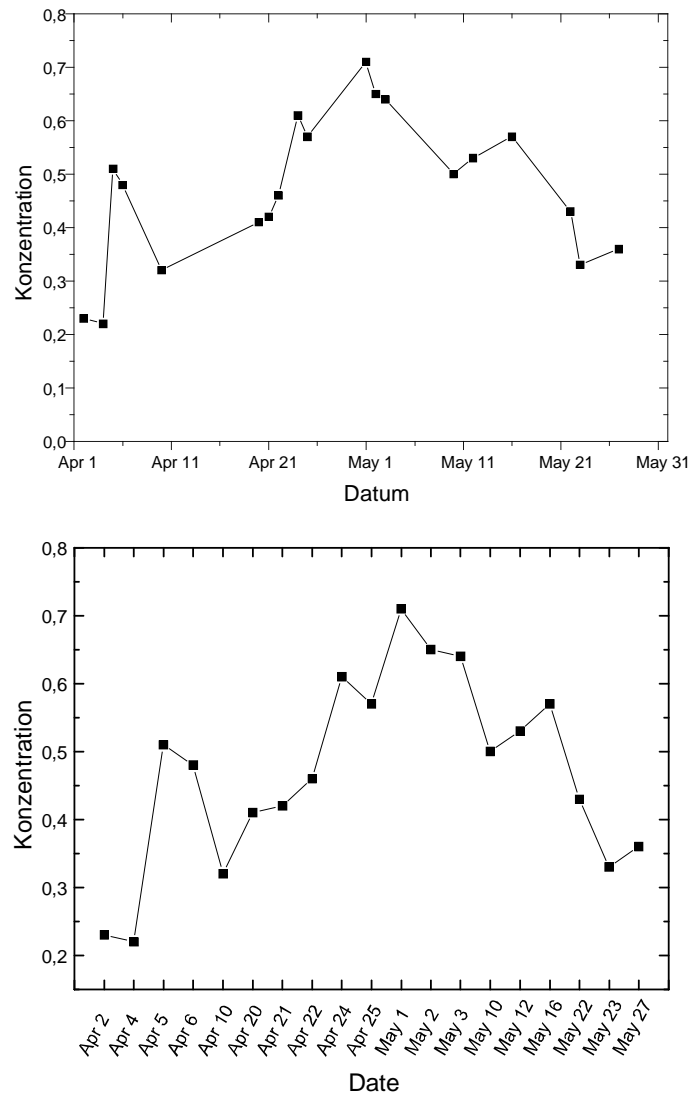


Abbildung 7.2.: Die obere Abb. zeigt eine Zeitreihe korrekt als Scatterplot dargestellt. Die untere Abb. hingegen stellt dieselben Daten dar, aber als Lineplot ausgeführt:

Man sieht sofort, daß die Werte einfach aneinandergereiht werden. Da die Meßintervalle aber unregelmäßig sind, führt dies zu einer dramatischen Verzerrung der Abbildung. **Die untere Abb. ist daher irreführend und muß unbedingt vermieden werden!** Ein weiterer gängiger Fehler wurde in der unteren Abb. gemacht: Die Werte an der y - Achse beginnen nicht — wie man erwarten würde — mit 0 sondern mit einem Wert größer 0. **Dies führt auch noch zu einer Verzerrung der y - Achse.**

$r = -1$	Es liegt ein negativer funktionaler Zusammenhang vor.
$-1 < r \leq 0$	Es liegt eine negative Korrelation vor.
$r = 0$	Die Variablen sind unabhängig voneinander.
$0 < r \leq 1$	Es liegt eine positive Korrelation vor.
$r = 1$	Die Variablen sind funktional abhängig.

Was bedeuten die Angaben in der Liste? Eine „Sonderstellung“ nehmen die Werte -1 ; 0 und 1 an. -1 und 1 stehen für einen sogenannten *funktionalen* Zusammenhang. D.h. eine Variable läßt sich aus der anderen mit einer linearen Funktion exakt errechnen. Graphisch bedeutet dies, daß im Scatter Plot die Fälle -1 und 1 eine Gerade ergeben, wobei die Gerade -1 eine negative Steigung (also von links oben nach rechts unten weist), die Gerade für $+1$ hingegen eine positive Steigung aufweist. 0 bedeutet, daß beide Variablen völlig unabhängig voneinander sind.

Diese drei Werte nehmen insofern eine Sonderstellung ein, als sie in der Praxis nie vorkommen; sie sind vielmehr „Extremwerte“, an die sich reale Ergebnisse annähern. Für die Praxis bedeutet dies, daß r *nahe* bei 1 oder -1 anzeigt, daß die Variablen korrelieren, Werte nahe bei 0 , daß sie unabhängig voneinander sind. Allerdings beobachtet man bei Meßdaten oft r -Werte, die relativ weit von 0 entfernt sind. In der Tat spricht man üblicherweise erst bei Werten von $r > 0,9$ bzw. $r < -0,9$ von einem Zusammenhang zwischen zwei Variablen.

Den Korrelationskoeffizienten r errechnet sich nach

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2] [n \sum y_i^2 - (\sum y_i)^2]}} \quad (7.1)$$

Manchmal wird auch das Quadrat des Korrelationskoeffizienten angegeben, wobei r^2 den Vorteil hat, nur positive Werte anzunehmen. r^2 wird auch als *Bestimmtheitsmaß* bezeichnet. Auch hier gilt: je näher r^2 bei 1 liegt, desto besser läßt sich die Abhängigkeit zweier Variable durch eine Gerade beschreiben. Das Bestimmtheitsmaß läßt sich auch so interpretieren, daß der Wert *den Prozentsatz der Varianz angibt, der durch das lineare Modell erklärt wird*. Salopp formuliert kann man das so verstehen: ist $r^2 = 0,97$ so können 97% der Daten durch das lineare Modell (die Gerade) erklärt werden.

7.5. Ergebnis

Das Ergebnis des „Krebsbeispiels“ wird hier nicht wie in den vorigen Beispielen explizit ausgerechnet, da die Summenbildung recht umfangreiche und wenig instruktive Formeln ergeben würde. Davon abgesehen ist sowieso bei Formeln diesen Umfangs und der heutigen Verbreitung von Computern und leistungsfähigen Taschenrechnern davon abzuraten, diesen Wert „mit der Hand“ zu berechnen.

Für Beispiel 1 ergibt sich also (mit einer Tabellenkalkulation aus Tab. 7.1 berechnet) ein Wert von: $r = 0,986$ und $r^2 = 0,973$. Wir können also schließen, daß beide Variablen sehr gut korrelieren, ein Zusammenhang kann also **vermutet** werden.

Praxis

Korrelationskoeffizient r

Bestimmtheitsmaß r^2

Beispiel 1

Kausalität? Auch hier gilt, wie schon in der Einleitung erwähnt, daß diese Korrelation niemals ein *Beweis* für eine Kausalität — also einen tatsächlichen Zusammenhang sein kann. Allerdings sollte dieses Ergebnis Anlaß sein, mit anderen Mitteln (biochemischen, medizinischen, ...) nachzuprüfen, ob tatsächlich ein Zusammenhang vorliegt.

Unser Beispiel besteht nur aus zwei Variablen. In der Praxis wird der Fall komplizierter liegen: Beispielsweise könnten die Anzahl der Krebserkrankungen und mehrere Dutzend Meßwerte verschiedener Schadstoffe vorliegen. Nun lassen sich unter Verwendung der Korrelationsrechnung die Korrelationskoeffizienten *aller* bekannten Schadstoffe mit der Krebsrate berechnen. Dann kann man alle diejenigen ausschließen, bei denen keine statistische Korrelation vorliegt. Alle anderen, die ähnlich korrelieren wie unser Beispiel, sollten dann für eine weitere Untersuchung vorgeschlagen werden. Das bedeutet:

Die Korrelationsrechnung kann uns helfen, Zusammenhänge AUSZUSCHLIESSEN aber nicht sie zu BEWEISEN!

Beispiel 2 Auch bei Beispiel 2 muß man bei der Interpretation von Korrelationen zwischen verschiedenen Ionen vorsichtig sein. Findet man z.B. hohe Korrelationen zwischen Calcium und Chlorid, nicht aber zwischen Calcium und anderen Anionen, so kann das ein **Hinweis** darauf sein, daß das Calcium eben als Calciumchlorid vorliegt. Oder man findet hohe Korrelationen zwischen Calcium und Magnesium und die Meßstationen liegen vorzugsweise in den Kalkalpen, so liegt der Schluß nahe, daß man möglicherweise vorwiegend $\text{CaMg}(\text{CO}_3)_2$, also Dolomit gemessen hat. Dies ist unter Umständen auch aus der Lage der Stationen und der Tatsache, daß Dolomit in großen Mengen als Baumaterial, Straßenunterlage, zur Kalkherstellung und vielfach in der chemischen Industrie verwendet wird, erklärbar [11]. Tritt eine solche Korrelation nur vereinzelt auf, könnte folglich auch Kontamination durch Industrie, Baustellen o.ä. vorliegen.

Bestimmen der Geraden Die nächste Frage, die sich aus diesem Abschnitt wohl zwangsläufig stellt, ist, ob man für den Fall daß zwei Variable korrelieren, die Gerade bestimmen kann, die diese Werte optimal beschreibt. Die Berechnung dieser Gerade wird als *Regression* bezeichnet und im nächsten Abschnitt anhand der Bestimmung einer Eichgeraden (= Kalibrierungsgerade) erklärt.

7.6. Korrelation = Kausalität ?

7.6.1. Einleitung

Es wurde schon mehrfach erwähnt, daß aus einem hohen Korrelationskoeffizienten **nicht** sofort auf einen direkten kausalen Zusammenhang zwischen den korrelierenden Variablen geschlossen werden darf (selbst wenn dies in Tageszeitungen und Illustrierten regelmäßig praktiziert werden mag). Statistische Analysen alleine können niemals entscheiden, ob ein kausaler Zusammenhang zwischen zwei oder auch mehreren Variablen vorliegt.

Es sind immer weiterführende (nicht-statistische) Verfahren nötig um einen *Verdacht* auf Kausalität zu überprüfen. Bevor man jedoch weitere (möglicherweise auf-

wendige) Untersuchungen vornimmt, sollte man sich überlegen, ob einer der folgenden Fälle zutrifft und sich somit weitere Tests erübrigen :

7.6.2. Korrelation zwischen Zeitreihen

Wenn es sich bei beiden Variablen x und y um Zeitreihen handelt, ist besondere Vorsicht geboten. Die Praxis zeigt, daß es nur sehr selten vorkommt, daß Zeitreihen *keinen* Trend zeigen. Für den Fall, daß bspw. x einen steigenden Trend aufweist und y ebenfalls, ist eine positive Korrelation, falls y einen absteigenden Trend zeigt, so ist negative zu erwarten. Dies kann unabhängig davon gelten, ob ein tatsächlicher Zusammenhang vorliegt oder nicht.

Beispiel: Die (ansteigende) Anzahl der Verkehrsunfälle korreliert mit Scheidungsrate, Folgerung: *Personen, die einen Verkehrsunfall erleiden werden öfter geschieden??*

Verkehrsunfälle
und Scheidungen

7.6.3. Formale Korrelation

Werden beide Variablen x und y durch dieselbe Zahl z dividiert, so korrelieren x/z mit y/z . Dies gilt z.B. bei sich ergänzenden Prozent-Angaben. Sind bspw. Variable A und Variable B in Prozent angegeben (Prozent Fett in Milch, ...) und ergänzen sie sich auf 100 %, also

$$A + B = 100\% \quad (7.2)$$

so korrelieren A und B selbstverständlich miteinander, obwohl kein Zusammenhang bestehen muß.

Beispiel: Prozent von Fett und Protein im Blut.

Fett und Protein

7.6.4. Korrelation durch Inhomogenität

Dieser Fall ist ein für Chemiker recht wichtiger, da er in der Meßpraxis recht häufig auftreten kann. Sind die Daten der beiden Variablen nicht homogen, sondern liegen z.B. zwei nicht zusammenhängende Gruppen vor, kann es zu diesem Fehler kommen.

Beispiel: Schuhgröße korreliert mit Einkommen! Lösung: Männer haben üblicherweise größere Füße als Frauen, weiters verdienen Männer im Schnitt mehr als Frauen. Diese beiden Variable haben aber logischerweise nichts miteinander zu tun. Zeichnet man den Scatterplot der Schuhgröße von Frauen und dem Einkommen von Frauen ergibt sich eine Punktwolke, eine Berechnung des Korrelationskoeffizienten ergibt keine Korrelation. Dasselbe Ergebnis erhält man bei den Daten für Männer. Wirft man jedoch beide Gruppen in einen Topf, so erhält man einen Scatterplot, der zwei Gruppen aufweist: Die eine bei kleineren Schuhgrößen und niedrigerem Einkommen (Frauen), die andere bei größeren Schuhgrößen und höheren Einkommen.

Schuhgröße
korreliert
mit Einkommen

Aus dem Scatterplot ist aber ebenso sofort klar, daß in Wahrheit *kein* Zusammenhang vorliegt, sondern eben zwei Gruppen. Errechnet man jedoch nur den Korrelationskoeffizienten, so findet man plötzlich eine Korrelation (Illustration siehe

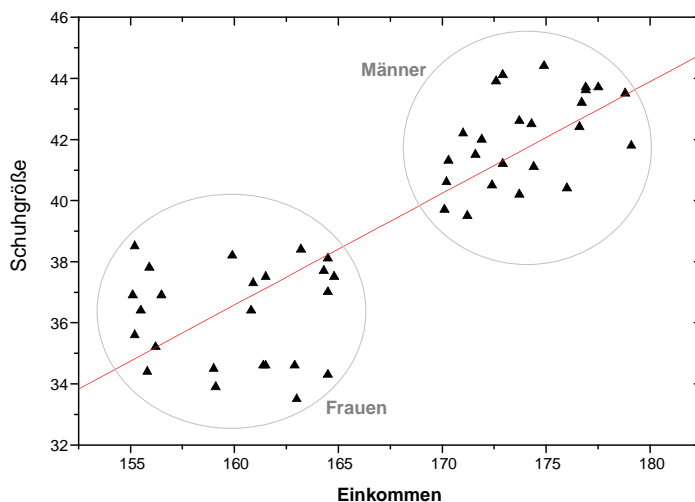


Abbildung 7.3.: Inhomogenitätskorrelation: Dieser Datensatz ist konstruiert und basiert nicht auf realen Daten, entspricht aber im Prinzip der Realität: Man sieht deutlich, daß „Frauen“ und „Männer“ zwei Gruppen bilden. Innerhalb der Gruppen gilt: $r_{\text{Frauen}} = 0,28$ und $r_{\text{Männer}} = -0,24$, also keine erkennbare Korrelation (dieses Ergebnis ist ja auch zu erwarten). Allerdings $r_{\text{beide Gruppen}} = 0,85$! Dies zu interpretieren wäre ein klarer Fehlschluß!

Abb. 7.3)! Anhand dieser Abbildung zeigt sich sofort die Aussagekraft der Scatterplots; man erkennt die Sinnlosigkeit der Annahme, es läge ein kausaler Zusammenhang vor. Diese ist eine sogenannte *Inhomogenitätskorrelation*.

Ein ähnliches Ergebnis kann man übrigens auch bei nicht korrelierenden Meßdaten, die jedoch über Ausreißer verfügen, erhalten: **dies ist ein in der analytischen Chemie durchaus häufiger Fall**. Bereits einige wenige Ausreißer reichen aus, um Korrelation vorzuteuschen, in manchen Fällen sogar ein einzelner! Da dies jedoch in einem Scatterplot sofort zu erkennen ist, wird klar, warum auf die korrekte Anwendung dieser Graphik großer Wert gelegt wird.

7.6.5. Korrelation aufgrund einer gemeinsamen Basis

Variable x korreliert mit y . Allerdings ist ein dritter Parameter z.B. z sowohl für x als auch für y verantwortlich. Zwischen x und y liegt jedoch kein Zusammenhang vor².

Verheiratete leben
länger!

Beispiel: Verheiratete Männer leben länger als ledige. Natürlich *könnte* ein Zusammenhang bestehen. Wahrscheinlicher ist jedoch, daß beide Variablen durch eine (oder mehrere) andere Größe(n) verursacht werden. Z.b. könnten charakterliche Ei-

²Weiterführende Analysen können in manchen Fällen durch Berechnung partieller Korrelationskoeffizienten erfolgen. Hierbei wird versucht, den Einfluß einer Größe auf andere auszuschalten. Diese Analysen gehen jedoch zu weit von über die hier erwähnten Probleme hinaus. Für Details siehe auch [21].

enschaften (höhere Risikobereitschaft, o.ä.) dazu führen, daß sie nicht heiraten *und* früher sterben (z.B. an einem Unfall).

Ein anderes Beispiel: Alle Ausländer sind Kriminelle, da die Anzahl der Ausländer mit der Kriminalität korreliert. Diese Aussage ist natürlich falsch und kann z.B. so zustande kommen: Große Städte verfügen über mehr Kriminalität (viele verschiedene Ursachen), gleichzeitig ziehen Städte aber auch Ausländer an (Arbeitsplätze, leichtere Integration, ...). Somit korrelieren diese beiden Größen, ohne daß ein direkter Zusammenhang vorliegen muß.

Ausländer sind
Kriminelle?

7.6.6. Kausalität?

Kann man alle oben erwähnten Fälle letztendlich ausschließen, so ist Korrelation immer noch **kein Beweis** für Kausalität. Allerdings hat man jetzt einen Hinweis auf einen möglichen kausalen Zusammenhang und sollte diesem mit anderen Methoden nachgehen. Bspw. sollten dann biochemische, medizinische Untersuchungen oder der Versuch chemische Reaktionen nachzuweisen, die eine solche statistische Korrelation untermauern, folgen.

7.7. Zusammenfassung

Vielleicht wirkt die Korrelationsrechnung nach dem bisher gesagten ein wenig verwirrend oder deren Ergebnisse nicht signifikant. Wie sollte man also vorgehen?

1. Zeichnen des Scatterplots: Mit Hilfe des Scatterplots kann man schon einige wesentliche Fehler ausschließen (vgl. Inhomogenitätskorrelation, Ausreißer)
2. Berechnen des Korrelationskoeffizienten: Ist dieser *nicht* nahe bei 1 oder -1, so kann man einen kausalen Zusammenhang ausschließen.
3. Ist der Korrelationskoeffizient (r) nahe bei 1 oder -1, so überlegt man sich, ob eine der in Abschnitt 7.6 beschriebenen Fälle zutrifft.
4. Für den Fall, daß weder Inhomogenitäten, gemeinsame Basis, formale Korrelation oder die beschriebenen Probleme bei Zeitreihen vorliegen, sollte man einen hohen Korrelationskoeffizienten als **Indiz** für einen kausalen Zusammenhang interpretieren und weitere Untersuchungen folgen lassen.

Weitere Beispiele wie Korrelation, Regression und Statistik im allgemeinen in Tageszeitungen, Fernsehen, Werbung, etc. falsch verwendet wird, findet sich in unterhaltsamer Form in dem sehr empfehlenswerten Buch „So lügt man mit Statistik“ [16]. Viele Diagramme und Zahlen mit denen man täglich konfrontiert ist erscheinen danach in neuem Lichte.

8. Bestimmen einer Eichgerade

8.1. Einleitung

Es ist oft wünschenswert, nicht nur Zusammenhänge zwischen Variablen aufzudecken, wie dies im Kapitel über Korrelation beschrieben wurde, sondern auch konkrete Modelle zu entwickeln, um diese Zusammenhänge zu beschreiben. Ein für den Chemiker wichtiges Beispiel ist die Bestimmung einer *Eichgeraden*.

Das Bilden eines linearen Modells bezeichnet man auch als *lineare Regression*: die Suche nach einer Geraden, die sich optimal an die vorhandenen Daten anpaßt. Diese Gerade wird auch *Ausgleichsgerade* genannt. Die beiden Variablen nennen wir x und y , wobei x die *unabhängige* und y die *abhängige* Variable ist. D.h. x ist „vorgegeben“ und y soll durch das lineare Modell (die Gerade) beschrieben werden. Mathematisch ausgedrückt heißt das, daß y eine Funktion von x ist: $y = f(x)$, also x in das Modell eingesetzt wird, um y zu berechnen. Dieses Modell kann dazu benutzt werden, für nicht bekannte x -Werte zugehörige y Werte zu bestimmen.

Ausgleichsgerade

In diesem Kapitel wird zunächst der wichtige Unterschied zwischen *Interpolation* und *Extrapolation* beschrieben. Anhand des Beispiels einer Eichgeraden wird dann die Vorgangsweise der Regressionsrechnung und der Einsatz verschiedener graphischer Verfahren wie Scatterplot und Analyse der Residuen beschrieben.

8.2. Interpolation, Extrapolation

Das beschriebene lineare Modell kann dazu benutzt werden um für einen bestimmten x -Wert, der *zwischen* zwei bekannten Werten liegt den dazugehörigen y -Wert abzuschätzen. Dies nennt man *Interpolation*

Setzt man in die Gleichung einen x -Wert ein, der *außerhalb* des Bereiches der bekannten Werte liegt, um den entsprechenden y -Wert zu bestimmen, spricht man von *Extrapolation*

Grundsätzlich ist beides möglich, allerdings muß vor Extrapolationen, v.a. wenn die Werte weit vom bekannten Bereich entfernt sind, gewarnt werden. Schließlich wurde das lineare Modell nur aus den bekannten Werten bestimmt und ebenso ist die Gültigkeit dieses Modells außerhalb des „bekannten Bereichs“ nicht gesichert. Führt man trotzdem Extrapolationen durch muß man sich im klaren darüber sein, daß der **zu erwartende Fehler umso größer sein wird, je weiter man sich vom „bekannten“ Bereich wegbewegt**. Und dies auch nur dann, wenn das lineare Modell auch außerhalb des bekannten Bereiches gültig ist.

Dies ist nicht selbstverständlich. Gerade in der Naturwissenschaft sind die meisten Prozesse nicht linearer Natur, können aber in einem kleinen, eingeschränkten Bereich durch ein lineares Modell angenähert werden. Wendet man nun die Extrapolation „großzügig“ an, kann man in einen Bereich vordringen, in dem diese lineare Annäherung keineswegs mehr richtig ist, und das Ergebnis kann vom „richtigen“ Wert ziemlich stark abweichen.

Beispiel Als drastisches Beispiel könnte man die Sinusfunktion anführen. In einem schmalen Bereich um den Nullpunkt kann man die Sinusfunktion $y = \sin(x)$ gut durch eine Gerade $y = x$ annähern¹ (siehe Abb. 8.1). Dies wird auch vereinzelt so gehandhabt. Verläßt man aber diesen engen Bereich und versucht mit dieser Geraden zu extrapolieren, wird man bald einen extrem großen Fehler begehen!

8.3. Beispiel

Analytik Ein Analytiker steht vor der Aufgabe eine neue Substanz S quantitativ aus Bodenproben zu bestimmen. Die Trennung und Identifikation dieser Substanz mittels *Hochdruck-Flüssigkeits-Chromatographie* (HPLC) ist bereits erfolgreich durchgeführt worden und die Analyse funktioniert zuverlässig. Die Menge einer Substanz, also die *Konzentration*, ist bekanntlich proportional zur Fläche des Peaks im Chromatogramm².

Peakfläche Um die Quantifizierung vornehmen zu können, muß bekannt sein, welche *Fläche* welcher *Konzentration* entspricht. Da es nicht sinnvoll möglich ist für jede mögliche Konzentration eine geeichte Probe herzustellen und die Fläche zu bestimmen, stellt man nur eine kleine Anzahl Eichproben her, die den Konzentrationsbereich abdecken,

¹Die ist auch durch eine Reihenentwicklung zu erklären. Nähert man die Sinusschwingung durch eine Taylorreihe an, und bricht diese nach dem ersten Glied ab, so erhält man eben diese Gerade.

²Unter einem Chromatogramm versteht man das Ergebnis einer chromatographischen Untersuchung. Dabei werden Substanzen anhand ihrer Affinität zur chrom. Säule getrennt. Je stärker diese Aff. ist, desto länger benötigt die Substanz um die Säule zu verlassen. Ein am Ende der Säule befindlicher Detektor registriert austretende Substanzen. Die Signale, die diese Substanzen am Detektor verursachen, werden aufgrund ihrer Form *Peaks* genannt.

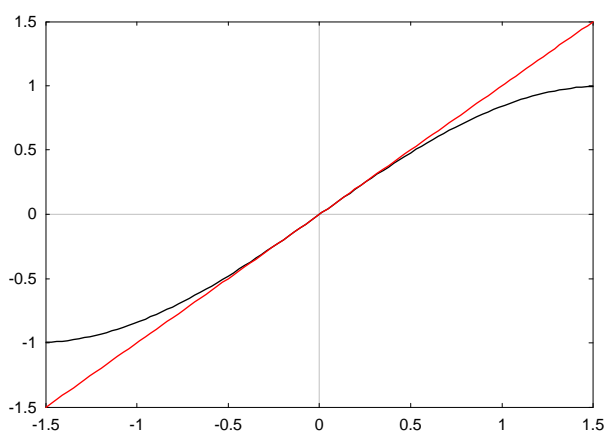


Abbildung 8.1.: Lineare Approximation der Sinusfunktion.

in dem man die Meßergebnisse erwartet. Diese werden mit der beschriebenen analytischen Methode gemessen.

Dann kann man einen Scatterplot zeichnen, wobei die eine Achse die *Konzentration* der Eichproben darstellt, die andere Achse die *Fläche* des entsprechenden Peaks im Chromatogramm. Mißt man dann eine Probe unbekannter Konzentration, kann man aus diesem Diagramm, oder noch einfacher aus der zugehörigen Geradengleichung aus der Fläche einfach auf die Konzentration schließen. Diese Kurve heißt auch *Eichkurve*, falls es sich um eine Gerade handelt: *Eichgerade*.

Eichkurve

Die Eichkurve sollte — wie oben erwähnt — den Konzentrationsbereich abdecken, der auch bei den „Realproben“ vorkommt. Es kann unter Umständen zu erheblichen Fehlern führen, wenn die Eichkurve beispielsweise nur relativ hohe Konzentrationen umfaßt, die Realproben aber bei niedrige Konzentrationen liegen. Das hätte nämlich zur Konsequenz, daß die niedrigen Konzentrationen durch *Extrapolation* der Eichkurve ermittelt werden würden. Dies sollte vermieden werden.

Achtung!
Extrapolation bei
Eichkurve

Die Daten der Messungen, die aus der Serie von Verdünnungen der Stammlösung hergestellt und gemessen wurden, finden sich in Tab. 8.1.

8.4. Scatter Plot

Der Scatterplot wurde schon in Abschnitt 7.3 auf Seite 64 erklärt. Was schon für die Korrelationsrechnung gilt, gilt noch in viel stärkerem Maße für die Regression. Wie in Abschnitt 7.6 auf Seite 68 gezeigt wurde, können auch bei Variablen, die keinerlei sinnvollen Zusammenhang aufweisen, recht hohe Korrelationskoeffizienten auftreten. Selbstverständlich ist auch die Berechnung der Ausgleichsgerade nicht sinnvoll, wenn einer der in Abschnitt 7.6 beschriebenen Fälle zutrifft. Diese Fälle lassen sich — wie schon erwähnt — oft aus den Scatterplots erkennen und ausscheiden. Dies ist auch der Grund, warum der korrekten Anwendung des Scatterplots hohe Bedeutung beigemessen wird.

Zeichnen wir den Scatterplot des Beispiels (Daten in Tab. 8.1), so stellen wir fest, daß die Daten durch eine Gerade gut angenähert werden können. Auch der Korrelationskoeffizient von $r = 0,997$ legt einen linearen Zusammenhang nahe. Abb. 8.2 zeigt den Scatterplot, auch die Ausgleichsgerade ist bereits eingetragen. Wie man diese Gerade berechnet, wird im nächsten Abschnitt im Detail erläutert.

Beispiel

Konzentration	Fläche
2,0	7,1
3,0	10,8
4,0	15,3
5,0	19,9
6,0	24,3
7,0	26,7
8,0	30,8

Tabelle 8.1.: Beispiel für Konzentration und Peakflächen der Eichproben. (Die Konzentrationen und die Fläche sind in beliebigen Einheiten.)

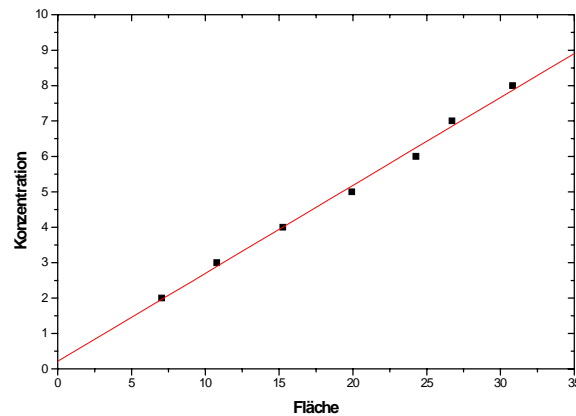


Abbildung 8.2.: Scatterplot der Daten in Tab. 8.1. An der x - Achse sind die aus dem Chromatogramm gemessenen Flächen, an der y - Achse die Konzentrationen der Eichlösungen aufgetragen. Man erkennt sofort, daß die Punkte sehr gut mit einer Ausgleichsgeraden angenähert werden können. ($r = 0,997$)

8.5. Regressionsrechnung

Da es sich, wie gesagt, um ein lineares Modell handelt, wollen wir folgende Gleichung ermitteln:

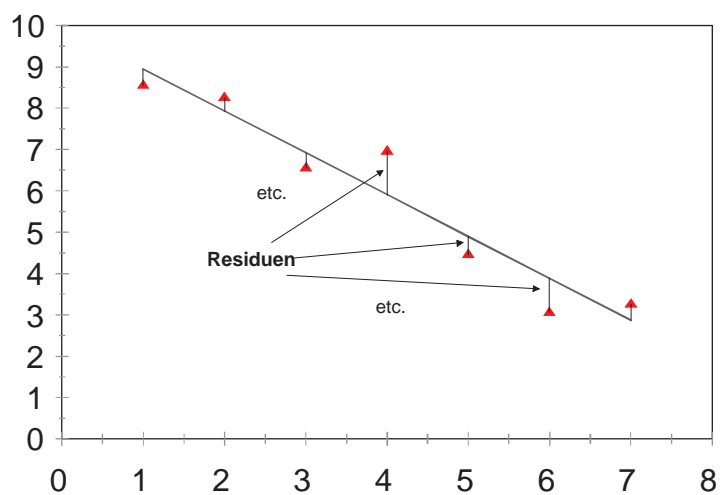


Abbildung 8.3.: Diese Illustration zeigt die Residuen, also die Abweichungen der einzelnen Meßpunkt von einer Geraden.

$$f(x) = kx + d \longrightarrow y = kx + d \quad (8.1)$$

Das ist die „einfache“ Gleichung, die eine Gerade beschreibt. Die Gerade wird durch zwei Parameter, nämlich k und d bestimmt. k gibt die Steigung der Geraden an und d den Schnittpunkt der Geraden mit der y -Achse. Sind diese beiden Parameter bekannt (ist die Gleichung also vollständig bestimmt), kann man sich den zu einem x -Wert gehörenden y -Wert durch einfaches Einsetzen dieses x -Wertes in die Gleichung ermitteln.

Geradengleichung

x und y sind die bekannten Variablen (also die gemessenen Werte). Da eine Gerade ermittelt werden soll, die möglichst optimal diese Daten repräsentiert, müssen die Parameter k und d so errechnet werden, daß eben diese Bedingung erfüllt wird. Das erfolgt üblicherweise nach dem Prinzip der *Minimierung der Summe der Fehlerquadrate*. Der englische Ausdruck lautet: *least sum of squared residuals*. Unter Residuen versteht man die Abweichungen der Meßpunkte von der Geraden. Dies wird in Abb. 8.3 illustriert.

least sum of squared residuals

Für jede mögliche Gerade kann man diese Residuen leicht bestimmen. Diese werden quadriert (u.a. um nur positive Werte zu erhalten) und summiert. Die Annahme ist, daß die Gerade optimal paßt, wenn diese Summe — die wir oben als Summe der Fehlerquadrate bezeichnet haben — minimal wird. Man versucht also die Gerade so zu legen, daß — salopp formuliert — der „Gesamtfehler“ minimal wird. Folgt man dieser Idee, kann man die beiden Parameter k_{min} und d_{min} ableiten [5] (Das Verständnis der Ableitung ist für die praktische Anwendung nicht unmittelbar erforderlich):

Ableitung

Ableitung:

$$e_i = kx_i + d - y_i, \quad i = 1, 2, \dots, n \quad (8.2)$$

sind die Residuen. Sind die *Residuen normalverteilt*, so kann man zeigen, daß die Gerade sich den Meßpunkten am besten anpaßt, wenn eben die Summe der quadrierten Residuen ein *Minimum* annimmt. Gesucht ist also das Minimum der Funktion f :

$$f(k, d) = \sum_{i=1}^n (kx_i + d - y_i)^2 \quad k, d \in R \quad (8.3)$$

Zunächst werden die relativen Extrema der Funktion f durch partielle Ableitung nach k und d bestimmt

$$\frac{\partial f}{\partial k} = 2 \sum_{i=1}^n ((kx_i + d - y_i)x_i) = 0 \quad \text{und} \quad (8.4)$$

$$\frac{\partial f}{\partial d} = 2 \sum_{i=1}^n (kx_i + d - y_i) = 0 \quad (8.5)$$

Um das Gleichungssystem mit Hilfe der *Cramerschen* Regel zu bestimmen, muß zunächst die Determinante errechnet werden:

$$D = n \sum_{i=1}^n x_i - \left(\sum_{i=1}^n x_i \right)^2 = \frac{1}{2} \sum_{i,k=1}^n (x_i - x_k)^2 \quad (8.6)$$

Das Gleichungssystem besitzt (logischerweise) genau eine Lösung, wenn mindestens zwei verschiedene Meßwerte von x_i vorliegen. Diese Lösung wird mit k_{min} und d_{min} bezeichnet:

$$k_{min} = \frac{1}{D} \cdot \begin{vmatrix} \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i & n \end{vmatrix} \quad (8.7)$$

$$d_{min} = \frac{1}{D} \cdot \begin{vmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n y_i \end{vmatrix} \quad (8.8)$$

Ein relatives Minimum liegt vor, da $f_{aa} \cdot f_{bb} - f_{ab}^2 > 0$ und $f_{aa}(k_{min}, d_{min}) > 0$. Die Regressionsgerade, also das gesuchte lineare Modell, hat dann die Form:

$$y = k_{min}x + d_{min} \quad (8.9)$$

Parameter k, d Das ist dasjenige Parameterpaar k, d , bei dem eben diese Summe der Fehlerquadrate minimal ist :

$$k_{min} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (8.10)$$

$$d_{min} = \frac{\sum y_i - k_{min} \sum x_i}{n} \quad (8.11)$$

und die Ausgleichsgerade ist folglich:

$$y = k_{min}x + d_{min} \quad (8.12)$$

k steht für die *Steigung* der Geraden, d ist der *Offset*, d.h. gibt den Punkt an, an dem die Gerade die y -Achse schneidet.

Beispiel Berechnen wir nun die Ausgleichsgerade der Daten in Tab. 8.1. Die unabhängige Variable ist die *Fläche*, die abhängige Variable (also diejenige die wir später mit dem Modell bestimmen wollen) ist die *Konzentration*. Folglich bezeichnen wir die Fläche mit x und die Konzentration mit y . Berechnen wir zunächst k_{min} :

$$k_{min} = \frac{7 \cdot (14,1 + 32,3 + 61,0 + 99,7 + 145,7 + 187,0 + 246,6) - 134,9 \cdot 35}{7 \cdot 3049,17 - 18198,01} \quad (8.13)$$

daraus folgt

$$k_{min} = \underline{\underline{0,24897}} \quad (8.14)$$

als nächstes errechnen wir d_{min}

$$d_{min} = \frac{35 - 0,248969 \cdot 134,9}{7} = \underline{\underline{0,2020}} \quad (8.15)$$

d.h. die Ausgleichsgerade, also die Eichgerade ergibt sich durch folgende Gleichung:

$$\text{Konzentration} = 0,24897 \cdot \text{Fläche} + 0,202 \quad (8.16)$$

8.6. Analyse der Residuen

Wie in der Einleitung erwähnt, ist auch die Anwendung der Regressionsrechnung an verschiedene Bedingungen geknüpft. Verschiedene Fehlerquellen wurden schon im Kapitel über die Korrelationsrechnung erwähnt (siehe im besonderen Abschnitt 7.6). Ein weiteres Hilfsmittel um „problematische“ Daten zu erkennen ist die *Analyse der Residuen* und wird in diesem Abschnitt kurz vorgestellt.

Betrachten wir die Gleichung der Ausgleichsgerade (siehe Gleichung 8.12), so müssen wir feststellen, daß sie eigentlich nicht ganz vollständig ist. Die gemessenen Daten liegen ja üblicherweise nicht *exakt* auf der Ausgleichsgeraden, sondern streuen in einem gewissen Rahmen. Diese Abweichung der Meßwerte von der Ausgleichsgeraden haben wir schon als Residuen bezeichnet. Wir haben bereits kurz in den vorigen Abschnitten davon gehört (Abb. 8.3 stellt die Residuen graphisch dar).

D.h. um die Gleichung zu vervollständigen, müssen die Residuen noch in die Gleichung mitaufgenommen werden (dies wurde auch schon in der Ableitung dargestellt, siehe z.B. Gleichung 8.2). Vollständig angeschrieben lautet die Gleichung der Ausgleichsgeraden daher:

$$y_i = kx_i + d + \varepsilon_i \quad (8.17)$$

Oder mit anderen Worten: Der i -te Wert der abhängigen Variable (y) ergibt sich aus der Multiplikation von k mit dem i -ten Wert der unabhängigen Variable (x) und Addition des „Offsets“ d . Als „Rest“ bleibt die Abweichung des gemessenen Wertes y_i von der Geraden. Dieser Rest ist das i -te Residuum ε_i und kommt normalerweise durch verschiedene Störungen wie Gerätauschen zustande und kann daher nicht berechnet werden³.

Damit die Berechnung der Ausgleichsgeraden durch Minimierung der Fehlerquadrate ein korrektes Ergebnis liefert, müssen folgende Bedingungen erfüllt sein:

³Dieser zufällige Aspekt ist sehr wichtig zu verstehen. Würde es sich um einen *berechenbaren* Faktor handeln, so sollte man ihn in das Modell miteinbeziehen und eben berechnen. Gewisse Unsicherheiten, Meßfehler, Rauschen etc. sind aber bei realen Daten immer zu finden und als solche unbestimmbar, trotzdem lassen sich gewisse Eigenschaften erkennen. Bspw. ist der Mittelwert des Rauschens meist gleich null.

$y = k_{min}x + d_{min}$
nicht vollständig?

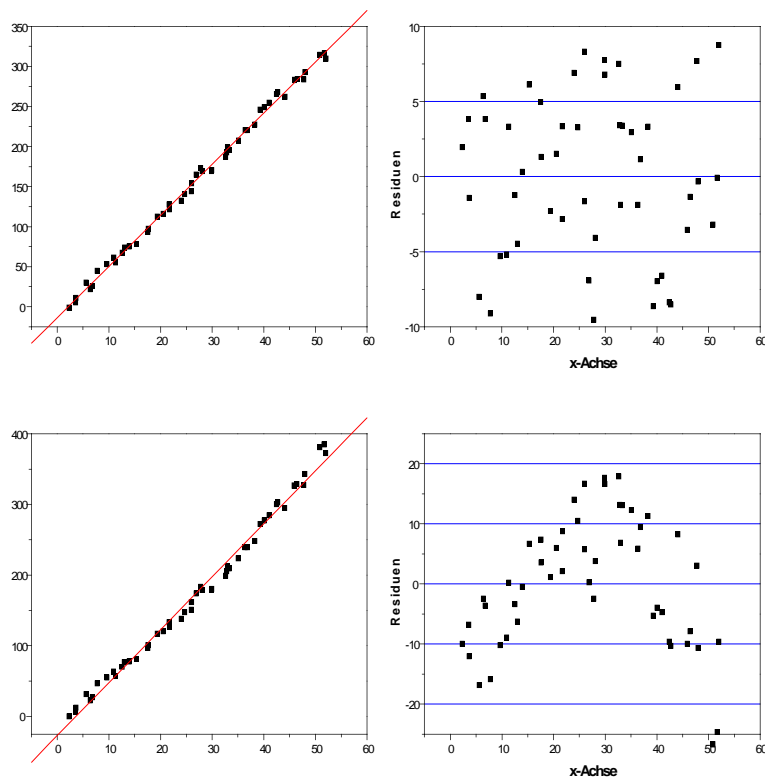


Abbildung 8.4.: Das obere Beispiel zeigt eine „korrekte“ Regression, im untere Beispiel liegt eine nicht-lineare Abhängigkeit der Daten vor. Eine *lineare* Regression wäre hier fehl am Platz!

- Der Mittelwert der Residuen ε_i muß gleich Null sein, mit der Standardabweichung s .
- s sollte annähernd normalverteilt sein.
- ε_m und ε_n mit $m \neq n$ (d.h. zwei beliebig herausgegriffene Residuen) sollten unabhängig voneinander sein.

Diese Bedingungen ergeben sich klar aus der Ableitung der Ausgleichsgeraden. Die Hauptfrage, die sich dem Praktiker stellt, ist, wie man möglichst einfach überprüfen kann, ob diese Bedingungen erfüllt sind. Eine schnelle und anschauliche Möglichkeit ist die graphische Analyse der Residuen. Man geht wie folgt vor:

graphische
Analyse

1. Berechnen der Ausgleichsgerade.
2. Berechnen der Residuen. D.h. $\varepsilon_i = y_{\text{gemessen}} - y_{\text{berechnet}}$
3. Zeichnen eines Scatterplots, wobei auf der x -Achse die abhängige oder unabhängige Variable (also x oder y) aufgetragen wird, auf der y -Achse hinge-

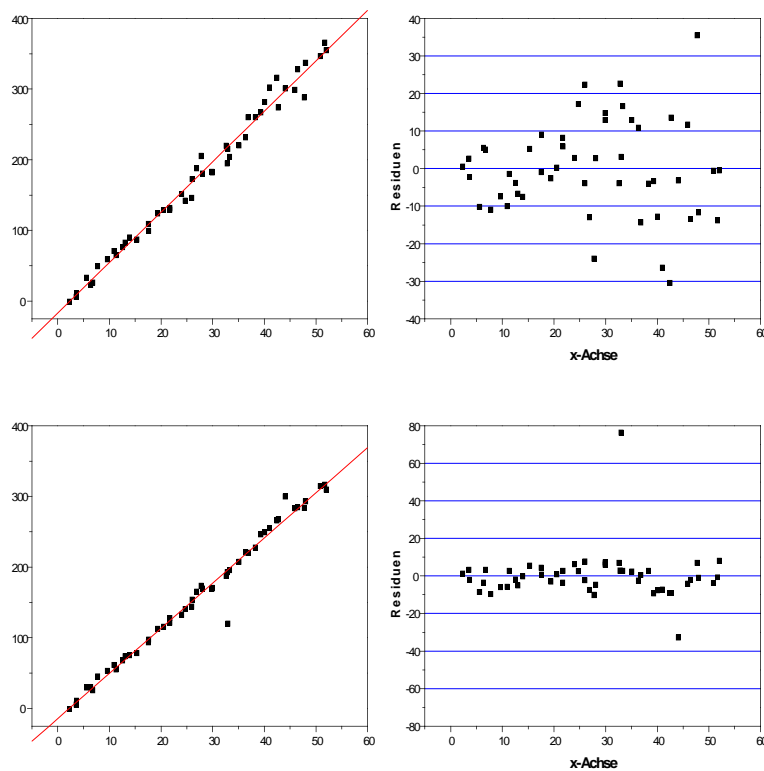


Abbildung 8.5.: Die Daten im oberen Beispiel zeigen nicht konstante Varianz der Residuen, im unteren Beispiel erkennt man in der Analyse der Residuen sofort vorliegende Ausreißer.

gen die Größe der Abweichung, also die Residuen. Ob man x oder y auf der x -Achse aufträgt ist für das Ergebnis unerheblich und daher beliebig.

Die Abbildungen 8.4 und 8.5 zeigen vier wichtige Fälle⁴: Das Beispiel in Abbildungen 8.4 oben zeigt eine korrekte Regression⁵. Die Residuen sind dementsprechend homogen. Das untere Beispiel zeigt einen Fall, wo die Daten nicht gut durch eine Gerade angenähert werden können. In der Abbildung, die die Regression darstellt ist dies noch nicht so stark zu bemerken. An den Residuen sieht man jedoch sofort, daß diese doch starke Inhomogenitäten zeigen.

Abbildung 8.5 oben zeigt den Fall, daß die Varianz (Standardabweichung) der Residuen nicht gleichbleibt, sondern sukzessive zunimmt. Man sagt auch: es liegt *Heteroskedastizität* vor. Das Gegenteil — also die Konstanz der Varianz — wird auch *Homoskedastizität* genannt. In diesem Beispiel erkennt man anhand der Residuen

⁴Nicht die **absolute Größe** der Residuen ist von Bedeutung, sondern vielmehr die Homogenität der Verteilung. Die Größe ist Abhängig von der Intensität des Rauschens. Auch Daten mit intensiverem Rauschen, als unser Beispiel zeigt, können gut durch lineare Regression beschrieben werden.

⁵Man beachte, daß die Skalierungen der Residuen-Achsen sich bei den verschiedenen Beispielen unterscheiden!

sofort: Bis zu einem x -Wert von etwa 25 bleiben die Residuen in einem Bereich zwischen -10 und 10, ab etwa 25 kann man ein deutliches Ansteigen bemerken. Im unteren Beispiel sind Ausreißer vorhanden. Auch diese sind im Plot der Residuen leicht zu erkennen.

Konsequenzen

Es ist wichtig diese Voraussetzungen zu beachten. Vor allem **Ausreißer** können das Ergebnis der Regressionsrechnung massiv verfälschen. Abb 8.6 zeigt den schon bekannten Datensatz aus Tab. 8.1 auf Seite 75, also die Daten der Eichgerade. Im linken Beispiel wurde ein mittlerer Wert durch einen Ausreißer ersetzt. Dies bewirkt, daß die Ausgleichsgerade nicht unbeträchtlich nach oben „gezogen“ wird. Es ist also in diesem Fall damit zu rechnen das v.a. der Parameter d einen erheblichen Fehler aufweisen wird. Im rechten Beispiel wurde der äußerste rechte Wert durch einen Ausreißer ersetzt. Auch hier erkennt man sofort, wie die Ausgleichsgerade dadurch verändert wird. In diesem Fall werden sich beide Parameter erheblich verändern.

Man sieht schon an diesen beiden Beispielen, daß unter Umständen schon **ein einziger Ausreißer ausreicht, um die Ausgleichsgerade massiv zu verfälschen.**

8.7. $A = kB + d$ und $B = kA + d$

Angenommen, es liegt eine Meßserie mit zwei Variablen A und B vor. Ein linearer Zusammenhang wird aus dem Scatterplot angenommen und die Ausgleichsgerade soll bestimmt werden. Es ist wichtig zu verstehen, daß sich die Parameter k und d unterscheiden, je nachdem ob als unabhängige Variable (x) A oder B gewählt wird. Etwas mathematischer formuliert lauten beiden Gleichungen dann

$$A = k_1 B + d_1 \quad (8.18)$$

und

$$B = k_2 A + d_2 \quad (8.19)$$

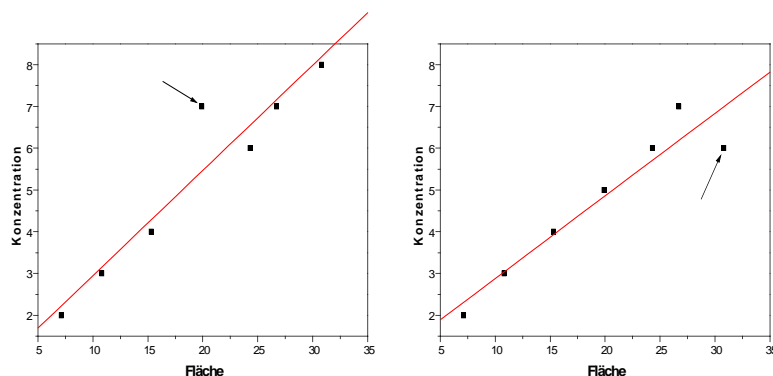


Abbildung 8.6.: Diese Abbildung zeigt, wie stark schon einzelne Ausreißer die Ausgleichsgerade verändern können!

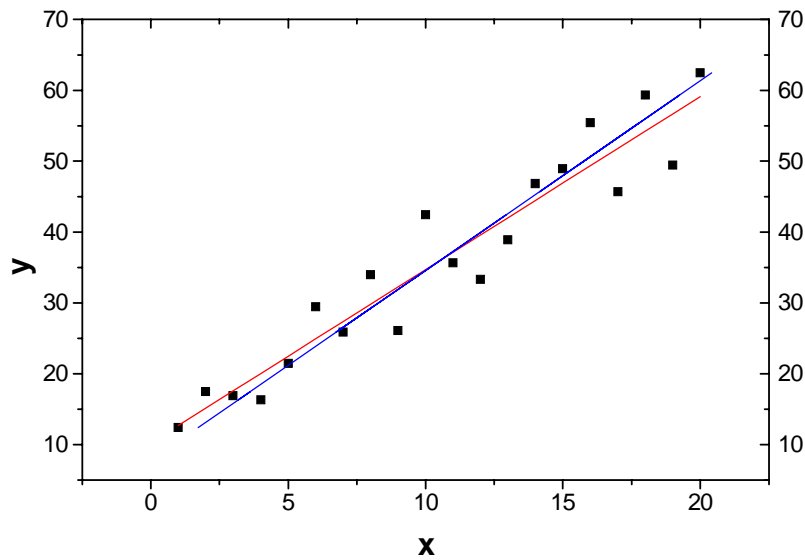


Abbildung 8.7.: Graphische Darstellung der beiden im Text beschriebenen Regressionsgeraden: Da der Winkel zwischen den Geraden recht klein ist, kann man auf eine gute Korrelation schließen ($r = 0,955$).

Die Parameter k_1 und k_2 sowie d_1 und d_2 sind **nicht gleich**:

$$k_1 \neq k_2 \quad \text{und} \quad d_1 \neq d_2 \quad (8.20)$$

das gilt, wie gesagt, obwohl die Daten in beiden Fällen dieselben sind. **Das Vertauschen der Variablen ist nicht einfach eine Umkehrung der Regressionsrechnung!** Denn im einen Fall werden die *vertikalen Abweichungen* der Meßwerte von der Geraden minimiert ($A = kB + d$) und im anderen Fall die *horizontalen Abweichungen* der Meßwerte von der Geraden ($B = kA + d$).

Man begeht daher einen unter Umständen erheblichen Fehler, wenn man die Parameter k und d aus der Gleichung $A = kB + d$ berechnet und dann die Gleichung zu $B = (A - d)/k$ umformt um B aus A zu berechnen!

In einem „Sonderfall“ gilt jedoch $k_1 = k_2$ und $d_1 = d_2$, nämlich genau dann, wenn ein *funktionaler Zusammenhang* zwischen den Variablen A und B besteht (siehe auch Abschnitt 7.4 auf Seite 65). Mit anderen Worten: Dies gilt **nur** dann, wenn alle Punkte *exakt* auf der Geraden liegen⁶.

Diese Zusammenhänge kann man auch zur Graphischen Interpretation des Korrelationskoeffizienten nutzen. Zeichnet man in einen Scatterplot sowohl die Regressionsgerade $A = kB + d$ als auch $B = kA + d$, so ist der Winkel zwischen diesen beiden Geraden ein Maß für die Güte der Korrelation. Dieser Zusammenhang ist in Abb. 8.7

Umformen der Gleichung

Korrelation graphisch

⁶Denn nur dann sind ja auch die vertikalen und die horizontalen Abweichungen beide gleich, nämlich null.

illustriert. Im Beispiel ist die Korrelation eine ziemlich gute ($r = 0,955$) und die beiden Geraden sind einander recht ähnlich. Trotzdem ist ein Unterschied deutlich festzustellen und unterstreicht das oben erwähnte Faktum, daß ein einfaches Vertauschen der Variablen nicht zulässig ist, und daß im Falle des Falles für beide Möglichkeiten getrennt eine Regression zu rechnen ist.

In den Extremfällen bedeutet das: Stehen die Geraden in einem rechten Winkel zueinander, so sind die Daten völlig unkorreliert, fallen die Geraden zusammen so ist $|r| = 1$. Im Falle unseres Beispiels der Eichgeraden ist die Korrelation bereits so gut, daß die beiden Geraden in einem Plot kaum zu unterscheiden wären.

8.8. Ergebnis/Zusammenfassung

Die Ergebnisse der Rechnungen in den vorigen Abschnitten kann man wie folgt zusammenfassen:

1. Zunächst werden Eichlösungen im Konzentrationsbereich der zu messenden Proben bereitet und mittels HPLC gemessen.
2. Die aus dem Chromatogramm gemessenen Flächen werden in einem Scatterplot den Konzentrationen gegenübergestellt.
3. Der Korrelationskoeffizient wird errechnet ($r = 0,997$). Aus Scatterplot und Korrelationskoeffizient kann man leicht folgern, daß als Eichkurve eine Gerade anzuwenden ist.
4. Zur Bestätigung der Voraussetzungen, die für die Berechnung einer Ausgleichsgerade erforderlich sind, sollten noch die Residuen analysiert werden. Die Analyse der Residuen ist an sich in jedem Fall empfehlenswert, da viele Abweichungen von den Voraussetzungen damit sehr empfindlich detektiert werden können. Im konkreten Beispiel wurde allerdings darauf verzichtet, da nur sieben Datenpunkte vorliegen, und da die Aussagekraft recht gering ist.
5. Die Parameter k und d der Ausgleichsgerade werden durch Minimierung der Fehlerquadrate errechnet und ergeben die Gleichung⁷: $\text{Konzentration} = 0,248 \cdot \text{Fläche} + 0,217$

Um nun aus einem neuen Meßwert die Konzentration zu errechnen, bestimmt man zunächst die Peakfläche und setzt diese in die Gleichung ein. Wäre die Fläche z.B. 23,48 so errechnet sich die Konzentration nach:

$$\text{Konzentration} = 0,249 \cdot 23,48 + 0,202 = 6,05 \quad (8.21)$$

Achtung! Es soll an dieser Stelle nochmals darauf hingewiesen werden, daß bei Anwen-

⁷Aus dem Scatterplot mit eingetragener Ausgleichsgerade erkennt man auch, daß die Gerade fast durch den Nullpunkt geht. Dies ist auch intuitiv verständlich: Eine Peakfläche von 0 bedeutet natürlich auch, daß die Konzentration der Substanz unter der Nachweisgrenze liegt, salopp ausgedrückt: ebenfalls 0 ist. Die Tatsache, daß die Gerade nicht *exakt* durch den Nullpunkt geht, ist durch immer vorhandene Störungen (Rauschen, ...) der Methode erklärbar.

derung der linearen Regression auf Basis der Minimierung der Summe der quadrierten Residuen unbedingt darauf zu achten ist, daß die erwähnten Voraussetzungen erfüllt sind. Also vor allem: **Die Residuen sollten homogen und normalverteilt sein und es sollten keine Ausreißer vorliegen.**

Bei Nichtbeachtung der Voraussetzungen der Regression kann die Ausgleichsgerade deutlich von den erwarteten Werten abweichen, da bspw. schon einzelne Ausreißer die Gerade deutlich beeinflussen können.

A. Tabellenanhang

FG	90 %	95 %	99 %
1	6,314	12,706	63,657
2	2,920	4,303	9,925
3	2,353	3,182	5,841
4	2,132	2,776	4,604
5	2,015	2,571	4,032
6	1,943	2,447	3,707
7	1,895	2,365	3,499
8	1,860	2,306	3,355
9	1,833	2,262	3,250
10	1,812	2,228	3,169
11	1,796	2,201	3,106
12	1,782	2,179	3,055
13	1,771	2,160	3,012
14	1,761	2,145	2,977
15	1,753	2,131	2,947
16	1,746	2,120	2,921
17	1,740	2,110	2,898
18	1,734	2,101	2,878
19	1,729	2,093	2,861
20	1,725	2,086	2,845
21	1,721	2,080	2,831
22	1,717	2,074	2,819
23	1,714	2,069	2,807
24	1,711	2,064	2,797
25	1,708	2,060	2,787
30	1,697	2,042	2,750
40	1,684	2,021	2,704
50	1,676	2,009	2,678
60	1,671	2,000	2,660
70	1,667	1,994	2,648
80	1,664	1,990	2,639
90	1,662	1,987	2,632
100	1,660	1,984	2,626
200	1,653	1,972	2,601
500	1,648	1,965	2,586
1000	1,646	1,962	2,581
∞	1,645	1,960	2,576

Tabelle A.1.: Student t Verteilung aus [21]. FG steht für Freiheitsgrad — wird im Text auch mit f bezeichnet.

n	P = 0,90	P = 0,95	P = 0,99
3	0,89	0,94	0,99
4	0,68	0,77	0,89
5	0,56	0,64	0,76
6	0,48	0,56	0,70
7	0,43	0,51	0,64
8	0,40	0,48	0,58

Tabelle A.2.: Q-Werte für Dean und Dixon Ausreißertest.

P = 0,90	f₁							
f₂	2	3	4	5	6	7	8	9
1	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86
2	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38
3	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24
4	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94
5	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32
6	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96
7	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72
8	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56
9	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44
10	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35
11	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27
12	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21
13	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16
14	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12
15	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09
16	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06
17	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03
18	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00
19	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98
20	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96
25	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89
30	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85
35	2,46	2,25	2,11	2,02	1,95	1,90	1,85	1,82
40	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79
45	2,42	2,21	2,07	1,98	1,91	1,85	1,81	1,77
50	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76
60	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74
70	2,38	2,16	2,03	1,93	1,86	1,80	1,76	1,72
80	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71
90	2,36	2,15	2,01	1,91	1,84	1,78	1,74	1,70
100	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69
200	2,33	2,11	1,97	1,88	1,80	1,75	1,70	1,66
500	2,31	2,09	1,96	1,86	1,79	1,73	1,68	1,64
1000	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64

Tabelle A.3.: F-Test, P = 0,90; alle F-Werte wurden mit Quattro Pro errechnet

P = 0,90	f₁							
	f₂	10	11	12	13	14	15	16
1	60,19	60,47	60,71	60,90	61,07	61,22	61,35	61,46
2	9,39	9,40	9,41	9,41	9,42	9,42	9,43	9,43
3	5,23	5,22	5,22	5,21	5,20	5,20	5,20	5,19
4	3,92	3,91	3,90	3,89	3,88	3,87	3,86	3,86
5	3,30	3,28	3,27	3,26	3,25	3,24	3,23	3,22
6	2,94	2,92	2,90	2,89	2,88	2,87	2,86	2,85
7	2,70	2,68	2,67	2,65	2,64	2,63	2,62	2,61
8	2,54	2,52	2,50	2,49	2,48	2,46	2,45	2,45
9	2,42	2,40	2,38	2,36	2,35	2,34	2,33	2,32
10	2,32	2,30	2,28	2,27	2,26	2,24	2,23	2,22
11	2,25	2,23	2,21	2,19	2,18	2,17	2,16	2,15
12	2,19	2,17	2,15	2,13	2,12	2,10	2,09	2,08
13	2,14	2,12	2,10	2,08	2,07	2,05	2,04	2,03
14	2,10	2,07	2,05	2,04	2,02	2,01	2,00	1,99
15	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,95
16	2,03	2,01	1,99	1,97	1,95	1,94	1,93	1,92
17	2,00	1,98	1,96	1,94	1,93	1,91	1,90	1,89
18	1,98	1,95	1,93	1,92	1,90	1,89	1,87	1,86
19	1,96	1,93	1,91	1,89	1,88	1,86	1,85	1,84
20	1,94	1,91	1,89	1,87	1,86	1,84	1,83	1,82
25	1,87	1,84	1,82	1,80	1,79	1,77	1,76	1,75
30	1,82	1,79	1,77	1,75	1,74	1,72	1,71	1,70
35	1,79	1,76	1,74	1,72	1,70	1,69	1,67	1,66
40	1,76	1,74	1,71	1,70	1,68	1,66	1,65	1,64
45	1,74	1,72	1,70	1,68	1,66	1,64	1,63	1,62
50	1,73	1,70	1,68	1,66	1,64	1,63	1,61	1,60
60	1,71	1,68	1,66	1,64	1,62	1,60	1,59	1,58
70	1,69	1,66	1,64	1,62	1,60	1,59	1,57	1,56
80	1,68	1,65	1,63	1,61	1,59	1,57	1,56	1,55
90	1,67	1,64	1,62	1,60	1,58	1,56	1,55	1,54
100	1,66	1,64	1,61	1,59	1,57	1,56	1,54	1,53
200	1,63	1,60	1,58	1,56	1,54	1,52	1,51	1,49
500	1,61	1,58	1,56	1,54	1,52	1,50	1,49	1,47
1000	1,61	1,58	1,55	1,53	1,51	1,49	1,48	1,46

Tabelle A.4.: F-Test, P = 0,90

P = 0,90	f₁							
	f₂	18	19	20	25	30	40	50
1	61,57	61,66	61,74	62,05	62,26	62,53	62,69	63,01
2	9,44	9,44	9,44	9,45	9,46	9,47	9,47	9,48
3	5,19	5,19	5,18	5,17	5,17	5,16	5,15	5,14
4	3,85	3,85	3,84	3,83	3,82	3,80	3,80	3,78
5	3,22	3,21	3,21	3,19	3,17	3,16	3,15	3,13
6	2,85	2,84	2,84	2,81	2,80	2,78	2,77	2,75
7	2,61	2,60	2,59	2,57	2,56	2,54	2,52	2,50
8	2,44	2,43	2,42	2,40	2,38	2,36	2,35	2,32
9	2,31	2,30	2,30	2,27	2,25	2,23	2,22	2,19
10	2,22	2,21	2,20	2,17	2,16	2,13	2,12	2,09
11	2,14	2,13	2,12	2,10	2,08	2,05	2,04	2,01
12	2,08	2,07	2,06	2,03	2,01	1,99	1,97	1,94
13	2,02	2,01	2,01	1,98	1,96	1,93	1,92	1,88
14	1,98	1,97	1,96	1,93	1,91	1,89	1,87	1,83
15	1,94	1,93	1,92	1,89	1,87	1,85	1,83	1,79
16	1,91	1,90	1,89	1,86	1,84	1,81	1,79	1,76
17	1,88	1,87	1,86	1,83	1,81	1,78	1,76	1,73
18	1,85	1,84	1,84	1,80	1,78	1,75	1,74	1,70
19	1,83	1,82	1,81	1,78	1,76	1,73	1,71	1,67
20	1,81	1,80	1,79	1,76	1,74	1,71	1,69	1,65
25	1,74	1,73	1,72	1,68	1,66	1,63	1,61	1,56
30	1,69	1,68	1,67	1,63	1,61	1,57	1,55	1,51
35	1,65	1,64	1,63	1,60	1,57	1,53	1,51	1,47
40	1,62	1,61	1,61	1,57	1,54	1,51	1,48	1,43
45	1,60	1,59	1,58	1,55	1,52	1,48	1,46	1,41
50	1,59	1,58	1,57	1,53	1,50	1,46	1,44	1,39
60	1,56	1,55	1,54	1,50	1,48	1,44	1,41	1,36
70	1,55	1,54	1,53	1,49	1,46	1,42	1,39	1,34
80	1,53	1,52	1,51	1,47	1,44	1,40	1,38	1,32
90	1,52	1,51	1,50	1,46	1,43	1,39	1,36	1,30
100	1,52	1,50	1,49	1,45	1,42	1,38	1,35	1,29
200	1,48	1,47	1,46	1,41	1,38	1,34	1,31	1,24
500	1,46	1,45	1,44	1,39	1,36	1,31	1,28	1,21
1000	1,45	1,44	1,43	1,38	1,35	1,30	1,27	1,20

Tabelle A.5.: F-Test, P = 0,90

P = 0,95	f₁							
	f₂	2	3	4	5	6	7	8
1	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
25	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
30	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
35	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16
40	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
45	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10
50	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07
60	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
70	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02
80	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00
90	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99
100	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97
200	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93
500	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90
1000	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89

Tabelle A.6.: F-Test, P = 0,95

f₂	f₁							
	10	11	12	13	14	15	16	17
1	241,9	243,0	244,0	244,7	245,4	246,0	246,5	247,0
2	19,40	19,40	19,41	19,42	19,42	19,43	19,43	19,44
3	8,79	8,76	8,74	8,73	8,71	8,70	8,69	8,68
4	5,96	5,94	5,91	5,89	5,87	5,86	5,84	5,83
5	4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,59
6	4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,91
7	3,64	3,60	3,57	3,55	3,53	3,51	3,49	3,48
8	3,35	3,31	3,28	3,26	3,24	3,22	3,20	3,19
9	3,14	3,10	3,07	3,05	3,03	3,01	2,99	2,97
10	2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81
11	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,69
12	2,75	2,72	2,69	2,66	2,64	2,62	2,60	2,58
13	2,67	2,63	2,60	2,58	2,55	2,53	2,51	2,50
14	2,60	2,57	2,53	2,51	2,48	2,46	2,44	2,43
15	2,54	2,51	2,48	2,45	2,42	2,40	2,38	2,37
16	2,49	2,46	2,42	2,40	2,37	2,35	2,33	2,32
17	2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,27
18	2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,23
19	2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,20
20	2,35	2,31	2,28	2,25	2,22	2,20	2,18	2,17
25	2,24	2,20	2,16	2,14	2,11	2,09	2,07	2,05
30	2,16	2,13	2,09	2,06	2,04	2,01	1,99	1,98
35	2,11	2,07	2,04	2,01	1,99	1,96	1,94	1,92
40	2,08	2,04	2,00	1,97	1,95	1,92	1,90	1,89
45	2,05	2,01	1,97	1,94	1,92	1,89	1,87	1,86
50	2,03	1,99	1,95	1,92	1,89	1,87	1,85	1,83
60	1,99	1,95	1,92	1,89	1,86	1,84	1,82	1,80
70	1,97	1,93	1,89	1,86	1,84	1,81	1,79	1,77
80	1,95	1,91	1,88	1,84	1,82	1,79	1,77	1,75
90	1,94	1,90	1,86	1,83	1,80	1,78	1,76	1,74
100	1,93	1,89	1,85	1,82	1,79	1,77	1,75	1,73
200	1,88	1,84	1,80	1,77	1,74	1,72	1,69	1,67
500	1,85	1,81	1,77	1,74	1,71	1,69	1,66	1,64
1000	1,84	1,80	1,76	1,73	1,70	1,68	1,65	1,63

Tabelle A.7.: F-Test, P = 0,95

P = 0,95	f₁							
	f₂	18	19	20	25	30	40	50
1	247,3	247,7	248,0	249,3	250,1	251,1	251,8	253,0
2	19,44	19,44	19,45	19,46	19,46	19,47	19,48	19,49
3	8,67	8,67	8,66	8,63	8,62	8,59	8,58	8,55
4	5,82	5,81	5,80	5,77	5,75	5,72	5,70	5,66
5	4,58	4,57	4,56	4,52	4,50	4,46	4,44	4,41
6	3,90	3,88	3,87	3,83	3,81	3,77	3,75	3,71
7	3,47	3,46	3,44	3,40	3,38	3,34	3,32	3,27
8	3,17	3,16	3,15	3,11	3,08	3,04	3,02	2,97
9	2,96	2,95	2,94	2,89	2,86	2,83	2,80	2,76
10	2,80	2,79	2,77	2,73	2,70	2,66	2,64	2,59
11	2,67	2,66	2,65	2,60	2,57	2,53	2,51	2,46
12	2,57	2,56	2,54	2,50	2,47	2,43	2,40	2,35
13	2,48	2,47	2,46	2,41	2,38	2,34	2,31	2,26
14	2,41	2,40	2,39	2,34	2,31	2,27	2,24	2,19
15	2,35	2,34	2,33	2,28	2,25	2,20	2,18	2,12
16	2,30	2,29	2,28	2,23	2,19	2,15	2,12	2,07
17	2,26	2,24	2,23	2,18	2,15	2,10	2,08	2,02
18	2,22	2,20	2,19	2,14	2,11	2,06	2,04	1,98
19	2,18	2,17	2,16	2,11	2,07	2,03	2,00	1,94
20	2,15	2,14	2,12	2,07	2,04	1,99	1,97	1,91
25	2,04	2,02	2,01	1,96	1,92	1,87	1,84	1,78
30	1,96	1,95	1,93	1,88	1,84	1,79	1,76	1,70
35	1,91	1,89	1,88	1,82	1,79	1,74	1,70	1,63
40	1,87	1,85	1,84	1,78	1,74	1,69	1,66	1,59
45	1,84	1,82	1,81	1,75	1,71	1,66	1,63	1,55
50	1,81	1,80	1,78	1,73	1,69	1,63	1,60	1,52
60	1,78	1,76	1,75	1,69	1,65	1,59	1,56	1,48
70	1,75	1,74	1,72	1,66	1,62	1,57	1,53	1,45
80	1,73	1,72	1,70	1,64	1,60	1,54	1,51	1,43
90	1,72	1,70	1,69	1,63	1,59	1,53	1,49	1,41
100	1,71	1,69	1,68	1,62	1,57	1,52	1,48	1,39
200	1,66	1,64	1,62	1,56	1,52	1,46	1,41	1,32
500	1,62	1,61	1,59	1,53	1,48	1,42	1,38	1,28
1000	1,61	1,60	1,58	1,52	1,47	1,41	1,36	1,26

Tabelle A.8.: F-Test, P = 0,95

P = 0,99	f₁							
	f₂	2	3	4	5	6	7	8
1	5000	5403	5625	5764	5859	5928	5981	6022
2	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66
5	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46
25	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22
30	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
35	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96
40	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89
45	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83
50	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78
60	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
70	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67
80	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64
90	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61
100	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59
200	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50
500	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44
1000	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43

Tabelle A.9.: F-Test, P = 0,99

P = 0,99	f₁							
	f₂	10	11	12	13	14	15	16
1	6056	6083	6106	6126	6143	6157	6170	6181
2	99,40	99,41	99,42	99,42	99,43	99,43	99,44	99,44
3	27,23	27,13	27,05	26,98	26,92	26,87	26,83	26,79
4	14,55	14,45	14,37	14,31	14,25	14,20	14,15	14,11
5	10,05	9,96	9,89	9,82	9,77	9,72	9,68	9,64
6	7,87	7,79	7,72	7,66	7,60	7,56	7,52	7,48
7	6,62	6,54	6,47	6,41	6,36	6,31	6,28	6,24
8	5,81	5,73	5,67	5,61	5,56	5,52	5,48	5,44
9	5,26	5,18	5,11	5,05	5,01	4,96	4,92	4,89
10	4,85	4,77	4,71	4,65	4,60	4,56	4,52	4,49
11	4,54	4,46	4,40	4,34	4,29	4,25	4,21	4,18
12	4,30	4,22	4,16	4,10	4,05	4,01	3,97	3,94
13	4,10	4,02	3,96	3,91	3,86	3,82	3,78	3,75
14	3,94	3,86	3,80	3,75	3,70	3,66	3,62	3,59
15	3,80	3,73	3,67	3,61	3,56	3,52	3,49	3,45
16	3,69	3,62	3,55	3,50	3,45	3,41	3,37	3,34
17	3,59	3,52	3,46	3,40	3,35	3,31	3,27	3,24
18	3,51	3,43	3,37	3,32	3,27	3,23	3,19	3,16
19	3,43	3,36	3,30	3,24	3,19	3,15	3,12	3,08
20	3,37	3,29	3,23	3,18	3,13	3,09	3,05	3,02
25	3,13	3,06	2,99	2,94	2,89	2,85	2,81	2,78
30	2,98	2,91	2,84	2,79	2,74	2,70	2,66	2,63
35	2,88	2,80	2,74	2,69	2,64	2,60	2,56	2,53
40	2,80	2,73	2,66	2,61	2,56	2,52	2,48	2,45
45	2,74	2,67	2,61	2,55	2,51	2,46	2,43	2,39
50	2,70	2,63	2,56	2,51	2,46	2,42	2,38	2,35
60	2,63	2,56	2,50	2,44	2,39	2,35	2,31	2,28
70	2,59	2,51	2,45	2,40	2,35	2,31	2,27	2,23
80	2,55	2,48	2,42	2,36	2,31	2,27	2,23	2,20
90	2,52	2,45	2,39	2,33	2,29	2,24	2,21	2,17
100	2,50	2,43	2,37	2,31	2,27	2,22	2,19	2,15
200	2,41	2,34	2,27	2,22	2,17	2,13	2,09	2,06
500	2,36	2,28	2,22	2,17	2,12	2,07	2,04	2,00
1000	2,34	2,27	2,20	2,15	2,10	2,06	2,02	1,98

Tabelle A.10.: F-Test, P = 0,99

P = 0,99	f₁							
	f₂	18	19	20	25	30	40	50
1	6192	6201	6209	6240	6261	6287	6303	6334
2	99,44	99,45	99,45	99,46	99,47	99,47	99,48	99,49
3	26,75	26,72	26,69	26,58	26,50	26,41	26,35	26,24
4	14,08	14,05	14,02	13,91	13,84	13,75	13,69	13,58
5	9,61	9,58	9,55	9,45	9,38	9,29	9,24	9,13
6	7,45	7,42	7,40	7,30	7,23	7,14	7,09	6,99
7	6,21	6,18	6,16	6,06	5,99	5,91	5,86	5,75
8	5,41	5,38	5,36	5,26	5,20	5,12	5,07	4,96
9	4,86	4,83	4,81	4,71	4,65	4,57	4,52	4,41
10	4,46	4,43	4,41	4,31	4,25	4,17	4,12	4,01
11	4,15	4,12	4,10	4,01	3,94	3,86	3,81	3,71
12	3,91	3,88	3,86	3,76	3,70	3,62	3,57	3,47
13	3,72	3,69	3,66	3,57	3,51	3,43	3,38	3,27
14	3,56	3,53	3,51	3,41	3,35	3,27	3,22	3,11
15	3,42	3,40	3,37	3,28	3,21	3,13	3,08	2,98
16	3,31	3,28	3,26	3,16	3,10	3,02	2,97	2,86
17	3,21	3,19	3,16	3,07	3,00	2,92	2,87	2,76
18	3,13	3,10	3,08	2,98	2,92	2,84	2,78	2,68
19	3,05	3,03	3,00	2,91	2,84	2,76	2,71	2,60
20	2,99	2,96	2,94	2,84	2,78	2,69	2,64	2,54
25	2,75	2,72	2,70	2,60	2,54	2,45	2,40	2,29
30	2,60	2,57	2,55	2,45	2,39	2,30	2,25	2,13
35	2,50	2,47	2,44	2,35	2,28	2,19	2,14	2,02
40	2,42	2,39	2,37	2,27	2,20	2,11	2,06	1,94
45	2,36	2,34	2,31	2,21	2,14	2,05	2,00	1,88
50	2,32	2,29	2,27	2,17	2,10	2,01	1,95	1,82
60	2,25	2,22	2,20	2,10	2,03	1,94	1,88	1,75
70	2,20	2,18	2,15	2,05	1,98	1,89	1,83	1,70
80	2,17	2,14	2,12	2,01	1,94	1,85	1,79	1,65
90	2,14	2,11	2,09	1,99	1,92	1,82	1,76	1,62
100	2,12	2,09	2,07	1,97	1,89	1,80	1,74	1,60
200	2,03	2,00	1,97	1,87	1,79	1,69	1,63	1,48
500	1,97	1,94	1,92	1,81	1,74	1,63	1,57	1,41
1000	1,95	1,92	1,90	1,79	1,72	1,61	1,54	1,38

Tabelle A.11.: F-Test, P = 0,99

Literaturverzeichnis

- [1] ATKINS, PETER W.: *Physikalische Chemie*. VCH, Weinheim, Cambridge, New York, Erste Auflage, 1990.
- [2] BACKHAUS, KLAUS, BERND ERICHSON, WULFF PLINKE und ROLF WEIBER: *Multivariate Analyse-Methoden*. Springer, Berlin, Heidelberg, New York, London, Paris, Tokyo, Siebente Auflage, 1994.
- [3] BRONSTEIN, I. N., K. A. SEMENDJAJEW, G. GROSCHE, V. ZIEGLER und D. ZIEGLER: *Teubner Taschenbuch der Mathematik*. B. G. Teubner, Stuttgart, Leipzig, 1996.
- [4] DOERFFEL, K. und R. GEYER: *Analytikum*. VEB, Leipzig, Siebente Auflage, 1987.
- [5] DORNINGER, D., G. EIGENTHALER und H. KAISER: *Mathematische Grundlagen für Chemiker II*. Prugg, Eisenstadt, 1981.
- [6] DUTTER, RUDOLF: *Einführung in die Wahrscheinlichkeitsrechnung und Statistik für Informatiker I*. <http://www.statistik.tuwien.ac.at/public/dutt/inf1/inf.html>, 1995.
- [7] DUTTER, RUDOLF: *Einführung in die Wahrscheinlichkeitsrechnung und Statistik für Informatiker II*. <http://www.statistik.tuwien.ac.at/public/dutt/inf2/inf.html>, 1995.
- [8] EIGEN, MANFRED und RUTHILD WINKLER: *Das Spiel, Naturgesetze steuern den Zufall*. Piper, München, Vierte Auflage, may 1996.
- [9] FAHRMEIR, LUDWIG, RITA KÜNSTLER, IRIS PIGEOT und GERHARD TUTZ: *Statistik*. Springer, Berlin, Heidelberg, New York, London, Paris, Tokyo, 1997.
- [10] FREUND, RUDOLF J. und WILLIAM J. WILSON: *Statistical Methods*. Academic Press, San Diego, London, New York, Tokyo, Revised Auflage, 1997.
- [11] GREENWOOD, NORMAN N., A. EARNSHAW und K. HÜCKMANN: *Chemie der Elemente*. VCH, Weinheim, Cambridge, New York, Erste Auflage, 1990.
- [12] HARTUNG, JOACHIM, BÄRBEL ELPELT und KARL-HEINZ KLÖSNER: *Statistik*. Oldenbourg, München, Sechste Auflage, 1987.

- [13] KETTELER, GUARDIAN: *Zwei Nullen sind keine Acht, Falsche Zahlen in der Tagespresse*. Birkhäuser, Basel, Boston, Berlin, 1997.
- [14] KÖCK, WOLFRAM K.: *Kognition — Semantik — Kommunikation*. In: SCHMIDT, SIEGFRIED J. (Herausgeber): *Der Diskurs des radikalen Konstruktivismus*, Seiten 340–373. Suhrkamp, Frankfurt am Main, Siebente Auflage, 1996.
- [15] KOLDITZ, LOTHAR (Herausgeber): *Anorganikum, Lehr und Praktikumsbuch der anorganischen Chemie mit einer Einführung in die physikalische Chemie*. VEB, Berlin, Zwölfte Auflage, 1989.
- [16] KRÄMER, WALTER: *So lügt man mit Statistik*. Campus, Frankfurt, New York, Sechste Auflage, 1994.
- [17] KUNZE, UDO R.: *Grundlagen der quantitativen Analyse*. Thieme, Stuttgart, Zweite Auflage, 1986.
- [18] OTTO, MATTHIAS: *Chemometrie. Statistik und Computereinsatz in der Analytik*. Wiley-VCH, Weinheim, 1997.
- [19] POPPER, KARL RAIMUND: *Auf der Suche nach einer besseren Welt — Vorträge und Aufsätze aus dreißig Jahren*. Piper, München, Siebente Auflage, mar 1994.
- [20] PRIGOGINE, ILYA und ISABELLE STENGERS: *Dialog mit der Natur — Neue Wege naturwissenschaftlichen Denkens*. Piper, München, Siebente Auflage, 1993.
- [21] SACHS, LOTHAR: *Angewandte Statistik*. Springer, Berlin, Heidelberg, New York, London, Paris, Tokyo, Siebente Auflage, 1992.
- [22] SHANNON, C. E. und W. WEAVER: *The mathematical theory of communication*. University of Illinois Press, Urbana and Chicago, 1949.

Index

Symbols

4-Sigma-Bereich *siehe* Ausreißertests,
4-Sigma ...

A

AAS 31
Abszisse 35
Abweichung von Normalverteilung 47
Alternativhypothese 25
Analyse der Residuen 81–84
 Ausreiser 84
 Graphiken 82
 Inhomogenitäten 83
 Varianz nicht homogen 83
Angabe eines Meßergebnisses
 Beispiel 51–53
Ausgleichsgerade 75
 Formel 80
Ausreißer 47, 55, 72, 84
 entfernen 55
Ausreißertest 48, 55–58
 4-Sigma-Bereich 55
 Dean und Dixon 56
Aussagen
 scharfe, unscharfe 46

B

Bestimmtheitsmaß 69
Blindwert 46
 Standardabweichung 47
Boxplot 15

C

Cramersche Regel 79

D

Dean und Dixon Test *siehe* Ausreißer-
tests, Dean und Dixon

E

ECD 31
Eichgerade *siehe* Regression, 77
 Nullpunkt 86
Eichkurve 77
Entropie 19, 21
Erfassungsgrenze 46–47
Extrapolation 75

F

F-Test 62–63
 Stichprobengröße 63
Fehler
 systematischer 31
 zufälliger 31, 42
Fehlerarten 31–32
Formale Korrelation 71
Freiheitsgrad
 Definition 42
funktionaler Zusammenhang 67

G

Gang 52
Gaußsche Glockenkurve 38
GC 31
Geradengleichung 79
gravimetrische Bestimmung 31
Grundgesamtheit 32
 endlich große 32
 unendlich große 33

H

Heteroskedastizität 83
Histogramm 15, 34–36
Homoskedastizität 83
Hypothesen 25

I

Informationsgehalt eines Zustandes 20

- Informationstheorie 19–24
 Inhomogenitätskorrelation 71
 Interpolation 75
 Irrtumswahrscheinlichkeit 26
- K**
- Kausalität 65, 70–73
 Ausschluß von 73
 Klassenhäufigkeit 35
 Kleinste Summe der Fehlerquadrate
 siehe least sum of squared residuals
 Korrelation 65, 67
 Ausreißer 72
 formale 71
 gemeinsame Basis 72
 in den Medien 73
 Inhomogenitäten 71
 Scatterplot 72
 Visualisierung 85
 weitere Untersuchungen 71
 Zeitreihen 71
 Korrelationskoeffizient 66
 Berechnung 69
 Definition 67
 Kausalität? 70
 Quadrat des 69
 Scatterplot 77
 Visualisierung 85
 Zusammenfassung 73
- L**
- least sum of squared residuals 79
 linearer Bereich 76
 Approx. Sinus 76
 Lineplots 66
- M**
- Meßergebnis, Angabe 29–53
 Überblick 30
 Median 48
 schiefe Verteilungen 49
 Vergleich mit Mittelwert 48
 Mittelwert 40
 falsche Anwendung 29
 Grundgesamtheit 41
 Stichprobe 41
- Modalwert 49
 multivariate Statistik 15
- N**
- Nachweisgrenze 46–47
 Normalverteilung 38–40
 Erwartungswert 40
 Mittelwert, Standardabweichung
 39
 Wahrscheinlichkeit 42
 Normalverteilung, testen 47
 Nullhypothese 25
- O**
- Ordinate 35, 66
- P**
- Perzentil 49
 Poissonverteilung 38
 Power *siehe* Tests, Power
 Prüfen auf Normalverteilung . . . *siehe*
 Normalverteilung, testen
 Präzision
 absolute 46
 relative 46
- Q**
- Q-Wert nach Dean und Dixon 56
 Quartil 49
- R**
- r *siehe* Korrelationskoeffizient
 Rückmeldungen 13
 Rauschen 81
 Regression 75–87
 Ableitung 79–80
 Scatterplot 77
 Scatterplot, Beispiel 77
 Vertauschen der Variablen . 84–86
 Voraussetzungen 81
 Zusammenfassung 86
 Reproduzierbarkeit *siehe* Fehler,
 zufälliger
 Residuen 15
 Definition 79, 81
 Illustration 78
 Normalverteilung 79

- Richtigkeit *siehe* Fehler,
systematischer
- robuste Statistik 40, 48–49
 Median 48
 Perzentil 49
 Quartil 49
- S**
- Scatterplot 15, 66, 72
scharfe Aussagen . . . *siehe* Aussagen,
scharfe, unscharfe
- Scheinkorrelation 71, 72
signifikante Stellen 49–50
 Addition, Subtraktion 50
 Multiplikation, Division 50
Signifikanzniveau 27
Software 13
 1-2-3 41
 Gnuplot 14
 R, S, SPlus 14
 SPSS 14
 Statistica 14
spontane Vorgänge 21
Stamm-Blatt-Diagramm 36–38
 Vergleich mit Histogramm 38
Standardabweichung 41
statistische Tests 25–27
Stichprobe 32
 analytische Praxis 41
Stichprobe einer Grundgesamtheit . 33
Streuung der Meßwerte 42
Subtraktion ähnlicher Größ en 50
systematischer Fehler . . . *siehe* Fehler,
systematischer
- T**
- t-Test 60–62
Tabellenkalkulation 66, 69
Taschenrechner 15
Tests
 Fehler 1.Art 26
 Fehler 2.Art 27
 konservative 27
 Power 26, 27
Thermodynamik 21
- U**
- univariate Statistik 15
- V**
- Varianz 44
Variationskoeffizient 44
Vergleich Mittelwert mit bekanntem
 Wert 61
Vergleich von Messungen 59–63
Vergleich von Mittelwerten 60–62
Vergleich von Standardabweichungen
 62–63
- Verteilung 33–40
 Dichtemaximum 52
 diskrete 36
 Fläche unter Kurve 38
 kontinuierliche 36
 kontinuierliche;Illustration 36
 Normierung 39
 schiefe 36, 48
 Wahrscheinlichkeit 33, 39
Vertrauensbereich 44–46
 z-Parameter 46
 größe der Grundgesamtheit 46
 Intervall 44
Vertrauensbereich und Standardab-
 weichung 45
Vertrauenswahrscheinlichkeit 26
- W**
- Wahrscheinlichkeit . . *siehe* Verteilung,
Wahrscheinlichkeit
Wahrscheinlichkeitstheorie 40
- Z**
- Zeitreihen 71
 graphische Darstellung 67
 Korrelation 71
 Trend 71
zentraler Grenzwertsatz 39
Zufallsfehler . . *siehe* Fehler, zufälliger