

CHEMOMETRIE

Skript zum Kurs Chemometrie
an der Fachhochschule Fresenius
März 1999

© Dr. habil. Wolfgang Durner

Vorwort

Dies ist kein Lehrbuch.

Dieses Skript für den Kurs *Chemometrie* enthält weder Abbildungen, noch Beispiele, noch Erläuterungen zu einzelnen Sachverhalten, noch Übungen, und auch keine statistischen Tabellen. All dies findet man in Lehrbüchern, von denen im Bereich der einführenden Statistik und Chemometrie genügend hervorragende Beispiele existieren.

Wozu also überhaupt ein Skript?

Der primäre Zweck des Skripts liegt darin, eine Orientierungshilfe darzustellen. Es soll den Aufbau der Veranstaltung übersichtlich und klar zu zeigen, und die Inhalte straff, aber vollständig aufzuführen. Das Skript befreit die TeilnehmerInnen von einem Großteil der standardmäßigen Mitschrift, so daß mehr Aufmerksamkeit auf das Verständnis der Vorlesung und auf die Notation von Besonderheiten und speziellen Hinweisen verwendet werden kann. Darüberhinaus kann das Skript von denjenigen, die den Stoff verstanden haben, als Formelsammlung verwendet werden.

Der Aufbau dieses Kurses folgt nicht einem einzelnen Lehrbuch. Der größte Teil basiert auf den Lehrbüchern von Khazanie [13], Köhler et al. [14], Sachs [22] und Kreyszig [16]. Das Skript selbst resultiert aus einem Entwurf, der als Begleitung zu einer allgemeinen Einführung in die Statistik für Geoökologen verfaßt wurde. Einige für die Chemometrie wesentliche Passagen wurden aus dem Vorgängerskript von Ludwig Ries oder aus dem Buch von Einax et al. [8] übernommen.

Das vorliegende Skript ist ein erster Entwurf und wird somit Fehler aufweisen. Es ist zudem recht heterogen aufgebaut: zu den meisten Themen sind nur die Stichworte aufgeführt, einige wenige andere Bereiche – etwa der Umgang mit Meßfehlern - wird etwas ausführlicher und subjektiv kommentiert. Ich hoffe dennoch, daß es die ihm zuge dachte Rolle erfüllen kann. Für Resonanz und kreative Verbesserungsvorschläge bin ich empfänglich und dankbar.

Bayreuth, im März 1999
Wolfgang Durner

Inhaltsverzeichnis

1	Einführung	1
1.1	Was ist Chemometrie ?	1
1.2	Motivation - Chemometrie als Werkzeug	2
1.3	Begriff und Gliederung der Statistik	3
2	Beschreibende Statistik	4
2.1	Ordnen und Organisieren von Daten	4
2.1.1	Grundgesamtheit, Stichprobe, Merkmalsausprägung	4
2.1.2	Skalenniveaus	5
2.1.3	Häufigkeiten und Häufigkeitsverteilungen	5
2.1.4	Klassifizieren von Daten	5
2.2	Maßzahlen zur Charakterisierung von Verteilungen	7
2.2.1	Momente und Zentralmomente	7
2.2.2	Lagemaße	7
2.2.3	Streuungsmaße	8
2.2.4	Schiefe und Wölbung	8
2.3	Darstellung univariater Stichproben	9
2.3.1	Tabellen	9
2.3.2	Grafische Darstellung von Daten	10
2.3.3	Histogramm	10
2.3.4	Empirische Verteilungsfunktion	11
2.3.5	Box-und-Whisker-Plot	11
2.4	Charakterisierung und Darstellung von Zusammenhängen	14
2.4.1	Bivariate Häufigkeitsverteilungen, Scatterdiagramme	14
2.4.2	Maßzahlen für Zusammenhänge: Kovarianz und Korrelation	14
2.4.3	Regression	15
3	Wahrscheinlichkeiten und W.-Verteilungen	18
3.1	Grundbegriffe der Wahrscheinlichkeitsrechnung	18
3.1.1	Experiment, Zufallsexperiment, Ereignis, Ereignisraum	18
3.1.2	Möglichkeiten: Permutation und Kombination	19
3.1.3	Der Wahrscheinlichkeitsbegriff	20
3.2	Wahrscheinlichkeitsverteilungen von Zufallsvariablen	21

3.2.1	Zufallsvariable, Verteilungsfunktion, Wahrscheinlichkeitsdichte . . .	21
3.2.2	Maßzahlen von Wahrscheinlichkeitsverteilungen	22
3.2.3	Rechenregeln für Erwartungswerte	23
3.2.4	Binomialverteilung	23
3.2.5	Poissonverteilung	25
3.2.6	Normalverteilung	26
3.2.7	Testverteilungen	29
3.3	Wahrscheinlichkeitsverteilungen als theoretische Modelle empirischer Häufigkeitsverteilungen	30
4	Messungen und Fehlerbehandlung	33
4.1	Datenerhebung und Verarbeitung	33
4.1.1	Kriterien für Messungen	33
4.1.2	Kriterien für Parameterschätzung	33
4.2	Fehler, Ausreißer und Fehlende Werte	34
4.2.1	Fehler bei Messungen	34
4.2.2	Fehler- und Ausreißerdetektion	35
4.2.3	Ausreißertests	36
4.2.4	Fehlende Daten	37
4.2.5	Fehlerfortpflanzung	38
4.3	Parameterschätzung	40
4.3.1	Schätzmethoden	40
4.3.2	Punktschätzung für Populationsparameter	41
4.3.3	Konfidenzintervalle	42
5	Tests	44
5.1	Grundlagen statistischer Tests	44
5.2	Parametrische Tests	46
5.2.1	t-Test zum Vergleich eines Mittelwerte mit einem theoretischen Wert . . .	46
5.2.2	t-Test zum Vergleich der Mittelwerte unabhängiger Stichproben	46
5.2.3	t-Test zum Vergleich der Mittelwerte verbundener Stichproben	47
5.2.4	t-Test zur Prüfung des Korrelationskoeffizienten	47

5.2.5	χ^2 -Anpassungstest zum Vergleich von beobachteten mit erwarteten Häufigkeiten	48
5.2.6	Kolmogorov-Smirnov-Anpassungstest zum Vergleich von beobachteten mit erwarteten Verteilungen	48
5.3	Nichtparametrische Tests	49
5.3.1	Vorzeichentest (Two Sample Sign Test) zum Vergleich zweier Mittelwerte verbundener Stichproben	50
5.3.2	Wilcoxon-Test zum Vergleich zweier Mittelwerte verbundener Stichproben	50
5.3.3	<i>U</i> -Test zum Vergleich zweier Mittelwerte unverbundener Stichproben	51
6	Analysen	52
6.1	Korrelationsanalyse	52
6.2	Regressionsanalyse	54
6.3	Varianzanalyse	57
7	Chemometrische Anwendungen	58
7.1	Präzision, Richtigkeit und Genauigkeit von Meßverfahren	58
7.2	Statistische Beurteilung von Analyseverfahren	59
7.2.1	Prüfung der Varianzhomogenität von Kalibrierfunktionen	59
7.2.2	Prüfung der Linearität der Eichgerade	60
7.2.3	Vertrauensbereiche um die Eichgerade	61
7.2.4	Nachweisgrenze	62
7.2.5	Erfassungsgrenze	63
7.2.6	Bestimmungsgrenze	63
7.3	Ringversuche	64

1 Einführung

1.1 Was ist Chemometrie ?

In den angewandten Naturwissenschaften besteht die Tendenz, daß nur mit einer weiteren Verfeinerung des wissenschaftlichen Instrumentariums die anstehenden, oftmals vielschichtigen Problemstellungen wissenschaftlich auf angemessene Weise behandelt werden können. Neben der fortschreitenden Verfeinerung der technischen Möglichkeiten bauen auch Auswerteverfahren auf zunehmend komplexen Datenanalyseverfahren auf. Diese beiden Tatsachen bedingen, daß die beruflichen Anforderungen an Chemiker und Chemieingenieure erheblich steigen.

Ein wesentliches Werkzeug zur Bewältigung dieser zunehmenden Komplexität ist die Statistik, die als elementares *Werkzeug* benötigt wird, um Daten zu erfassen, zu ordnen und zu präsentieren, und zu bewerten. Entscheidend beim Einsatz der Statistik ist demgemäß nicht nur in die Ordnung, Reduktion und Beschreibung von Datenmaterial, sondern der Umgang mit Unsicherheiten. Dies ergibt sich aus der folgenden Definition von Statistik, in der die Suche nach *wesentlichen und allgemeingültigen* Zusammenhängen betont wird:

Statistik ist die Wissenschaft, die sich mit allgemeinen Regeln und Methoden des Erfassens, Verarbeitens, Darstellens und Auswertens von zahlenmäßigen Informationen über Massenerscheinungen beschäftigt. Sie sucht nach wesentlichen und allgemeingültigen Erkenntnissen über Niveau, Struktur, Zusammenhang und Entwicklung dieser Erscheinungen.

Die Anwendung der Statistik vorwiegend auf die Thematik eines speziellen Wissenschaftsgebietes brachte in der Vergangenheit neue wissenschaftliche Disziplinen wie Biometrie, Psychometrie, medizinische Statistik, Ökonometrie etc. hervor. Diese Entwicklungen spiegeln sich wider in der Herausgabe von Zeitschriften wie "Biometrika" (1901), "Psychometrika" (1936), "Technometrics" (1959). Die grundlegenden Methoden dieser -metrien sind dieselben. Unterschiede finden sich in der Ausrichtung und Gewichtung der angewandten statistischen Verfahren. Die folgende Definition soll den Begriff der Chemometrie verdeutlichen:

Chemometrie umfaßt die wissenschaftliche Planung, Durchführung, Auswertung und Interpretation von chemischen Untersuchungen unter systematischer Zuhilfenahme mathematischer Methoden.

Die junge wissenschaftliche Disziplin "Chemometrie" hat sich in den letzten beiden Dekaden rasch entwickelt. Während ursprünglich die den limitierenden Schritt im analytischen Prozeß darstellte, führten seit den 50er Jahren Fortschritte in der Laborinstrumentierung und -Automation zu einer zunehmenden Flut von Rohdaten, die durch den Einsatz von Computern und benutzerfreundlicher Software bewältigt werden müssen. Chemometrie ist heute somit ein wichtiges Werkzeug der Datenanalyse und Qualitätssicherung in der Chemie, insbesondere im Bereich analytischer Chemie.

1.2 Motivation - Chemometrie als Werkzeug

Viele praktische Anforderungen an die Statistik bestehen im Lösen von Aufgaben, bei denen Ungewißheit eine Rolle spielt: solche Aufgaben können sein:

- Beurteilen von (scheinbaren) Widersprüchen
- Fällen von Entscheidungen in ungewissen Fällen
- Filtern von Informationen
- Reduzieren von Ungenauigkeiten
- Finden optimaler Versuchsbedingungen

Im Bereich der Wissenschaft und Forschung muß die Statistik bei folgenden elementaren statistische Prozessen eingesetzt werden:

- Hypothesenformulierung (empirisch)
- Hypothesenprüfung
- Optimierung von Versuchsanordnungen
- Optimierung von Datenerfassung
- Versuchsauswertung

Praktische Aufgabenstellungen der *Chemometrie* sind:

- Entwickeln einer Laboranalytik zur Praxisreife
- Entwicklung und Anwendung objektiver Kriterien zum Vergleich von Labormethoden
- Qualitätssicherung von Laboranalysen und Synthesen
- Steuerung chemischer und technischer Prozesse
- Optimierung von industriellen Produktionsprozessen

Ziel dieses Kurses

Dieser Kurs soll Ihnen zeigen, wie Sie mit wissenschaftlich fundierten statistischen Techniken Ihre Arbeitsweise verbessern können und so Ihren Ergebnissen größere Wirksamkeit verhelfen. Dabei werden Ihre analytischen und synthetischen Denkfähigkeiten genauso gefordert, wie Ihr Vermögen, präzise zu formulieren.

Sie sollten Ihren Sachverstand schulen, wie sprachlich formulierte Aufgabenstellungen möglichst präzise in einen Versuchsaufbau übersetzt werden können. Und Sie sollten Lernen, gemessene Werte in eine adäquat formulierte Aussage zu übersetzen. Konkret sollen Sie am Ende diese Kurses

- statistische Annahmen erkennen, die einer gegebenen Analysensituation unterliegen
- geeignete Methoden zum Auswerten von Messungen und Testen von Hypothesen finden können, und
- statistische Berichte anderer beurteilen und interpretieren können.

1.3 Begriff und Gliederung der Statistik

- *Objekt* der Statistik: Massenerscheinungen, die “zufällige” oder “zufallsartige” Elemente enthalten.
- *Verwendung des Begriffs* = $\left\{ \begin{array}{l} \text{a) für Ergebnisübersichten} \\ \text{b) für die Gesamtheit statistischer Methoden} \end{array} \right.$

Gliederung der Statistik

- *Beschreibende Statistik* (deskriptive-, deduktive-):
Präsentation und Auswertung von Daten
- *Beurteilende Statistik* (schließende-, induktive-):
Schlußfolgerungen über Grundgesamtheiten

Der Übergang von der beschreibenden zur schließenden Statistik erfolgt auf Basis von wahrscheinlichkeitstheoretischen Überlegungen.

Weiterhin läßt sich die Statistik nach der Art des Datenmaterials, das statistisch verarbeitet wird, in *uni-*, *bi-* und *multivariate* Statistik einteilen.

“Lügt” die Statistik ?

- Vorbehalte gegenüber Statistik entstehen aus vorsätzlichem Mißbrauch statistischer Methoden zum Zweck der Meinungsmanipulation. Ein solcher Mißbrauch ist nur möglich, wenn die “Kunden” über die verwendeten statistischen Verfahren, ihre Grenzen und Voraussetzungen nicht Bescheid wissen, oder wenn Datenmaterial durch bewußtes Weglassen oder Ergänzungen manipuliert wird.
- Grundsätzlich immanente Unsicherheit: keine exakte Aussage über einen *Einzelwert* möglich.
- Grundsätzlich kein Nachweis von Kausalzusammenhängen (“Statistik beweist nichts”)

Literaturhinweis: Huff [12] “*How to lie with statistics*”.

2 Beschreibende Statistik

Die Berechnung von statistischen Kennwerten und die Anfertigung von Graphen und Tabellen sind ein unverzichtbarer Teil der *primären Datensichtung*. Die gebräuchlichsten statistische Kennwerte für Datensätze umfassen den Mittelwert und die Standardabweichung für jede Variable, die Korrelation zwischen Datenpaaren, sowie die Regressionsberechnung, wenn Variablen von einander abhängig sind.

2.1 Ordnen und Organisieren von Daten

2.1.1 Grundgesamtheit, Stichprobe, Merkmalsausprägung

- Grundgesamtheit (*population*) — Stichprobe (*sample*)
(finit oder infinit) (immer finit)
- Parameter (*parameter*) — Stat. Zahl (*statistic*)
- Individuum — Merkmal — Merkmalsausprägung (auch: M.-Wert)

Bemerkungen:

Der **Grundgesamtheit** gilt in der Regel das Interesse. Es ist entscheidend, daß sie klar und eindeutig abgegrenzt wird in sachlicher, räumlicher und zeitlicher Hinsicht.

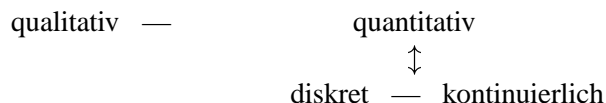
Die **Stichprobe** soll ein verkleinertes, repräsentatives Abbild der Gesamtheit darstellen. Sie dient in der Regel zur *Abschätzung* der Eigenschaften der Grundgesamtheit.

Eigenschaften der Grundgesamtheit werden durch **Parameter** gekennzeichnet, Eigenschaften der Stichprobe, die zur Schätzung der Parameter dienen, heißen **Statistische Zahl** (oder kurz "Statistik")

Das **Individuum** (auch: statistische Einheit, Einzelobjekt, Einzelprobe, Element, Untersuchungseinheit) steht in der Statistik niemals im Zentrum des Interesses; es ist jedoch der "Lieferant" der Einzelinformationen, die *zusammen* das Bild der Masse – des Untersuchungsobjekts – prägen.

Merkmale sind die Eigenschaften des Individuums, die untersucht werden. Die konkreten Werte, die von Merkmalen angenommen werden können, werden als **Merkmalsausprägung** oder Realisierung bezeichnet.

- Typen von Merkmalsausprägungen (**Datentypen**)



2.1.2 Skalenniveaus

<i>Skala</i>	<i>Datentyp</i>
Nominalskala	Häufigkeiten
Ordinalskala	Ränge
Intervallskala	metrische Daten (Meßwerte)
Verhältnisskala	metrische Daten mit absolutem Nullpunkt

2.1.3 Häufigkeiten und Häufigkeitsverteilungen

- absolute Häufigkeit f_i
(*frequency*)

relative Häufigkeit $f_i^{rel} = \frac{f_i}{\sum f_i}$
(*relative frequency*)

prozentuale Häufigkeit $f_i^{\%} = f_i^{rel} \cdot 100$
(*proc. frequency*)

- Häufigkeitsverteilung $f(x_i)$
(*frequency distribution*)

- Summenhäufigkeit $F_i = \sum_{j=1}^i f_j$
(*cumulative frequency*)

- Häufigkeitssumme (= Stichprobenumfang) $n = \sum f_i$
(*sample size*)

- Dichtefunktion $f(x)$
(*frequency density distribution*)

- Verteilungsfunktion $F(x) = \int_{-\infty}^x f(\xi) d\xi$
(*cumulative frequency distribution*)
(auch: Summenhäufigkeitsfunktion)

2.1.4 Klassifizieren von Daten

Um die Häufigkeitsverteilung von kontinuierlichen Daten visualisieren zu können, müssen diese klassifiziert werden. Auch bei diskret verteilten Daten, die viele verschiedene Werte annehmen können (so daß Einzelwerte nur in sehr geringen Häufigkeiten vorkommen), ist die Klassifizierung ein notwendiges Werkzeug, um die Verteilungsform darzustellen.

Ziel: Mit Minimum an Klassen ein Maximum an Information aus den Daten holen.

1. Scannen der Urdaten, um die Spannweite $V = x_{max} - x_{min}$ zu finden
2. Bestimmung der geeigneten Klassenzahl m
3. Festlegen der Klassengröße (Klassenbreite)
4. Festlegen der Klassengrenzen und -mitten

5. Finden der Klassenhäufigkeiten

Faustregel nach *Sturges*:

- Klassenzahl $m = 1 + 3.3 \lg n$
- Klassenbreite $w = \frac{(x_{max} - x_{min})}{m}$

Beschreibung der klassifizierten Daten durch f und F .

HINWEIS: Zur Frage der Klasseneinteilungen beachte auch die Aussagen zur Konstruktion von Histogrammen, Kap. 2.3.3

Transformieren von Daten

Oftmals müssen vor der Anwendung statistischer Verfahren Datensätze (oder Teile davon) einer Datentransformation unterzogen werden (um Voraussetzungen von statistischen Verfahren zu genügen).

Die wichtigste Transformation ist die sogenannte *Standardisierung* von Daten: Durch eine Lineartransformation werden die Daten auf eine Verteilung mit dem Mittelwert 0 und der Varianz 1 transformiert:

$$z = \frac{x - \bar{x}}{s}$$

wobei \bar{x} = arithmetisches Mittel und s = Standardabweichung der Daten ist (vg. Kap. 2.2).

Unter den nichtlinearen Transformationen ist die Logarithmierung von Originaldaten sehr gebräuchlich, wenn diese eine linkssteile Häufigkeitsverteilung aufweisen, oder deren Wertebereich mehrere Größenordnungen umfaßt.

2.2 Maßzahlen zur Charakterisierung von Verteilungen

Vorbemerkung:

In diesem Kapitel werden Datensätze mit Hilfe statistischer Maßzahlen charakterisiert. Wir interessieren uns an dieser Stelle nicht für eine Grundgesamtheit, die über den Datensatz hinaus geht. Wie zu verfahren ist, wenn die Maßzahlen der Grundgesamtheiten über Stichproben geschätzt werden, wird in Kapitel 4 behandelt werden.

2.2.1 Momente und Zentralmomente

(moments and central moments)

- k -tes Moment ist definiert durch $m_k = \frac{1}{n} \sum_{j=1}^n x_j^k$
(moment about the origin)
- k -tes Zentralmoment ist definiert durch $zm_k = \frac{1}{n} \sum_{j=1}^n (x_j - \mu)^k$
(moment about the mean)

2.2.2 Lagemaße

(measures of central tendency)

- arithmetisches Mittel $\mu = \bar{x} (= m_1)$
(mean)
- geometrisches Mittel $\bar{x}_{\text{geo}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \left(\prod_{j=1}^n x_j \right)^{1/n}$
(geometric mean)
- harmonisches Mittel $\bar{x}_{\text{har}} = \frac{1}{\frac{1}{n} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right)} = \left(n / \sum_{j=1}^n \frac{1}{x_j} \right)$
(harmonic mean)
- Modus (auch: Dichtemittel) $D = x_i$ für $f_i = \max$.
(mode)
Der Modus wird grundsätzlich an gruppierten Daten erhoben.
- Median (auch: Zentralmaß) $Z = x$ für $F(x) = 0.5$
(median)
Der Median wird grundsätzlich an ungruppierten Daten erhoben.

$$Z = \begin{cases} x_{\frac{n+1}{2}} & \text{für ungerade } n \\ \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{für gerade } n \end{cases}$$

- Perzentile
Allgemein: das p -te Perzentil einer Verteilung ist der Wert, unterhalb dessen ein Anteil p aller Beobachtungen liegt. Oft verwendete Spezialfälle sind die *Quartile*
 - Erstes Quartil $Q_1 = x$ für $F(x) = 0.25$

- Zweites Quartil $Q_2 = x$ für $F(x) = 0.50$
- Drittes Quartil $Q_3 = x$ für $F(x) = 0.75$

Perzentile haben u.a. eine praktische Bedeutung in der Toxikologie bei der Grenzwertediskussion. Hohe Perzentile wie z.B. das 95-er oder 98-er Perzentil werden in der Umweltüberwachung als Kennwerte für die Regelung von Grenzwerten eingesetzt. Hohe und niedrige Perzentilwerte sind wesentlich unempfindlicher gegen Ausreißer als es Maximalwerte bzw. Minimalwerte sind.

2.2.3 Streuungsmaße

(*measures of dispersion*)

- Variationsbreite $V = x_{max} - x_{min}$
(*range*)
- (Stichproben-)Varianz $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ ($= zm_2$)
(*variance*)
- Standardabweichung $\sigma = \sqrt{\sigma^2}$ bzw. $s = \sqrt{s^2}$
(*standard deviation SD*)
- Variationskoeffizient $cv = \frac{\sigma}{\mu}$ bzw. $cv = \frac{s}{\bar{x}}$
(*coefficient of variation*)
- Interquartilabstand $IQ = Q_3 - Q_1$
(*interquartile range*)

2.2.4 Schiefe und Wölbung

- Schiefe (*skewness*)

$$Sf = \frac{zm_3}{\sigma^3} = \frac{\frac{1}{n-1} \sum (x_i - \bar{x})^3}{\left(\frac{1}{n-1} \sum (x_i - \bar{x})^2 \right)^{3/2}}$$

linkssteil – symmetrisch – rechtssteil

$$\bar{x} > D \qquad \qquad \qquad \bar{x} < D$$

$$Sf > 0 \qquad \qquad \qquad Sf < 0$$

- Wölbung (*kurtosis*)

$$Ex = \frac{zm_4}{\sigma^4} - 3$$

relativ zur Normalverteilung:

positiv – “normal” – negativ

$$Ex > 0 \qquad \qquad Ex = 0 \qquad \qquad Ex < 0$$

HINWEIS: Das arithmetische Mittel kann nur von metrischen Daten berechnet werden. Es gilt nicht für ordinale Daten, für logarithmierte Werte (z.B. pH-Werte) oder für zirkulare Daten (z.B. Winkelangaben). Soll z.B. das arithmetische Mittel aus pH-Werten berechnet werden, so sind diese Werte zuerst zu delogarithmieren, dann wird das arithmetische Mittel daraus berechnet. Das Ergebnis wird anschließend wieder logarithmiert, um den pH-Mittelwert zu erhalten. Dasselbe Ergebnis ergibt sich auch unmittelbar aus dem geometrischen Mittel der pH-Werte. Ein wesentliches Merkmal des arithmetischen Mittels ist seine Empfindlichkeit gegenüber einzelnen großen Zahlen (etwa Ausreißern). Demgegenüber stellt der Median ein sehr robustes Maß dar: seine Berechnung wird durch einzelne extreme Werte nicht beeinflusst. Bei schiefen Verteilungen besitzen Modus, Median und arithm. Mittel eine charakteristische Abfolge (*“Fechnersche Lageregel”*). Der Median ist bei solchen Stichproben oft das geeignetste Maß zur Charakterisierung der mittleren Tendenz.

2.3 Darstellung univariater Stichproben

2.3.1 Tabellen

Man sollte in Manuskripten Datenmaterial tabellarisch präsentieren, wenn immer dies irgendwie sinnvoll möglich ist. Die Anfertigung klarer, übersichtlicher Tabellen erfordert jedoch besondere Sorgfalt. Die folgenden Hinweise nach *Chatfield* [5]) sind insbesondere für Tabellen zu beachten, bei denen Zahlenmaterial in Zeilen und Spalten präsentiert wird:

1. Zahlen sollten auf zwei signifikante Ziffern gerundet werden. Es ist nur verwirrend, wenn zu viele signifikante Ziffern gezeigt werden (wie es oft in Computerausdrucken der Fall ist). Standardmäßige Computerlisten müssen deshalb für Präsentationszwecke überarbeitet werden.
2. Die Maßeinheiten stets mitauführen.
3. Mittelwerte der Zeilen und Spalten sollten mit aufgeführt werden, sofern dies Sinn macht. Man überlege, ob in manchen Fällen eher die Angabe des Medians oder der Spaltensumme nützlich ist.
4. Die Abfolge von Variablen sollte sinnvoll gewählt werden. Sofern sich eine spezifische Anordnung nicht anbietet, sollten die Variablen z.B. nach der Größe ihrer Werte geordnet werden.
5. Es sollte stets geprüft werden, ob die Umkehrung von Zeilen und Spalten zu einer übersichtlicheren Präsentation führt. Spalten sind leichter zu lesen als Zeilen. Die Zahl der Zeilen sollte also größer sein als die Zahl der Spalten.
6. Für das generelle Layout der Tabelle sollten die Abstände von Zeilen und Spalten optimal gewählt werden. Zu enge Abstände wirken abschreckend, andererseits sollte ein übermäßiges Spreizen der Tabelle (um eine Seite auszufüllen) vermieden werden. Zu große Abstände zwischen Spalten sind generell schlecht. Bei einer sehr großen Zahl von Spalten sollten Untergruppen gebildet werden, die durch größere Abstände voneinander getrennt sind.
7. Eine Tabelle soll einen selbsterklärenden Titel besitzen.

2.3.2 Grafische Darstellung von Daten

Grafiken bilden die schnellste und beste Methodik, wenn es darum geht, einen qualitativen Überblick über (umfangreiches) Datenmaterial zu bekommen. Wenn umfangreiche Datensätze also durch Grafiken darstellbar sind, sollten sie in wissenschaftlichen Arbeiten auch so dargestellt werden. Tabellen haben dagegen den Vorteil, daß sie *exaktes* Zahlenmaterial aufweisen. In Diplomarbeiten und Dissertationen ist es sinnvoll, Daten im Hauptteil grafisch zu präsentieren und die zugehörigen Tabellen in einen Anhang zu verlegen. Gängige Darstellungen sind:

- Stab- (Balken-, Block-) Diagramm
- Pictogramme (Vergleich absoluter Häufigkeiten)
- Komponenten Stabdiagramm (Vergleich relativer Häufigkeiten; insb. bei Platznot)
- Kreisdiagramm (Darstellung relativer Häufigkeiten)
- Histogramm (Häufigkeitsverteilung kontinuierlicher Variablen)
- Polygonzug (Zeitreihen; Darstellung bivariater Daten)
- stetiger Ausgleich (ev. geglättet; zur Visualisierung von Trends)
- Wahrscheinlichkeitsplot (vergleichende Darstellung streuender Meßdaten)
- Boxplot (Datenreihen oder Vergleiche von streuenden Daten)

2.3.3 Histogramm

(Aus Ries [21], in Anlehnung an Nagel et al. [18], S. 37) Durch das Histogramm erhält der Anwender eine Vorstellung von der Häufigkeitsverteilung der Daten. Erkennbar werden: der Wertebereiche und die Form der Verteilung (symmetrisch, linkssteil, rechtssteil, ein- oder mehrgipflig), Bereiche größerer oder kleinerer Datenhäufigkeit, und die Variation der Werte in der untersuchten Datenmenge. Die Konstruktion des Histogramms beinhaltet wieder das Problem, daß die Anwender die Klassenbreite bzw. die Klassenzahl und die Lage der Klassen vorgeben müssen.

Erstellung eines Histogramms:

1. Die Fläche des Histogrammbalkens muß proportional zur Häufigkeit sein. Aus diesem Grunde kann nur bei konstanter Klassenweite der Wert der (Klassen)häufigkeit als Ordinate verwendet werden.
2. Bei ungleicher Klassenweite (davon ist grundsätzlich abzuraten) verwendet man für die Höhen der Balkens die Klassenhäufigkeiten dividiert durch die Klassenweiten. Dieser Quotient wird auch Dichte der Klassenhäufigkeit genannt.
3. Für die Auswahl der Klassenbreite empfiehlt es sich, neben der Anzahl der Daten auch die Variabilität der Daten zu berücksichtigen.
4. Die Klassenmitten werden mit Vorteil so gewählt, daß sie auf runde Zahlen fallen.
5. Wird das Histogramm mit dem Computer gezeichnet, so empfiehlt es sich, mehrere Klassenzahlen durchzurechnen und darzustellen. Die Anwender müssen dabei entscheiden zwischen einer zu geringen Klassenzahl, die wesentliche Strukturen der Häufigkeitsverteilung nivelliert und einer zu hohen Anzahl von Klassen, die die generelle Struktur der Häufigkeitsverteilung nicht mehr erkennen läßt.

Da jedes Histogramm nur eine mögliche Darstellung der Häufigkeitsverhältnisse der Stichprobe ist, und es bisher kein Verfahren zur Bestimmung des am besten geeigneten Histogramms gibt, ist in Zweifelsfällen das wiederholte Durchrechnen verschiedener Fälle eine Möglichkeit, mit der Anwender eine größere Sicherheit über die Häufigkeitsstruktur der untersuchten Daten gewinnen können.

2.3.4 Empirische Verteilungsfunktion

Werden die *kumulativen* relativen Klassenhäufigkeiten gegen x aufgetragen und durch eine Treppenlinie verbunden, so erhält eine monoton ansteigende Funktion, die bei null beginnt, und beim Wert eins endet. Diese Funktion heißt *empirische Verteilungsfunktion* der klassifizierten Daten.

Erstellung der empirischen Verteilungsfunktion (ungruppierte Daten):

1. Es sei n die Anzahl der Daten, so daß jeder Datenpunkt eine relative Häufigkeit $f_{rel} = 1/n$ besitzt.
2. Man sortiere alle vorkommenden Werte der Größe nach.
3. Die Verteilungsfunktion besitzt von $x = -\infty$ bis zum kleinsten Wert $x = x_1$ den Ordinatenwert $y = f(x) = 0$. An der Stelle x_1 springt $f(x)$ auf den neuen Wert $y = 1/n$.
4. Zwischen $x = x_1$ und $x = x_2$ bleibt die Verteilungsfunktion bei $y = 1/n$, um dann an der Stelle $x = x_2$ wiederum um den Betrag $\Delta y = 1/n$ zu wachsen.
5. Diese Prozedur wird an allen Stellen x_i wiederholt, so daß nach Passieren der Stützstelle $x = x_n$ die Ordinate den Wert 1 erreicht.
6. Von $x = x_n$ bis $x = +\infty$ bleibt die Funktion auf dem Wert 1.

Die Verteilungsfunktion ist zur Charakterisierung der einer Verteilung und zum Ablesen von robusten Lage- und Streuungsmaßen (Medien, Quartile, Perzentile, Interquartilabstand) ideal geeignet.

HINWEIS: Man beachte, daß im Gegensatz zum Histogramm die Verteilungsfunktion immer auf Basis der *ungruppierten* (=Original-) Daten erstellt werden sollte! Liegen keine Rohdaten, sondern bereits klassifizierte Daten vor, so sind die kumulativen Häufigkeiten jeweils am Intervallende der Klassen aufzutragen, und durch einen Polygonzug zu verbinden.

2.3.5 Box-und-Whisker-Plot

Eine sehr komprimierte und nützliche Darstellung der Verteilung von Daten ergibt sich in Form eines Box-und-Whisker-Plots. Im wesentlichen werden hierbei die Variationsweite und Quartile in der folgenden Weise grafisch dargestellt:

1. Man zeichnet eine Ordinate, die die größten und kleinsten vorkommenden Werte abdeckt.
2. Man zeichnet rechts von der Ordinate ein Rechteck (*box*) von Q_1 bis Q_3 .
3. Der Median wird innerhalb des Rechtecks durch eine waagrechte Linie gekennzeichnet.

4. Von Q_1 bis zum kleinsten Wert, und von Q_3 bis zum größten Wert werden durchgezogenen Linien (*whiskers* = “Barthaare”) gezeichnet.

HINWEIS: Es existieren einige Varianten des Box-und-Whisker-Plots, bei denen die *whiskers* nicht die gesamte Spannweite abdecken, und einzelne Extremwerte als gesonderte Punkte aufgetragen sind. Im Zweifelsfall sollte die Grafik kurz erklärt werden.

Manipulation durch grafische Darstellungen

- Verzerrung von Achsen
- Nicht-flächenproportionale Darstellung
- Unklare Pictogramme

Abschließende Bemerkungen zur graphischen Darstellung

Ein exzellentes Buch über statistische Grafiken ist *The Visual Display of Quantitative Information* von [25]. In diesem Buch werden nicht nur Beispiele von guten und schlechten Grafiken gezeigt, sondern zusätzlich erläutert *warum* eine Abbildung gelungen oder warum sie mißraten ist.

Es ist eine traurige Tatsache, daß trotz der Offensichtlichkeit der Grundregeln für gute Grafiken diese in wissenschaftlichen Publikationen und in Vorträgen gern ignoriert werden, und dafür bunter Murks serviert wird. Zu diesen wenigen Grundregeln gehört etwa, daß Grafiken und Tabellen klare, selbsterklärende Titel haben, daß die Achsen mit Titeln und Maßeinheiten versehen sind, und daß alle Schriften und Zahlen genügend groß und in einem sinnvollen Format dargestellt werden.

Ein Grund für das bestehende Qualitätsproblem mag darin liegen, daß Tabellen und Grafiken oft durch Computerprogramme produziert werden, die nicht automatisch erkennen können, welches Zahlenformat zum Beispiel für eine Darstellung optimal ist. Es sollte selbstverständlich sein, daß solche Outputs vor einer Präsentation entsprechend bearbeitet werden. Eine korrekte Darstellung der Verteilungsfunktion oder des Histogramms ist z.B. in Tabellenkalkulationsprogrammen wie EXCEL nicht als voreingestelltes Grafikmakro implementiert.

Ein weiterer Grund für die vielfältige Verbreitung schlechter Grafiken liegt sicher in der Leichtigkeit, mit der durch Computerprogramme auf den ersten Blick beeindruckende Farb- oder 3D-Effekte herbeigezaubert werden können. Solche Effekte führen jedoch in der Praxis sehr schnell dazu, daß die essentielle Botschaft einer Grafik verschleiert wird. Sie sollten deshalb sehr sparsam und überlegt eingesetzt werden.

Wirklich gute Grafiken und Tabellen fallen nicht einfach vom Himmel (oder aus dem Computer), sondern müssen mit Erfahrung und Gespür in Feinabstimmung erarbeitet werden.

2.4 Charakterisierung und Darstellung von Zusammenhängen

2.4.1 Bivariate Häufigkeitsverteilungen, Scatterdiagramme

Analog zum univariaten Fall können Daten klassifiziert und die resultierenden Häufigkeiten f_{ij} über 3-D Balkengrafik oder 3-D Histogramm visualisiert werden. Moderne Grafikverarbeitungssysteme bieten die Möglichkeit, diskrete oder kontinuierliche bivariate Häufigkeitsfunktionen $f(x, y)$ dreidimensional abzubilden.

Eine bessere Alternative ist oft die zweidimensionale Abbildung, bei der die Häufigkeit oder Häufigkeitsdichte über eine Farbcodierung (Graufärbung in schwarz-weißen Darstellungen) dargestellt wird.

Scatterdiagramm

Im Scatterdiagramm wird jedem Individuum im kartesischen Koordinatensystem ein Punkt zugeordnet. Die Achsen des Systems sind die Merkmalsachsen X_1 und X_2 (bzw. X und Y). Die Lage der Punkte ergibt sich aus den Werten x_1 und x_2 (bzw. x und y).

2.4.2 Maßzahlen für Zusammenhänge: Kovarianz und Korrelation

Kovarianz

Die Kovarianz σ_{xy} ist Maß für den Zusammenhang zweier streuender Variablen X und Y :

$$\sigma_{xy} = \frac{1}{n} \sum (x_i - \mu_x)(y_i - \mu_y)$$

bzw.¹

$$s_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

Pearson'sche Produkt-Moment-Korrelation

Der Korrelationskoeffizient ρ (bzw. der empirische Korrelationskoeffizient r) ist ein Maß, das geeignet ist, einen *linearen* Zusammenhang zwischen den zwei Variablen zu quantifizieren.

$$\text{linearer Korrelationskoeffizient } \rho = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \quad \text{bzw.} \quad r = \frac{s_{xy}}{s_x \cdot s_y}$$

mit σ_x bzw. s_x : Standardabweichung der x-Werte

σ_y bzw. s_y : Standardabweichung der y-Werte

σ_{xy} bzw. s_{xy} : Kovarianz der Variablen

Eigenschaften des linearen Korrelationskoeffizienten:

¹Obwohl zwei Abweichungen miteinander multipliziert werden, und die Kovarianz ein Pendant zur Varianz ist, schreiben wir s_{xy} , und nicht s_{xy}^2

- Es gilt immer: $-1 \leq r \leq 1$
- Nur dann, wenn $r = 1$ oder $r = -1$ liegen alle Punkte genau auf einer Geraden
- Das Vorzeichen von r ergibt sich aus der Art des Zusammenhangs
Wenn $r > 0$, dann ist die Korrelation positiv
Wenn $r < 0$, dann ist die Korrelation negativ
Wenn $r = 0$, dann sind die Stichprobenwerte *unkorreliert*
- Sind zwei Variablen voneinander unabhängig, so ist der Korrelationskoeffizient der Grundgesamtheit $\rho = 0$.

Bemerkung 1: Die Korrelation ist ein reines Zusammenhangsmaß für zwei Variablen. Sie geht *nicht* davon aus, daß eine der Variablen unabhängig, die andere abhängig ist.

Bemerkung 2: Soll der Korrelationskoeffizient einer bivariaten Stichprobe berechnet werden, so sind mindestens intervallskalierte Meßdaten erforderlich. Für eine sinnvolle Interpretation müssen die Daten außerdem bivariat normalverteilt sein (siehe Kapitel 6.1). Sind die Daten nur ordinalskaliert, so kann der Spearman'sche Rangkorrelationskoeffizient berechnet werden.

Spearman'scher Rangkorrelationskoeffizient

Der Rangkorrelationskoeffizient R verlangt zum einen nur ordinalskalierte Daten für X_1 und X_2 , zum anderen braucht kein linearer Zusammenhang vorausgesetzt werden. Es genügt Monotonie.

$$R = 1 - \frac{6 \sum d_i^2}{n \cdot (n^2 - 1)}$$

wobei

d_i Die Differenz des i -ten Rangplatzpaares

n Anzahl der untersuchten Individuen

i Laufindex

Für $n < 5$ sollte *kein* Rangkorrelationskoeffizient bestimmt werden, da er kaum Aussagekraft besitzt.

2.4.3 Regression

Lineare Regression: Least-squares Anpassung

Die Darstellung einer linearen Relation zwischen zwei Variablen ist durch eine Geradengleichung möglich.

Wird eine Variable y als von x abhängig definiert, so führt die Anpassung einer Ausgleichsgeraden $\hat{y}(x)$ in das (x, y) -Datenfeld nach der Methode der kleinsten Quadrate (*least squares fitting*)

$$\text{Min} \sum_{i=1}^n (y_i - \hat{y}(x_i))^2$$

zur Regressionsgeraden

$$\hat{y} = a + bx$$

mit den Koeffizienten

$$b = \frac{s_{xy}}{s_x^2} \quad \text{und} \quad a = \bar{y} - b\bar{x}$$

Bemerkung: Wird die angepaßte Gerade nur als Beschreibung eines Zusammenhangs zweier Variablen angesehen, wobei nicht entschieden werden kann, welche von welcher als abhängig anzusehen ist, so können die Rollen von x und y vertauscht werden. Man erkenne, daß die Gerade $\hat{x}(y) = a_2 + b_2 y$ nicht mit der Geraden $\hat{y}(x)$ übereinstimmt. Man spricht in diesem Zusammenhang auch von der "Schere der Regressionsgeraden". Die Geradensteigung $b_2 = \frac{s_{xy}}{s_y^2}$ hängt mit der Geradensteigung b über die einfache Beziehung $b \cdot b_2 = r^2$ zusammen.

Linearisierte Regression

Besitzen Daten einen Zusammenhang, der durch eine Datentransformation linearisiert werden kann, so kann die Lineare Regression für die linearisierten Daten durchgeführt werden. e -Funktionen, \ln -Funktionen, Potenzfunktionen und Polynome werden oft durch lineare Regression gefittet.

VORSICHT: Da die least-squares Anpassungskriterien auf die transformierten Daten angewandt werden, können teilweise überraschende oder unbeabsichtigte Ergebnisse erzielt werden!

Im folgenden wird die Anpassung eines Polynoms zweiten Grades an ein Datenfeld gezeigt (nach [10], S. 49ff).

Fragestellung:

Gesucht werden die Koeffizienten a_0 , a_1 und a_2 die den funktionellen Zusammenhang zwischen einer unabhängigen Variablen X und einer abhängigen Y nach folgender Gleichung beschreiben:

$$y = a_0 + a_1 \cdot x + a_2 \cdot x^2$$

Berechnung der Hilfsgrößen:

$$Q_{xx} = \sum x_i^2 - \frac{1}{n} \left(\sum x_i \right)^2$$

$$Q_{xy} = \sum (x_i y_i) - \frac{1}{n} \left(\sum x_i \right) \cdot \left(\sum y_i \right)$$

$$Q_x^3 = \sum x_i^3 - \frac{1}{n} \left(\sum x_i \right) \cdot \left(\sum x_i^2 \right)$$

$$Q_x^4 = \sum x_i^4 - \frac{1}{n} \left(\sum x_i^2 \right)^2$$

$$Q_{xy}^2 = \sum \left(x_i^2 y_i \right) - \frac{1}{n} \left(\sum x_i^2 \right) \cdot \left(\sum y_i \right)$$

Berechnung der Koeffizienten:

$$a_2 = \frac{Q_{xy} \cdot Q_x^3 - Q_{xy}^2 \cdot Q_{xx}}{(Q_x^3)^2 - Q_{xx} \cdot Q_x^4}$$

$$a_1 = \frac{Q_{xy} - a_2 \cdot Q_x^3}{Q_{xx}}$$

$$a_0 = \frac{1}{n} \left[\sum y_i - a_1 \cdot \sum x_i - a_2 \cdot \sum x_i^2 \right]$$

Aufgrund der Streuung der Daten um die angepaßte Funktion stellt diese lediglich eine Schätzung dar, deren Präzision durch die *Residualstandardabweichung* s_y ausgedrückt wird:

$$s_y = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 3}}$$

Die Anzahl der Freiheitsgrade ergibt sich als $FG = n - 3$, da drei aus den Daten abgeleitete Größen (a_0 , a_1 , a_2) in die Berechnung von s_y eingehen.

Nichtlineare Regression

Die nichtlineare Regression (Anpassung beliebiger Kurven an bivariate Datensätze) ist ein Bereich der mathematischen Disziplin "Nichtlineare Optimierung". Verschiedene Optimierungsstrategien sind verfügbar ("Steepest Descent", "Newton", "Levenberg-Marquardt"), die resultierenden Gleichungssysteme müssen iterativ gelöst werden. Die Verfahren werden z.B. in [19] näher beschrieben.

3 Wahrscheinlichkeiten und W.-Verteilungen

Die Wahrscheinlichkeitsrechnung ist die Theorie von zufälligen Ereignissen. Ein Basisverständnis der grundlegenden Wahrscheinlichkeitstheorie und einfacher Wahrscheinlichkeitsmodelle ist notwendig, um die Methodik statistischer Schließweise zu verstehen. Wahrscheinlichkeitsrechnung und Statistik sind komplementäre Gebiete. Während Statistik sich mit *induktiver* Schlußweise (DATEN \rightarrow MODELL) beschäftigt, untersucht die Wahrscheinlichkeitsrechnung die *deduktive* Schlußweise (MODELL \rightarrow Verhalten des Systems).

Es gibt einige philosophische Probleme bei der Frage darüber, wie der Begriff "Wahrscheinlichkeit" genau aufzufassen ist, und wie mit ihnen gerechnet werden kann. Verschiedene Auffassungen von Wahrscheinlichkeit umfassen gleichartige Wahrscheinlichkeiten (klass. Def.), relative Häufigkeiten in sehr langen Versuchsreihen (empirische Def. der Wahrscheinlichkeit), und einen subjektiven Wahrscheinlichkeitsbegriff. Für eine Diskussion dieser Auffassungen wird auf die Literatur verwiesen ([7, 15]).

3.1 Grundbegriffe der Wahrscheinlichkeitsrechnung

3.1.1 Experiment, Zufallsexperiment, Ereignis, Ereignisraum

Experiment (*experiment*)

= jeglicher Vorgang (Messung, Befragung, Beobachtung), der zu einem beobachtbaren Ergebnis führt (also sehr viel weiter gefaßt als *sensu strictu*).

deterministisch — *stochastisch*

Ein Vorgang ist *deterministisch*, wenn bei gegebenen Einflußgrößen das Resultat eindeutig und sicher prognostizierbar ist, und die Genauigkeit allein durch die Meßgenauigkeit der Einflußgrößen bestimmt ist.

In allen anderen Fällen ist ein Vorgang *stochastisch*. Die *Stochastik* als Teilbereich der Statistik ist demnach die Theorie zufälliger Vorgänge. Sie schließt "zufallsartige" Vorgänge mit ein.

Es kommt häufig vor, daß sich ein zufallsartiger Vorgang aus der Überlagerung eines stochastischen und eines deterministischen Vorgangs hervorgeht. Es läßt sich dann ein stochastischer Teil von einem deterministischen abspalten.²

Ereignisraum oder Ereignismenge (*sample space*)

= Menge aller denkbaren (möglichen) Ergebnisse; finit oder infinit. Der Ereignisraum umfaßt mindestens 2 Elemente.

Ereignis (*event*)

= aktuelles Resultat eines Experiments. Untermenge der Ereignismenge. Ein Elementarereignis besteht aus einem einzigen Element der Ereignismenge. Zusammen-

²In den Geowissenschaften ist das Problem der Größenskala oft entscheidend dafür, ob ein Vorgang als deterministisch oder stochastisch behandelt werden muß. Auf kleiner Skala deterministische Prozesse werden als subskalig aufgefaßt und für die Modellierung parametrisiert. Stochastisch verteilte *Subskalige Prozesse* stecken durch Mittelwertbildung im Datenmaterial; bei der Interpretation muß dies berücksichtigt werden. *Supraskalige Prozesse* dagegen können Nichtstationarität erzeugen und allmählich die Rahmenbedingungen verändern. *Literaturhinweis: [23]*.

gesetzte Ereignisse können aus Elementarereignissen kombiniert werden. Bei Wiederholung eines Experiments beeinflussen sich die Ereignisse gegenseitig nicht.

Zufallsexperiment (*random experiment*)

= Experiment, dessen Ausgang zufallsbehaftet ist. Der exakte Ausgang eines Zufallsexperiments ist also nicht exakt vorhersagbar, sondern besteht aus einer Ereignismenge, deren einzelne Elemente (mit einer definierten Wahrscheinlichkeit) verifiziert werden können.

Jedem Element einer diskreten Ereignismenge, und damit jedem Ereignis, läßt sich eine *Wahrscheinlichkeit* $P(E)$ zuordnen, mit der es bei Ausführung des Zufallsexperiments eintritt. Hierbei ist ein *einfaches* Ereignis dadurch gekennzeichnet, daß nur ein Element der Ereignismenge realisiert wird.

Ereignisse heißen *unabhängig*, wenn das Eintreffen des einen Ereignisses in seiner Wahrscheinlichkeit durch das Eintreffen des anderen nicht beeinflußt wird.

3.1.2 Möglichkeiten: Permutation und Kombination

Für zwei grundlegende Experimente sollen die Zahl der möglichen Ereignisse dargestellt werden:

1. **Permutation**

Seien n Objekte gegeben. Die Anzahl der Möglichkeiten, daraus eine Untermenge von r Objekten zu entnehmen ist

$$k = \frac{n!}{(n-r)!}$$

2. Eine Anzahl von r Objekten kann auf $r!$ verschiedene Weisen angeordnet werden.

3. **Kombination**

Seien n Objekte gegeben. Die Anzahl der *verschiedenen* Untermengen von r Objekten, die daraus kombiniert werden können, ist

$$k = \frac{n!}{(n-r)!r!} = \binom{n}{r}$$

3.1.3 Der Wahrscheinlichkeitsbegriff

- **Klassische Fassung**³

(Bernoulli, Laplace)

$$P(E) = \frac{n_{\text{Erfolg}}}{n_{\text{total}}}$$

Voraussetzg.: alle $P(E)$ sind gleich wahrscheinlich (a priori).

- **A posteriori Fassung** (Richard von Mises)

Beruht auf der Beobachtung: Bei häufigen Wiederholungen zeigen relative Häufigkeiten im allgemeinen eine auffallende Stabilität

$$P(E) = \text{dessen relative Häufigkeit in einer sehr großen Versuchsreihe}$$

- **Subjektive (a priori) Fassung** (Savage)

Die Wahrscheinlichkeit des zufälligen Ereignisses A ist eine auf Einsicht bzw. Erkenntnis einer oder mehrerer Personen beruhende Zahlenangabe über das voraussichtliche Eintreffen des entsprechenden Ereignisses.

- **Axiomatische Fassung**⁴ (Kolmogorow)

1. Die Wahrscheinlichkeit ist eine Zahl zwischen 0 und 1

$$0 \leq P \leq 1 \quad \text{Nichtnegativ}$$

2. Das *Sichere Ereignis* hat die Wahrscheinlichkeit 1

$$P(S) = 1 \quad \text{Normierung}$$

3. Die Wahrscheinlichkeit, daß von mehreren, *sich paarweise einander ausschließenden* Ereignissen, eines auftritt, ist gleich der Summe der Einzel-Wahrscheinlichkeiten

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) \quad \text{Additivität}$$

4. Die Eintrittswahrscheinlichkeit eines zusammengesetzten Ereignisses ist bei *stochastischer Unabhängigkeit* der Einzelereignisse gleich dem Produkt der Einzel-Wahrscheinlichkeiten

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$

Produktdefinition der stochastischen Unabhängigkeit

³Die Wahrscheinlichkeit wird durch das Symbol p (*probability*) dargestellt

⁴Axiom = Grundsatz, der keines Beweises bedarf. Axiome definieren die mathematischen Eigenschaften der Wahrscheinlichkeit, nicht deren Wesen.

3.2 Wahrscheinlichkeitsverteilungen von Zufallsvariablen

3.2.1 Zufallsvariable, Verteilungsfunktion, Wahrscheinlichkeitsdichte

- **Zufallsvariable**⁵ X
(*random variable*)

= Variable, die den Ausgang eines Zufallsexperiments charakterisiert, also alle Werte eines Ereignisraums annehmen kann (qualitativ oder quantitativ; diskret oder kontinuierlich). Schreibweise: Die Zufallsvariable X nimmt mit einer bestimmten Wahrscheinlichkeit $P(x_i)$ den Wert x_i an. Die Wahrscheinlichkeiten, mit denen die Werte des Ereignisraums angenommen werden, sind durch die Verteilungsfunktion (bzw. die Wahrscheinlichkeitsdichtefunktion oder die Wahrscheinlichkeitsfunktion) gegeben. Ist eine Zufallsvariable X normalverteilt mit dem Mittelwert μ und der Varianz σ^2 (siehe Kap. 3.5), so schreiben wir $X \sim N(\mu, \sigma^2)$, wobei das Symbol \sim bedeutet 'ist verteilt'.

- **Verteilungsfunktion** $F(x)$
(*cumulative distribution function, cdf*)

Zur eindeutigen Festlegung einer Zufallsvariablen dient die Verteilungsfunktion; Sie gibt die Wahrscheinlichkeit an, mit der X einen Wert *kleiner oder gleich* x annimmt:

$$F(x) = \text{Prob}(X \leq x)$$

Im Fall diskreter Zufallsvariablen ist die Verteilungsfunktion eine Treppenfunktion, die an den Stellen x_i Sprungstellen besitzt, wobei die Sprunghöhen durch $f(x_i)$ gegeben sind. Aus den Axiomen der Wahrscheinlichkeitsrechnung ergibt sich zwangsläufig, die Verteilungsfunktion die Werte 0 bis 1 durchläuft.

- **Wahrscheinlichkeitsfunktion** $f(x)$
(*probability distribution function*)

Für diskrete Zufallsvariable können den verschiedenen möglichen Werten des Ereignisraums zugehörige Wahrscheinlichkeiten angegeben werden.

$$f(x_i) = \text{Prob}(X = x_i)$$

Die Wahrscheinlichkeitsfunktion ist für beliebige x_i definiert und nimmt überall einen Wert zwischen 0 und 1 an. Sie ist somit das theoretische Pendant zur (empirischen) relativen Häufigkeitsverteilung.

Zwischen Verteilungsfunktion und Wahrscheinlichkeitsfunktion besteht folgende funktionale Beziehung:

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

⁵Konvention zur Schreibweise: Zufallsvariablen werden immer mit Großbuchstaben bezeichnet; Einzelne Werte ihres Wertebereichs dagegen mit Kleinbuchstaben (z.B. hat die Zufallsvariable X an irgendeiner Stelle x eine bestimmte Wahrscheinlichkeit $P = \text{Prob}(X = x)$.)

- **Wahrscheinlichkeitsdichte** $f(x)$
(probability density function, pdf)

Eine stetige Zufallsvariable liegt dann vor, wenn die Variable in einem bestimmten Intervall jeden beliebigen Wert annehmen kann. Die Angabe einer Wahrscheinlichkeit für das Eintreten eines einzelnen Wertes $P(X = a)$ ist zwangsläufig immer Null (!); Wahrscheinlichkeiten können immer nur für Intervalle angegeben werden (die natürlich beliebig klein sein können).

Die Wahrscheinlichkeitsdichte ist definiert als Ableitung der kontinuierlichen Verteilungsfunktion $F(x)$.

$$f(x) = \frac{dF(x)}{dx}$$

Umgekehrt läßt sich die Verteilungsfunktion aus der Wahrscheinlichkeitsdichtefunktion durch Integration berechnen:

$$F(x) = \int_{-\infty}^x f(\xi) d\xi$$

Die Wahrscheinlichkeit, daß ein Ereignis einen Wert zwischen a und b (einschließlich) annimmt, ergibt sich aus:

$$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(\xi) d\xi$$

3.2.2 Maßzahlen von Wahrscheinlichkeitsverteilungen

Das arithmetische Mittel einer Zufallsvariablen (oder eines Ausdrucks, in dem eine solche Variable involviert ist) heißt *Erwartungswert*. Da Wahrscheinlichkeitsverteilungen als theoretische Modelle für Grundgesamtheiten dienen, werden die gebräuchlichsten Maßzahlen analog zu den Parametern von Grundgesamtheiten mit griechischen Buchstaben bezeichnet: Mittel μ , Varianz σ^2 , Standardabweichung σ .

Für die Berechnung von Mittelwert und Varianz von Wahrscheinlichkeitsfunktionen gilt

- diskrete Wahrscheinlichkeitsverteilungen

$$E(X) = \mu = \sum x_i f(x_i)$$

$$VAR(X) = E(X - \mu)^2 = \sigma^2 = \sum (x_i - \mu)^2 f(x_i)$$

- kontinuierliche Wahrscheinlichkeitsverteilungen

$$E(X) = \mu = \int_{-\infty}^{\infty} x f(x) dx$$

$$VAR(X) = E(X - \mu)^2 = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

3.2.3 Rechenregeln für Erwartungswerte

Wenn man mit X bzw. X_1, X_2, \dots, X_n verschiedene Zufallsvariable und mit k eine Konstante bezeichnet ($k \neq 0$), dann gelten folgende Regeln für Erwartungswerte:

1. $E(X \pm k) = E(X) \pm k$
2. $E(X \cdot k) = k \cdot E(X)$
3. $E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)$
 $E(X_1 - X_2) = E(X_1) - E(X_2)$
4. $E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i)$ nur bei stochastischer Unabhängigkeit der X_i

Für Erwartungswerte des zweiten Zentralmoments (Varianzen) gelten folgende Rechenregeln (sei $VAR(X) = E[(X - \mu)^2]$):

1. $VAR(X \pm k) = VAR(X)$ weil $VAR(k) = 0$
2. $VAR(X \cdot k) = k^2 \cdot VAR(X)$
3. $VAR\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n VAR(X_i)$ nur bei stoch. Unabhängigkeit der X_i
4. $VAR(X_1 \pm X_2) = VAR(X_1) + VAR(X_2) \pm 2 \cdot COVAR(X_1, X_2)$

3.2.4 Binomialverteilung

Die Binomialverteilung wurde für Zufallsexperimente mit einem dichotomischen Ereignisraum entwickelt. Es gibt bei praktischen Untersuchungen viele Fälle, in denen der Ereignisraum aus zwei sich ausschließenden Ereignissen A mit einer Eintreffwahrscheinlichkeit von p und \bar{A} einer Eintreffwahrscheinlichkeit von $q = (1 - p)$ besteht. Wird ein solches Experiment oft (n mal) wiederholt, wobei der Ausgang jeweils unbeeinflusst vom letzten Ausgang bleibt, so heißt dies *Bernoulli - Experiment*:

1. Das Experiment besteht aus unabhängigen Ereignissen
2. Die Ereignisse sind entweder als Erfolg (success) oder Nichterfolg (failure) einzuordnen
3. Die Erfolgswahrscheinlichkeit des einzelnen Ereignisses ist bekannt und bleibt während des Experiments konstant

Die Wahrscheinlichkeitsfunktion für dieses Experiment wird durch die Binomialverteilung wiedergegeben:

In einem Binomial-Experiment mit einer konstanten Erfolgswahrscheinlichkeit p je Versuch ist die Wahrscheinlichkeit von x_i Erfolgen bei n Versuchen durch

$$f(x_i) = P_{(x_i \text{ Erfolge})} = \frac{n!}{x_i! (n - x_i)!} \cdot p^{x_i} \cdot q^{n - x_i}$$

gegeben, wobei $x_i = 0, 1, \dots, x_n$ eine ganze Zahl, und $q = 1 - p$ die Wahrscheinlichkeit eines einzelnen Mißerfolgs ist.

Eigenschaften der Binomialverteilung

1. Die Binomialverteilung ist eine diskrete Verteilung, da die Ereignisse nur ganzzahlige Werte annehmen können.
2. Die Binomialverteilung ist eine zweiparametrische Verteilung. Sie wird durch n =Umfang des Zufallsexperiments (bzw. der Stichprobe) und p = Wahrscheinlichkeit des Erfolges vollständig beschrieben.
3. Die Binomialverteilung gilt nur für den Fall, daß die Wahrscheinlichkeit p bei allen Realisierungen des Zufallsexperiments gleich groß ist. Das ist bei unendlichen Gesamtheiten und bei endlichen Gesamtheiten beim Ziehen mit Zurücklegen der Fall.
4. Der Erwartungswert der Binomialverteilung ist $E(X) = \mu = n \cdot p$
5. Die Varianz einer binomialverteilten Zufallsvariable ist $VAR(X) = n \cdot p \cdot q$
6. In Abhängigkeit von der Höhe der Wahrscheinlichkeit p bzw. der Gegenwahrscheinlichkeit q verändert sich die Form der Binomialverteilung. Bei $p < 0.5$ ist sie linkssteil, bei $p > 0.5$ rechtssteil, bei $p = q = 0.5$ symmetrisch. Bei $p = 0.5$ und großen n nimmt die Binomialverteilung eine glockenförmige Gestalt an.
7. Die Binomialverteilung kann unter bestimmten Bedingungen durch andere Verteilungen approximiert werden.
 - Durch die Normalverteilung, wenn sowohl die Bedingung $n \cdot p \geq 5$ als auch $n \cdot q \geq 5$ erfüllt ist.
 - Durch die Poissonverteilung, wenn p sehr klein und n groß ist, aber die Bedingungen zur Approximation durch die Normalverteilung nicht erfüllt sind.
8. Die grafische Darstellung der Binomialverteilung erfolgt mittels Stabdiagramm (Wahrscheinlichkeitsfunktion) und mittels Treppenfunktion (Verteilungsfunktion).

Die einzelnen Binomial-Wahrscheinlichkeiten sind für kleine n tabelliert. Zur Schreibweise: $b_{(4;10,0.9)}$ bedeutet Wahrscheinlichkeit von vier Erfolgen in einer Reihe von zehn Versuchen mit der Erfolgswahrscheinlichkeit 0.9 .

3.2.5 Poissonverteilung

Der französische Mathematiker *Poisson* führte diese Verteilung 1837 ein. Er leitete sie durch Grenzwertbetrachtungen aus der Binomialverteilung ab. Die Binomialverteilung strebt unter der Bedingung, daß p gegen Null und n gegen unendlich geht und der Mittelwert $\mu = n \cdot p$ einen endlichen Wert annimmt, gegen die Funktion

$$f(x) = \frac{\mu^x}{x!} e^{-\mu} \quad \text{für } x = 0, 1, 2, \dots, n$$

Diese Funktion gibt für ein Zufallsexperiment, bei dem die Wahrscheinlichkeit eines bestimmten Ereignisses p sehr klein, die Anzahl der Versuche aber sehr groß ist, an, mit welcher Wahrscheinlichkeit das Ereignis genau x mal eintritt.

Eigenschaften der Poissonverteilung

1. Die Poissonverteilung ist eine diskrete Verteilung.
2. Die Poissonverteilung ist eine einparametrische Verteilung. Sie wird durch μ vollständig beschrieben.
3. Der Erwartungswert der Poissonverteilung ist $E(X) = \mu = n \cdot p$
4. Die Varianz einer poissonverteilten Zufallsvariable ist gleich ihrem Erwartungswert: $VAR(X) = \sigma^2 = n \cdot p$
5. Die Poissonverteilung ist wegen der kleinen Werte von p stark linkssteil. Mit zunehmendem Wert von μ nimmt die Asymmetrie ab. Für $\mu > 7$ ist sie nahezu symmetrisch.
6. Die Poissonverteilung ist eine gute Approximationsverteilung für die Binomialverteilung. Bei $n > 100$ und $p < 0.01$ unterscheiden sich die Werte der Poisson- und der Binomialverteilung nur unwesentlich.
7. Die Poissonverteilung kann ihrerseits bei $\mu > 7$ durch die Normalverteilung approximiert werden.
8. Die grafische Darstellung der Poissonverteilung erfolgt mittels Stabdiagramm (Wahrscheinlichkeitsfunktion) und mittels Treppenfunktion (Verteilungsfunktion).

Die Poissonverteilung liegt tabelliert vor (in vielen Lehrbüchern allerdings nur die Verteilungsfunktion).

3.2.6 Normalverteilung

Die Normalverteilung ist die wichtigste Wahrscheinlichkeitsverteilung in der Statistik, weil bei vielen Anwendungen in der Praxis damit gerechnet werden kann, daß die Werte der untersuchten Grundgesamtheit annähernd normalverteilt sind, und weil die Normalverteilung als Approximation für andere (schwieriger zu berechnende) Verteilungen benutzt werden kann. Weiterhin erwächst die überragende Rolle der Normalverteilung aus dem zentralen Grenzwertsatz der Statistik (s.u.).

Die **Wahrscheinlichkeitsdichte** (auch: *Gauß'sche Glockenkurve*) der Normalverteilung ist allgemein durch den Ausdruck

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

gegeben.

Eigenschaften der Normalverteilung

1. Die Normalverteilung ist eine stetige Wahrscheinlichkeitsverteilung.
2. Die Normalverteilung ist eine zweiparametrische Verteilung. Sie wird vollständig beschrieben durch den Mittelwert μ und die Varianz σ^2 der Zufallsvariablen X .
3. Die Normalverteilung ist symmetrisch und hat eine glockenförmige Gestalt.
4. Das Dichtemaximum liegt an der Stelle μ .
5. Die Wendepunkte der W.Dichte liegen bei $\mu \pm \sigma$.
6. Für $\sigma = 1$ ist das Dichtemaximum $f(x) \simeq 0.4$; je größer die Standardabweichung ist, desto flacher wird die W.Dichte.
7. Die Flächenanteile unter der Normalverteilung nehmen für das z -fache der Standardabweichung folgende Werte an:

$$P(\mu \pm \sigma) = 0.683$$

$$P(\mu \pm 2\sigma) = 0.955$$

$$P(\mu \pm 3\sigma) = 0.998$$

Ein Flächenanteil von 95% wird durch den Bereich $\mu \pm 1.96\sigma$ eingenommen, ein Anteil von 90% durch den Bereich $\mu \pm 1.65\sigma$.

Für die praktische Arbeit ist die Normalverteilung mit dem Mittelwert $\mu = 0$ und der Varianz $\sigma^2 = 1$ tabelliert. Diese spezielle Verteilung heißt *Standardnormalverteilung*:

$$f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$$

Die Transformation einer normalverteilten Zufallsvariablen X zu einer Standardnormalverteilten Variablen Z erfolgt durch Standardisierung:

$$Z = \frac{X - \mu}{\sigma} \quad .$$

Die **Verteilungsfunktion** der Normalverteilung ist durch das Integral der Wahrscheinlichkeitsdichte definiert:

$$F(z) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^z \exp\left(-\frac{1}{2}\left(\frac{\xi - \mu}{\sigma}\right)^2\right) d\xi$$

Die Werte der Verteilungsfunktion der Standardnormalverteilung sind in allen Statistikbüchern tabelliert. Die Verteilungsfunktion kann durch

$$B = \frac{1}{2} \left[1 + 0.196854|z| + 0.115194|z|^2 + 0.000344|z|^3 + 0.019527|z|^4 \right]^{-4}$$

approximiert werden [2], wobei $|z|$ den Absolutwert von z bezeichnet und $F(z)$ durch

$$F(z) = \begin{cases} B & \text{für } z < 0 \\ 1 - B & \text{für } z \geq 0 \end{cases}$$

gegeben ist. Der Fehler für $F(z)$ ist in dieser Approximation kleiner als 0.00025.

Zentraler Grenzwertsatz der Statistik

Der Grenzwertsatz der Statistik gilt als bedeutendster Satz der Wahrscheinlichkeitslehre. Auf ihm fußt die induktive Statistik:

Werden aus unabhängigen, identisch verteilten Grundgesamtheiten von beliebiger Verteilungsform mit dem Mittelwert μ und der Varianz σ^2 Stichproben des Umfangs n entnommen, so sind die Mittelwerte dieser Stichproben approximativ normalverteilt mit $\mu_{\bar{x}} = \mu$ und $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$.

Die Approximation wird besser mit wachsendem n . Die Mittelwerte sind auch bei kleinem n bereits normalverteilt, wenn sie einer normalverteilten Grundgesamtheit entstammen.

Die Standardabweichung

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{1}{\sqrt{n}} \cdot \sigma$$

heißt *Standardfehler des Mittelwerts* oder mittlerer Fehler.

Erzeugung normalverteilter Zufallswerte

Für **Simulationsstudien** ist es wichtig zu wissen, wie man auf einem Rechner Realisationen normalverteilter Zufallsvariabler erzeugen kann. In der Regel bieten Programmiersprachen (oder Programme) einen Zufallszahlengenerator, der unabhängige Realisationen x_i einer im Intervall $[0 \dots 1]$ gleichverteilten Zufallsgröße bereitstellt. Bildet man nun

$$y = \frac{\sum_{i=1}^n x_i - \frac{n}{2}}{\sqrt{n/12}}$$

so hat man es – wegen der Gültigkeit des zentralen Grenzwertsatzes – für genügend großes n näherungsweise mit der Realisation einer $N(0, 1)$ -verteilten Zufallsgröße zu tun. Oft reicht dabei schon $n = 12$ aus, so daß

$$y = \sum_{i=1}^{12} x_i - 6$$

ist. Allerdings können hierbei keine Extremwerte auftauchen da $-6 \leq y \leq 6$ gilt. Eine weitere Möglichkeit zur Erzeugung von $X \sim N(0, 1)$ findet sich z.B. in *Hartung*, S. 147.

Aufgrund der Bedeutung der Normalverteilung in der Statistik sei dieser Abschnitt mit einer etwas ungewöhnlichen approximativen Darstellung ihrer Dichteverteilung abgeschlossen:

THE
NORMAL
LAW OF ERROR
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALIZATIONS OF NATURAL
PHILOSOPHY • IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES
IN THE PHYSICAL AND SOCIAL SCIENCES AND
IN MEDICINE AGRICULTURE AND ENGINEERING •
IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT

W.J. Youden

Annäherung der Binomialverteilung durch die Normalverteilung

Die Berechnung der Binomialkoeffizienten wird für große n zu einem sehr aufwendigen Unternehmen; dies besonders dann, wenn keine einzelne Wahrscheinlichkeit, sondern die Summe aller Wahrscheinlichkeiten größer oder kleiner gleich einem bestimmten Wert verlangt werden.

Für große n und wenn p nicht zu nahe an 0 oder 1 liegt, läßt sich die Binomialverteilung durch die Normalverteilung mit dem Mittelwert $\mu = np$ und der Standardabweichung $\sigma = \sqrt{npq}$ approximieren.

$$b_{(x_i; n, p)} \approx P(z_1 \leq z \leq z_2) = F(z_2) - F(z_1)$$

mit

$$z_1 = \frac{x_i - \frac{1}{2} - np}{\sqrt{npq}}$$

und

$$z_2 = \frac{x_i + \frac{1}{2} - np}{\sqrt{npq}} .$$

Als Faustregel gilt, daß sowohl das Produkt np als auch das Produkt nq größer als 5 sein müssen.

3.2.7 Testverteilungen

Einige Wahrscheinlichkeitsverteilungen spielen in der in der schließenden Statistik als sogenannte Prüf- oder Testverteilungen eine tragende Rolle. Dies deshalb, weil sie theoretische Modelle für die Verteilung von Parametern darstellen, die aufgrund von Stichproben berechnet werden. Die wichtigsten Testverteilungen sind die Normalverteilung, die χ^2 -Verteilung, die t -Verteilung und die F -Verteilung.

Zur mathematischen Formulierung der Dichte- und Verteilungsfunktionen sowie zur Kurvendiskussion dieser Verteilungen sei auf die einschlägigen Lehrbücher verwiesen. Anwendungen der Testverteilungen im Rahmen der schließenden Statistik werden in den folgenden Kapiteln gezeigt werden.

χ^2 - Verteilung

Die χ^2 -Verteilung wurde von F. Robert Helmert (1876) eingeführt und durch K. Pearson (1900) benannt ([22], S. 213). Sie ist die Wahrscheinlichkeitsverteilung einer Summe quadrierter, standardisierter, normalverteilter Zufallsvariablen. Die Wahrscheinlichkeitsdichte der χ^2 -Verteilung ist

$$f(x) = K_n x^{\frac{n-2}{2}} e^{-\frac{x}{2}} \quad ,$$

mit n = Anzahl der Freiheitsgrade, und K_n = Konstante, die unter Nutzung der Gammafunktion ermittelt wird.

Die χ^2 -Verteilung hat folgende Eigenschaften

1. Die χ^2 -Verteilung ist nur für positive Werte definiert.
2. Die χ^2 -Verteilung ist linkssteil; bei zunehmender Anzahl n asymptotisch symmetrisch.
3. Die Gestalt der χ^2 -Verteilung hängt von der Zahl der Freiheitsgrade ab. Für $n = 1$ und $n = 2$ fallen die Kurven monoton. Für $n > 2$ haben sie ein Maximum bei $x = n - 2$.
4. Mit wachsendem n wird die χ^2 -Verteilung symmetrischer und konvergiert gegen eine Normalverteilung.
5. Die χ^2 -Verteilung hat den Erwartungswert

$$E(X) = n$$

und die Varianz

$$VAR(X) = 2n \quad .$$

Die χ^2 -Verteilung ist für die wichtigsten Werte der Aussagesicherheit und für eine begrenzte Zahl von Freiheitsgraden in Lehrbüchern tabelliert.

t - Verteilung

Die t -Verteilung oder Studentverteilung wurde von W.S. Gosset eingeführt, der unter dem Pseudonym Student veröffentlicht hat. Bildet man aus einer standardnormalverteilten Zufallsvariablen X und einer χ^2 -verteilte Zufallsvariablen Y^2 mit n Freiheitsgraden einen

Quotienten der Form

$$T = \frac{X}{\sqrt{\frac{Y^2}{n}}} = \frac{X}{Y} \sqrt{n} \quad ,$$

so ergibt sich eine neue Zufallsvariable, die einer Wahrscheinlichkeitsverteilung folgt, die man t -Verteilung oder Studentverteilung nennt. Auf die Angabe der Dichtefunktion wird hier verzichtet und auf die Lehrbücher verwiesen. Die Dichtefunktion ist sehr mühsam zu berechnen; man greift somit in der Regel auf tabellierte Werte zurück, die für eine begrenzte Zahl von Freiheitsgraden in jedem Lehrbuch zu finden sind.

Die t -Verteilung hat folgende Eigenschaften:

1. Die t -Verteilung ist wie die Normalverteilung glockenförmig und symmetrisch zum Erwartungswert.
2. Die Gestalt der t -Verteilung hängt von der Anzahl der Freiheitsgrade ab. Mit zunehmendem n wird die Wahrscheinlichkeitsdichte steiler, und die Funktion strebt gegen die Standardnormalverteilung.
3. Für $n = 1$ hat die t -Verteilung keinen Mittelwert, für $n \geq 2$ ist der Erwartungswert $E(T) = 0$.
4. Für $n = 1$ und $n = 2$ hat die t -Verteilung keine Varianz, für $n \geq 3$ ergibt sich die Varianz als

$$VAR(T) = \frac{n}{n-2} \quad .$$

3.3 Wahrscheinlichkeitsverteilungen als theoretische Modelle empirischer Häufigkeitsverteilungen

Im Bereich der Umweltwissenschaften sind einige weitere Verteilungen als theoretische Modelle von beobachteten Häufigkeitsverteilungen interessant. Erwähnenswert sind die einparametrische Exponentialverteilung (z.B. als Modell für die Abstandszeit zwischen Niederschlagsereignissen oder radioaktiven Zerfällen), die zweiparametrische Weibullverteilung (z.B. als Modell für die Beschreibung von Materialermüdungen), die Lognormalverteilung (als Modell für viele natürlich vorkommende Meßgrößen) sowie die Pearson III- und die Gumbel-Verteilung (zur Beschreibung von Hochwasserspitzen). Drei dieser Verteilungen sollen hier kurz vorgestellt werden. Die Ausführungen dieses Abschnitts richten sich nach [6].

Normal - Verteilung

Aufgrund des zentralen Grenzwertsatzes ist zu erwarten, daß Variablen, die sich als Summe vieler unabhängiger Ereignisse summieren, normalverteilt sind. Im Bereich der Hydrologie trifft dies z.B. auf die Jahressumme der Niederschläge zu. Für viele andere hydrologische Größen, die nur positive Werte annehmen können und oft linkssteile Verteilungen aufweisen, ist die Normalverteilung allerdings nicht optimal geeignet (s.u.).

Lognormal - Verteilung

Viele in der Natur vorkommende Größen, die nur positive Werte annehmen können, sind nicht symmetrisch, sondern linkssteil verteilt, insbesondere wenn sich der Wertebereich über mehrere Größenordnungen erstreckt. In solchen Fällen kann oft durch logarithmieren erreicht werden, daß die Verteilung symmetrisch wird und sich durch eine Normalverteilung gut approximieren läßt.

Ist $\ln X$ eine $N(\mu, \sigma^2)$ -verteilte Zufallsvariable, so heißt X selbst *logarithmisch normalverteilt*. Oft wird davon ausgegangen, daß Größen, die durch Multiplikation normalverteilter Variablen zustandekommen, lognormalverteilt sind. Empirische Beispiele solcher Variablen sind die hydraulische Leitfähigkeiten von porösen Medien, oder die Größe von Regentropfen während eines Niederschlagsereignisses. Die Dichtefunktion der Lognormalverteilung ist für $x > 0$

$$f(x) = \frac{1}{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} .$$

Der Erwartungswert der lognormalverteilten Zufallsvariablen X ist

$$E(X) = e^{\mu + \frac{\sigma^2}{2}} ,$$

ihre Varianz ist

$$VAR(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) ,$$

und der Median und die Schiefe berechnen sich zu

$$D = S_f = \sqrt{(e^{2\mu + \sigma^2} - 1)} (e^{\sigma^2} + 2) .$$

Exponentialverteilung

Manche Reihen hydrologischer Ereignisse, wie z.B. das Auftreten von Niederschlägen oder das Auftreten von Oberflächenabschwemmungen, können als sogenannte *Poisson Prozesse* angesehen werden, bei denen die Ereignisse unabhängig voneinander auf einer (Zeit-)Achse auftreten. Die Zeit zwischen dem Auftreten der Ereignisse, die *interarrival time*, wird durch eine Exponentialverteilung beschrieben, deren Parameter λ die durchschnittliche Zeit zwischen zwei Ereignissen beschreibt. Die Exponentialverteilung ist eine einparametrische Verteilung, mit der Wahrscheinlichkeitsdichte

$$f(x) = \lambda e^{-\lambda x} \quad \text{für } x \geq 0$$

Der Erwartungswert ist

$$E(X) = \frac{1}{\lambda} ,$$

die Varianz

$$VAR(X) = \frac{1}{\lambda^2} .$$

Man beachte bei Anwendungen der Exponentialverteilung als Wahrscheinlichkeitsmodell für Prozesse, daß *Unabhängigkeit* der Ereignisse vorausgesetzt wird. Diese Voraussetzung ist z.B. bei Niederschlägen aus Fronten nicht erfüllt (vgl. [6]).

Stetige Gleich- oder Rechteckverteilung

Treten innerhalb eines Bereichs a bis b Werte mit gleich hoher Wahrscheinlichkeit auf, während sie außerhalb überhaupt nicht auftreten, so liegt eine *Gleich- oder Rechteckverteilung* (uniform distribution) vor. Beispiele sind die erzielten Augen beim gleichmäßigen Würfelwurf (diskrete Gleichverteilung). Zufallszahlengeneratoren von Computern liefern oft gleichverteilte Werte im Intervall $[0,1]$.

Die konstante Wahrscheinlichkeitsdichte der stetigen Gleichverteilung ist gegeben durch

$$y = f(x) = \begin{cases} 1/(b - a) & \text{für } a < x < b \\ 0 & \text{für } x \leq a \text{ oder } x \geq b \end{cases}$$

Der Erwartungswert ist

$$E(X) = \frac{a + b}{2} ,$$

die Varianz

$$VAR(X) = \frac{(b - a)^2}{12} .$$

Für den wichtigen Spezialfall $0 < x < 1$ ergibt sich ein Erwartungswert von 0.5 und eine Varianz von $1/12$, d.h. eine Standardabweichung von 0.28867.

4 Messungen und Fehlerbehandlung

Die zentralen Aufgaben der schließenden Statistik sind die Parameterschätzung und die Hypothesenprüfung. Grundsätzlich geht es darum, aus Meßwerten einer Stichprobe mit einer geschätzten Unsicherheit auf Eigenschaften der Grundgesamtheit rückzuschließen.

In diesem Kapitel soll in einem ersten Block zunächst beleuchtet werden, welche Regeln bei der Auswahl von Messgrößen, bei der Datenerfassung und Datenverarbeitung beachtet werden sollten. Im nächsten Schritt werden Fehlerarten und Möglichkeiten zur Aufspürung dieser Fehler diskutiert. In einem dritten Schritt wird aufgezeigt, wie sich Fehler (oder Unsicherheiten) bei Berechnungen auf das Endergebnis durchschlagen.

In einem zweiten Block wird schließlich dargestellt, nach welchen Kriterien unbekannte Parameter von Grundgesamtheit geschätzt werden können. Es geht dabei nicht nur um die Schätzung von Parameterwerten, sondern auch zugehöriger Vertrauensbereiche (Konfidenzintervalle), innerhalb derer sich die wahren Parameter mit einer vorgewählten Wahrscheinlichkeit befinden.

4.1 Datenerhebung und Verarbeitung

4.1.1 Kriterien für Messungen

Bei der Auswahl von Merkmalen und Merkmalsausprägungen zur statistischen Analyse müssen folgende Kriterien berücksichtigt werden:

- **Objektivität**
Objektiv ist ein Meßergebnis, wenn es nicht durch den Bearbeiter beeinflusst ist, wenn mehrere Bearbeiter durch einen analogen Meßvorgang zu dem selben Ergebnis gelangen.
- **Reliabilität** (Zuverlässigkeit)
Zuverlässigkeit der Meßwerte ist gegeben, wenn wiederholte Messungen ein- und desselben Merkmals am selben Objekt übereinstimmen.
- **Validität** (Gültigkeit)
Gültigkeit beweisen Meßergebnisse und damit Meßverfahren, wenn sie in der Lage sind, genau den Sachverhalt wiederzugeben, der von der Aufgabenstellung her gefordert war.

4.1.2 Kriterien für Parameterschätzung

Ein aus einer Stichprobe berechneter Schätzer $\hat{\theta}$ für einen Parameter θ muß folgende Kriterien erfüllen:

1. Erwartungstreue (*unbiasedness*)
Die Verteilungsfunktion des Schätzers $\hat{\theta}$ hat den Erwartungswert θ .
2. Wirksamkeit (*efficiency*)
Die Varianz des Schätzers aufgrund der Unsicherheit der Stichprobenwerte ist möglichst klein.

3. Konsistenz (*consistency*)

Mit wachsendem Stichprobenumfang nähert sich der Schätzer beliebig nahe an den zu schätzenden Parameter (*Konvergenz*).

4. Vollständigkeit (*sufficiency*)

Die Berechnung des Schätzers nutzt alle nutzbare Information der Stichprobe und hat somit eine minimale Varianz.

4.2 Fehler, Ausreißer und Fehlende Werte

Es gibt drei Typen von Problemdaten: Fehler, Ausreißer und fehlende Werte. Ein *Fehler* ist eine Beobachtung die falsch aufgezeichnet wurde, möglicherweise aufgrund eines Geräte-defekts, oder eines Übertragungs- oder Kopierfehlers. Ein *Ausreißer* ist dagegen ein Extremwert, der nicht so recht zum Rest der Daten passen will. Ausreißer gehen auf eine Reihe von möglichen Ursachen zurück und können viel Ärger bereiten. Eine grundlegende Behandlung des Problems incl. Lösungsvorschläge zu seiner Behandlung findet man bei [3].

4.2.1 Fehler bei Messungen

Allgemein wird davon ausgegangen, daß sich ein Meßwert aus einer wahren Information und einem fehlerhaften Anteil zusammensetzt. Hat man keine weiteren Informationen über die verwendete Methode und die ggf. analysierte Substanz, so ist die Größe des fehlerhaften Anteils zunächst unbekannt. Dieser fehlerhafte Anteil kann, soweit nichts weiteres über ihn bekannt ist, als Gesamtfehler bezeichnet werden. Der bei einer Messung entstehende Gesamtfehler setzt sich als Teilfehlern mit sehr unterschiedlichen Ursachen zusammen. Für eine Analyse und eine mögliche Reduzierung oder Beseitigung bestimmter Fehler ist ein allgemeines Modell der Fehlerursachen nötig:

$$\begin{array}{ccccccc} \text{Messwert} = & \text{wahrer} & + & \text{grober} & + & \text{systemat.} & + & \text{zufälliger} \\ & \text{Wert} & & \text{Fehler} & & \text{Fehler} & & \text{Fehler} \\ & & & \text{vermeidbar} & & \text{! gefährlich} & & \text{unvermeidbar} \end{array}$$

Grobe Fehler

Hierunter versteht man zum Beispiel einen Ablesefehler oder das Notieren einer falschen Zahl. Fehlerursachen sind häufig in menschlicher Fahrlässigkeit zu finden. Durch Plausibilitätskontrollen, mehrfache Messungen und die erneute Durchführung eines Versuchs kann die Wahrscheinlichkeit eines groben Fehlers sehr stark reduziert werden. Grobe Fehler sind in jedem Fall zu vermeiden.

Systematische Fehler

Kennt man einen systematischen Fehler, so ist dessen Behebung aus Sicht der Datenanalyse in einer Vielzahl von Fällen relativ einfach. Da (wie der Name andeutet) eine feste funktionelle Ursache diesen Fehler hervorruft, kann in vielen Fällen eine Funktion zur Datenkorrektur gefunden werden. Unter diesen Voraussetzungen ist die Behebung eines systematischen Fehlers auch nach bereits erfolgter Messung auf rechnerischem Weg aus den Daten

möglich. Entsprechend der Sachlage kann ein systematischer Fehler auch am Gerät oder am Versuchsaufbau direkt behoben werden. Systematische Fehler können *nicht* aus der statistischen Streuung der Daten heraus erkannt oder abgeschätzt werden. Sie können deshalb sehr gefährlich sein, insbesondere dann, wenn sie klein ausfallen und keine Parallelanalysen (mit anderen Verfahren, Geräten, Bearbeitern) vorgenommen werden.

Zufällige Fehler

Gilt der Versuchs- oder Meßaufbau als überprüft und richtig (=frei von systematischen Fehlern), so verbleibt noch ein zufälliger Fehler, der sich in kleineren Abweichungen bei wiederholten Messungen äußert. Die Ursachen des zufälligen Fehlers liegen in vielen kleinen, unvermeidbaren und in der Regel voneinander unabhängigen Fehlern. Die Größe des zufälligen Fehlers ist ausschlaggebend für die Reproduzierbarkeit (precision) einer Meßmethode. Unsicherheiten aufgrund zufälliger Fehler lassen sich durch statistische Mittel berechnen und ausdrücken.

4.2.2 Fehler- und Ausreißerdetektion

Was ist Fehler, was ist Ausreißer?

Fehler und Ausreißer werden oft durcheinandergebracht. Es gibt Situationen, in denen Fehler keine Ausreißer verursachen, und andere, in denen Ausreißer nicht auf Fehler zurückzuführen sind. Besonders problematisch sind Extremwerte, die aus linkssteilen Verteilungen entstammen, und somit tatsächlich berücksichtigt werden sollten.

Die Suche nach möglichen Fehlern ist ein wichtiger Teil der frühen Datensichtung. Einige Checks können von Hand durchgeführt werden, aber Computer können leicht so programmiert werden, daß sie einen großen Teil der routinemäßigen Prüfungen durchführen können. Es sollten die *Vertrauenswürdigkeit*, die *Konsistenz* und die *Vollständigkeit* der Daten geprüft werden.

Plausibilitätsprüfung von Daten

Beim Check auf Glaubwürdigkeit wird geprüft ob die Daten sich innerhalb eines zulässigen Wertebereichs bewegen. Unmögliche Werte oder sonstige extreme Ausreißer sollten dadurch detektiert werden.

Ein einfacher, aber sehr sinnvoller Check liegt darin, einen Ausdruck der Daten oder ihrer grafische Darstellung zu sichten. Obwohl es nicht möglich ist, die Zahlen im einzelnen zu lesen, ist das Auge recht effizient wenn es darum geht, Werte aufgrund einer veränderten Struktur zu entdecken — vorausgesetzt die Daten stehen in einem sinnvollen Format und strikt ausgerichtet in Spalten.

Beobachtet man bei der ersten Datensichtung einen oder mehrere Werte, die den Eindruck erwecken, daß sie Ausreißer sind, so ist als erstes zu klären, ob diese durch Meß-, Rechen-, Schreib-, oder Datenerfassungsfehler bedingt sein können, d.h., ob sie *systematische* Fehler sind. Oft kann man fehlerhafte Daten, die auf einen Datenverarbeitungsfehler zurückgehen, nachträglich korrigieren.

Der Umgang mit Extremwerten – ob diese nun auf Fehler zurückzuführen sind oder nicht – ist schwierig. So kann man mittels sogenannter *Ausreißertests* überprüfen, ob diese Daten überhaupt zur präzisierten Stichprobe zu rechnen sind. Sind sie dies signifikant nicht, so wird man sie bei einer Auswertung der Daten gar nicht berücksichtigen, um so eine Verfälschung der Ergebnisse zu vermeiden. Andererseits ist beim Ausschluß von Daten, die linkssteilen Verteilungen entstammen, höchste Zurückhaltung geboten, da solche Werte ja tatsächlich vorkommen können.

Können suspekte Daten nicht klar als Ausreißer eliminiert werden, so empfiehlt *Chatfield* (1994) einen praktischen Ansatz: Sofern kein Hinweis auf einen Fehler besteht, sollte eine statistische Prozedur mit dem Extremwert durchgeführt werden und sodann ohne ihn wiederholt werden. Wenn die Schlußfolgerungen aus den Tests durch die Berücksichtigung dieses Wertes maßgeblich verändert werden, so sollten Entscheidungen, die damit auf der Existenz eines oder zweier Werte beruhen, nur mit höchster Vorsicht gefällt werden.

4.2.3 Ausreißertests

Ausreißertests werden benutzt, um (1) routinemäßig die Zuverlässigkeit von Daten zu kontrollieren, (2) rechtzeitig gewarnt zu werden, die Datengewinnung besser zu kontrollieren und um (3) Beobachtungen, die extrem liegen und bedeutungsvoll sein könnten, zu erfassen.

Bei Normalverteilung der Daten eröffnet die Normalverteilungsfunktion einen direkten Weg mit dem die Wahrscheinlichkeit von seltenen aber korrekten sehr großen (extremen) Abweichungen vom Mittelwert bestimmt werden kann. Entsprechend der in Abschnitt 3.2.6 angegebenen Wahrscheinlichkeitsintervalle

- a) $P(\mu - \sigma < x < \mu + \sigma) \approx 68.3\%$
- b) $P(\mu - 2\sigma < x < \mu + 2\sigma) \approx 95.5\%$
- c) $P(\mu - 3\sigma < x < \mu + 3\sigma) \approx 99.7\%$
- d) $P(\mu - 4\sigma < x < \mu + 4\sigma) \approx 99.994\%$

ergibt sich, daß sich 68.3% der normalverteilten Werte im Intervall $\mu \pm 1\sigma$ befinden. Die Wahrscheinlichkeit einer Abweichung des Meßwerts x um mehr als 1σ von μ ist etwa einmal in drei Versuchen zu erwarten.

$$\text{a) } P(|x - \mu| > \sigma) \approx 0.317$$

Entsprechend gelten für die weiteren σ -Schranken folgende Wahrscheinlichkeiten für extreme Abweichungen

- b) $P(|x - \mu| > 2\sigma) \approx 0.0455$ (einmal in ca. 22 Versuchen)
- c) $P(|x - \mu| > 3\sigma) \approx 0.0027$ (einmal in ca. 370 Versuchen)
- d) $P(|x - \mu| > 4\sigma) \approx 0.000064$ (einmal in ca. 15625 Versuchen)

Dies zeigt, daß ein extremer Wert um so unwahrscheinlicher ist, je kleiner die Stichprobe. Einer allgemeinen Regel folgend (siehe [22], S. 364), darf bei einer Stichprobe mit dem Mindestumfang $n = 10$ ein Wert dann als Ausreißer verworfen werden, wenn er außerhalb des Intervalls $\bar{x} \pm 4s$ liegt. Mittelwert und Standardabweichung sind hierbei ohne den verdächtigen Wert zu berechnen.

Ausreißertests für nicht normalverteilte Daten

Es existieren viele Tests die dazu dienen, Ausreißer zu detektieren. Für kontinuierliche Variablen wurde bislang der DIXON-Test für $n \leq 29$ und der GRUBBS-Test für $n \geq 30$ empfohlen. Nach den letzten Empfehlung den der Standardisierungsorganisationen wird heute der DIXON-Test in modifizierter Form allgemein empfohlen. Die Testvorschrift sowie die tabellierten Prüfwerte findet man z.B. bei *Einax et al.* [8], S. 41f. Eine Übersicht über Testverfahren bietet *Rechenberg* [20].

HINWEIS: Die Anwender sollten sich bewußt sein, daß Ausreißer in einer Meßreihe entweder einen neuen Hinweis auf unerwartete aber real vorhandenen Variabilität des untersuchten Phänomens geben oder fehlerhafte Werte darstellen. Aus diesem Grund sollte man stets einen sehr sorgfältigen Blick auf festgestellte und entfernte Ausreißer legen und das Entfernen von Ausreißern nicht ausschließlich Computerprogrammen überlassen.

BEISPIEL: Hätten sich die maßgebenden Forscher nicht nur auf die Macht laufender Rechenprogramme verlassen, so wäre das Ozonloch über der Arktis vermutlich um einige Jahre früher entdeckt worden.

4.2.4 Fehlende Daten

Sofern Daten fehlen, ist zunächst zu klären, warum sie fehlen. Es ist ein entscheidender Unterschied, ob die fehlenden Daten auf völlig zufallsbedingte Ausfälle zurückzuführen sind, oder ob diese Ausfälle stets in ganz bestimmten Situationen auftreten. Im ersten Fall können die fehlenden Daten in einfachen Analysen einfach weggelassen werden, in komplexeren Verfahren (Varianzanalysen) durch "gefittete" Stellvertreter ersetzt werden. Es muß auf jeden Fall strikt davor gewarnt werden, fehlende Daten durch spezielle Zahlen zu codieren, da dies bei automatischen Auswertungen zu den seltsamsten Ergebnissen führen kann.

4.2.5 Fehlerfortpflanzung

Eine völlig genaue Messung einer kontinuierlichen Größe ist nicht möglich. Es existiert immer eine Abweichung $\Delta x = x_a - x_b =$ absoluter Fehler zwischen einem abgelesenen und dem realen Wert. Man unterscheidet konstante Fehler, systematische Fehler und zufällige Fehler.

Konstante Fehler lassen sich durch Differenzenbildung eliminieren. Sie sind keine “Fehler” im Sinne der Statistik, da ihnen keine Unsicherheit innewohnt.

Systematische Fehler einer Eingangsgröße sind ähnlich wie konstante Fehler durch eine gerichtete (aber unbekannt) Verschiebung eines Meßwerts gegenüber dem wahren Wert definiert. Die Abschätzung der möglichen Größe eines systematischen Fehlers aus Mehrfachmessungen ist mit statistischen Mitteln nicht möglich! Kann die Größe systematischer Fehler Δx_i in den Messungen der Eingangsvariablen anderweitig angegeben werden, so errechnet sich das “Durchpausen” auf das Ergebnis wie folgt:

$$\Delta y = \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i} \right| \Delta x_i$$

wobei

$\Delta y =$ Fehler der resultierenden Variablen

$\frac{\partial f}{\partial x_i} =$ partielle Ableitung der Funktion $y = f(x_1, x_2, \dots, x_n)$ nach x_i .

Diese partielle Ableitung wird *Sensitivität* (-skoeffizient) von y nach x_i genannt.

$\Delta x_i =$ Fehler der Eingangsgröße x_i .

Zufällige Fehler sind klassische Meßfehler in dem Sinn, daß Messungen um den wahren Wert mal mehr, mal weniger streuen. Oft sind diese Messungen um den wahren Wert normalverteilt.

Für zufällige Fehler ergibt sich das *Fehlerfortpflanzungsgesetz nach Gauß*, das eine mildere Schätzung des Gesamtfehlers eines Experiments als Folge verschiedener zufälliger Fehler in den gemessenen Eingangsgrößen gibt.

$$s_y = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 s_i^2}$$

mit $s_y =$ Standardabweichung des (normalverteilten) Ergebnisses und $s_i^2 =$ Varianzen der Eingangsgrößen x_i .

Die Gleichung besagt, daß sich die Fehler der Eingangsvariablen multipliziert mit den entsprechenden Sensitivitätskoeffizienten quadratisch (“vektoriell”) addieren.

Fortpflanzungsregeln für zufällige Fehler

Im folgenden gilt folgende Symbolregelung:

absoluter Fehler einer Variablen x : s_x

(= Stabw. einer wiederholten Messung)

relativer Fehler einer Variablen x : $\delta_x = s_x/\bar{x}$

Die Anwendung des *Fehlerfortpflanzungsgesetzes nach Gauss* führt bei unkorrelierten(!!!) Meßfehlern zu folgenden einfachen Regeln:

- **Fehlerreduktion durch wiederholte Messung**

Der Fehler eines Mittelwerts reduziert sich durch Wiederholungsmessungen.

$$\bar{x} = \frac{1}{n} \sum x; \quad s_{\bar{x}} = s_x/\sqrt{n}$$

Der wahre zu messende Wert liegt mit einer Wahrscheinlichkeit von ca. 68% im Intervall $[\bar{x} - s/\sqrt{n}, \bar{x} + s/\sqrt{n}]$.

- **Summe oder Differenz**

Die absoluten Fehler addieren sich vektoriell.

$$y = x_1 + x_2 \text{ oder } y = x_1 - x_2; \quad s_y = \sqrt{s_{x_1}^2 + s_{x_2}^2}$$

Merke: der relative Fehler kann sich bei der Differenzenbildung zweier etwa gleichgroßer Zahlen extrem vergrößern.

- **Produkt oder Quotient**

Die relativen Fehler addieren sich vektoriell.

$$y = x_1 \cdot x_2 \text{ oder } y = x_1/x_2; \quad \delta_y = \sqrt{\delta_{x_1}^2 + \delta_{x_2}^2}$$

- **Potenzen**

Der relative Fehler vervielfacht sich entsprechend der Potenz

$$y = x^n; \quad \delta_y = n \cdot \delta_x$$

Es empfiehlt sich in der Praxis, in größeren Gleichungen für jeden Term die Variable mit dem größten relativen Fehler zu identifizieren. Dieser bestimmt oft in erster Näherung den relativen Fehler des gesamten Terms.

Bei korrelierten Fehlern muß die Kovarianz der Meßfehler mit berücksichtigt werden.

Literatur: [16], S. 333.

4.3 Parameterschätzung

4.3.1 Schätzmethoden

Es gibt einige generelle Verfahren, um Parameter zu schätzen.

1. Momentenmethode (*method of moments*)
Die (parameterabhängigen) Momente der Wahrscheinlichkeitsverteilung werden den Momenten der Stichprobe gleichgesetzt.
2. Maximum-Likelihood-Methode (*method of maximum likelihood*)
Grundidee: Der beste Parameterwert einer Wahrscheinlichkeitsverteilung ist derjenige Wert, für den sich am ehesten der tatsächlich beobachtete Datensatz einstellt. Mathematisch: Man findet die verbundene Wahrscheinlichkeitsfunktion (joint probability oder joint pdf im kontinuierlichen Fall) der Daten in Bezug auf den unbekannt Parameter θ und maximiert diese sogenannte Likelihood-Funktion hinsichtlich θ . Da die Wahrscheinlichkeitsverteilungen oft auf Exponentialfunktionen basieren, ist es attraktiv, den Log der Likelihood-Funktion, d.h. die sogenannte Log-Likelihood-Funktion zu maximieren.
3. Methode der kleinsten quadratischen Abweichungen (*method of least squares*)
Die unbekannt Parameter werden geschätzt, indem die Summe der quadratischen Abweichungen zwischen den beobachteten und den (durch das angenommene Modell) berechneten Werte minimiert werden.
4. Methode der kleinsten absoluten Abweichungen (*method of least absolute deviations*)
Wie oben, nur wird die Summe der Abständsbeträge minimiert. Die Methode ist in der Regel analytisch sehr diffizil, aber durch Computer leicht durchführbar.

Die verschiedenen Verfahren führen oft zu den selben Formeln für die Schätzfunktionen. Da sie in der Regel durch fertige Computerprogramme bereitgestellt werden, werden ihre Details hier nicht weiter diskutiert.

4.3.2 Punktschätzung für Populationsparameter

Nach den genannten Kriterien ergeben sich⁶:

<i>Parameter</i>	<i>Optimaler Schätzer</i>
Mittelwert μ	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Varianz σ^2	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
Kovarianz σ_{xy}	$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
Korrelationskoeffizient ρ	$r = \frac{s_{xy}}{s_x s_y}$
Steigung der Regressionsgeraden β	$b = \frac{s_{xy}}{s_x^2}$
Achsenabschnitt der Regressionsgeraden α	$a = \bar{y} - b\bar{x}$

Der Stichprobenmittelwert ist also ein optimaler Schätzer für den Populationsmittelwert. Dagegen muß für die Schätzung der Populationsvarianz berücksichtigt werden, daß in der Berechnungsformel ein Wert vorkommt, der aus der Stichprobe bereits abgeschätzt werden mußte: der wahre Populationsmittelwert μ ist ja unbekannt. Dies kommt dadurch zum Ausdruck, daß die Summe der berechneten quadratischen Abweichungen nicht durch die Zahl der Messungen n , sondern durch eine kleinere Zahl, $(n - 1)$ geteilt wird. Diese kleinere Zahl nennt man *Anzahl der Freiheitsgrade* (FG oder *df=degrees of freedom*). Generell wird die Zahl der Freiheitsgrade durch die Zahl der Messungen (bzw. Meßpaare) gegeben, reduziert um 1 für jeden in der Gleichung auftretenden Wert der aus den Daten selbst berechnet wurde. Im Fall der Varianz wird z.B. der Mittelwert μ durch den aus den Daten errechneten Wert \bar{x} ersetzt, die Zahl der Freiheitsgrade ist somit $n - 1$.

HINWEIS: Da die Parameterberechnung für Stichproben praktisch stets in Hinblick auf die Parameterschätzung für die Grundgesamtheit vorgenommen wird, werden in der Regel stets die Schätzer berechnet, ohne daß dies in der Namensgebung deutlich wird. So wird z.B. s^2 (eigentlich irreführend) als *Stichprobenvarianz* bezeichnet.

⁶Zur Schreibweise: Populationsparameter sowie Parameter ihrer theoretischen Modelle, der Wahrscheinlichkeitsverteilungen, werden mit griechischen Buchstaben bezeichnet, die statistischen Kennzahlen der Stichproben dagegen mit den korrespondierenden lateinischen Buchstaben.

4.3.3 Konfidenzintervalle

Konfidenzintervalle (oder Vertrauensintervalle) sind Wertebereiche, die einen gesuchten Parameter mit einer gewissen Konfidenz (ähnlich, aber nicht identisch mit Wahrscheinlichkeit) einschließen. Intervallschätzungen sollten Punktschätzungen in der Regel vorgezogen werden.

Man hüte sich vor Angaben wie $a \pm b$ wenn nicht klar gesagt wird, ob b ein Standardfehler, zwei Standardfehler, ein 95% Konfidenzintervall, oder was auch immer ist.

- Allgemeine Fragestellung: Finde zwei Zahlen, die einen gesuchten unbekanntem Parameter mit einer vorgegebenen Konfidenz umschließen.
- Allgemeine Vorgehensweise: Der gesuchte Parameter ist eine Zufallsvariable, d.h. er streut in irgendeiner Weise um den "wahren" Mittelwert (z.B. normalverteilt, t -verteilt, oder χ^2 -verteilt).

Wir müssen also den Typ dieser *Testverteilung* kennen, und ihre Parameter aus der Stichprobe abschätzen. Sodann können wir aufgrund der Kenntnis der Streueigenschaften dieser Testverteilung ausrechnen, in welchem Intervall der unbekannt Parameter mit einer Konfidenz γ von 90%, 95% oder 99% eingeschlossen ist. Naturgemäß hängt die Weite des Intervalls vom Stichprobenumfang ab, da mit einer größeren Stichprobe der unbekannt Wert mit größerer Sicherheit geschätzt werden kann.

Beispiele für Konfidenzintervalle

1. Mittelwert bei bekannter und unbekannter Varianz

Wie oben angeführt, ist die Punktschätzung des Mittelwerts durch den Stichprobenmittelwert gegeben.

$$\hat{\mu} = \bar{x}$$

Der Hut über dem μ bedeutet "Schätzung für". Es kann gezeigt werden, daß die Stichprobenverteilung des Mittelwerts normalverteilt ist, d.h. $\bar{X} \sim N(\mu, \sigma^2/n)$. Die Streuung dieses Mittelwertes, σ/\sqrt{n} , wird *Fehler des Mittelwerts* genannt. Da wir σ kennen (dürfte in der Praxis eher selten vorkommen) können wir ein 95% Konfidenzintervall für μ berechnen

$$\bar{x} \pm 1.96\sigma/\sqrt{n}$$

Wenn σ dagegen aus der Stichprobe geschätzt werden muß, benutzen wir s/\sqrt{n} als geschätzten Standardfehler von \bar{x} und erhalten das 95% Intervall aus der t -Verteilung

$$\bar{x} \pm t_{0.025, n-1} s/\sqrt{n}$$

wobei t den Wert der t -Verteilung angibt, oberhalb dessen $2\frac{1}{2}\%$ der Verteilung liegt (Da die Normalverteilung wie die t -Verteilung symmetrisch ist, kann dieser Wert natürlich für die Ober- und Untergrenze verwendet werden, um die Bereiche zu kennzeichnen, die zusammen 5% aller Werte ausschließen).

2. Varianz einer Population

Die Stichprobenvarianz s^2 ist ein erwartungstreuer Schätzer für die Populationsvarianz σ^2 . Das 95%-Konfidenzintervall für σ^2 ist gegeben durch

$$\frac{(n-1)s^2}{\chi_{0.025, n-1}^2} \quad \text{bis} \quad \frac{(n-1)s^2}{\chi_{0.975, n-1}^2}$$

(Beachte, daß die χ^2 -Verteilung unsymmetrisch ist, und deshalb die Stellen, die die unteren $2\frac{1}{2}\%$ und die oberen $2\frac{1}{2}\%$ abschneiden, explizit aus der Tabelle entnommen werden müssen).

3. Mittelwertsdifferenz zweier Populationen

Wir haben zwei Stichproben, die normalverteilten Grundgesamtheiten entstammen, die durch verschiedene Mittelwerte, aber gleiche Varianz gekennzeichnet sind. Die Stichprobengrößen seien n_1 und n_2 . Wenn die Stichprobenparameter durch \bar{x}_1 , \bar{x}_2 , s_1^2 und s_2^2 gegeben sind, so ist die Differenz $(\bar{x}_1 - \bar{x}_2)$ eine erwartungstreue Schätzung für $(\mu_1 - \mu_2)$ mit einem Standardfehler von

$$\sqrt{\frac{\sigma}{n_1} + \frac{\sigma}{n_2}} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Das 95%-Konfidenzintervall für $(\mu_1 - \mu_2)$ bei bekanntem σ ergibt sich dann aus

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Wenn σ aus den Stichproben geschätzt werden muß

$$s^2 = \left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 \right] / (n_1 + n_2 - 2)$$

ergibt sich ein 95%-Konfidenzintervall für $(\mu_1 - \mu_2)$ bei unbekanntem, aber gleichem σ aus

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.025, n_1+n_2-2} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Die Voraussetzungen für die obigen parametrischen Schätzverfahren (Normalverteiltheit der Daten, gleiche Varianzen) sind leider oft nicht gegeben. Die Methoden gelten allerdings wg. des zentralen Grenzwertsatzes asymptotisch für beliebige Verteilungen, wobei die Näherung mit wachsendem n besser wird.

Für viele Fälle, in denen die Verteilungsvoraussetzung absolut nicht gewährleistet werden kann, existieren *robustere* nichtparametrische Schätzverfahren, die etwa Ausreißern weniger Gewicht zuordnen. Generell sollte nach der Philosophie verfahren werden, eher ein robustes Verfahren mit sicheren Voraussetzungen zu verwenden, als ein "schärferes" Verfahren, das auf unsicheren Beinen steht. Für nähere Informationen über nichtparametrische Verfahren sei auf die Spezialliteratur verwiesen (z.B. [24]).

5 Tests

Unter einem *Statistischen Test* verstehen wir einen Algorithmus, der dazu führt eine bestimmte Aussage über eine Grundgesamtheit zu akzeptieren oder abzulehnen.

Die Annahme oder Ablehnung ist gemäß dem Wesen der Statistik keine sichere Aussage, sondern mit einer Irrtumswahrscheinlichkeit behaftet. Wir können in sensiblen Fällen diese Irrtumswahrscheinlichkeit zwar klein wählen, dennoch kann mit statistischen Tests nichts *bewiesen* werden. Zudem muß berücksichtigt werden, daß man sich trotz der *Nichtablehnung* einer Hypothese nicht auf deren tatsächliche Gültigkeit verlassen kann, solange die Teststärke (Macht, Trennschärfe) des Tests unbekannt ist (Erklärung der Begriffe s.u.)!

5.1 Grundlagen statistischer Tests

Nullhypothese und Alternativhypothese

Eine statistische Hypothese ist eine Behauptung über eine Eigenschaft der Grundgesamtheit.⁷ Die Hypothese wird *Nullhypothese* H_0 genannt, wenn sie eine Behauptung über eine konkrete Eigenschaft der Grundgesamtheit enthält, die statistisch getestet werden kann. Die Nullhypothese wird mit der Absicht getestet, sie zurückzuweisen, sofern sich anhand einer Stichprobe ein Resultat ergibt, das stark gegen die Richtigkeit der Nullhypothese spricht (*proof by contradiction*)⁸.

Die Hypothese, gegen die die Nullhypothese getestet wird, heißt *Alternativhypothese* H_A . Im Gegensatz zur Nullhypothese muß die Alternativhypothese nicht zu einer eindeutigen, testbaren Aussage über die Grundgesamtheit führen. Die Alternativhypothese wird selbst nicht getestet; ihre Formulierung entscheidet jedoch darüber, in welcher Form die Nullhypothese getestet wird (einseitig oder zweiseitig).

Die Wertebereiche von Null- und Alternativhypothese dürfen sich nicht überlappen, es kann also nur eine der beiden Hypothesen angenommen werden. Die Alternativhypothese wird genau dann angenommen, wenn die Nullhypothese als Folge des Tests zurückgewiesen wird.

Ein- und zweiseitiger Test

Ein Test heißt *einseitig*, wenn der Wertebereich der Alternativhypothese zusammenhängend ist. Ansonsten ist ein Test *zweiseitig*.

Fehler 1. und 2. Art

Es gibt vier mögliche Konstellationen in einem Test:

1. Die H_0 ist richtig, und sie wird aufgrund des Tests nicht verworfen ($P = 1 - \alpha$).

⁷Formuliert als Annahme über die Verteilung einer Zufallsvariablen.

⁸Man beachte: Die Nicht-Zurückweisung einer Nullhypothese bedeutet somit keineswegs, daß sie richtig ist!

2. Die H_o ist richtig, wird aber aufgrund des Tests verworfen (Fehler 1. Art, $P = \alpha$).
3. Die H_o ist falsch, wird aber aufgrund des Tests nicht verworfen (Fehler 2. Art, $P = 1 - \beta$).
4. Die H_o ist falsch, und wird aufgrund des Tests tatsächlich verworfen (Macht des Tests, $P = \beta$).⁹

Während die erste und vierte Konstellation erwünscht ist, weil sie das richtige Ergebnis liefert, treten die Fehler 1. und 2. Art mit den Wahrscheinlichkeiten α und β auf.

Signifikanz und Macht

Die *Aussagesicherheit* oder *Signifikanz* eines Tests ist die Wahrscheinlichkeit, mit der ein Irrtum erster Art in Kauf genommen wird. Wegen der Methodik des Tests kann die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, durch die testende Person frei gewählt werden.

Die Wahrscheinlichkeit β , H_o abzulehnen wenn die H_A richtig ist, wird *Macht* (Teststärke, Trennschärfe) eines Tests genannt. Ist die Signifikanz α eines Tests festgelegt, so ergibt sich die Wahrscheinlichkeit $(1 - \beta)$ eines Fehlers 2. Art aus der Formulierung der Alternativhypothese, dem Umfang der Stichprobe, der Streuung der Daten und der Seitigkeit des Tests. Ist die Alternativhypothese konkret formuliert (so daß sie auch als Nullhypothese getestet werden könnte), so ist es möglich, den Fehler 2. Art zu quantifizieren.

Bei kleineren Stichprobenumfängen und kleinem α ist die Möglichkeit, tatsächlich vorhandene Unterschiede nachzuweisen, gering. Das Ergebnis "*Es wurde kein signifikanter Unterschied gefunden*" muß dann mit Vorsicht beurteilt werden.

Durchführung eines Tests

Die Durchführung von statistischen Tests folgt einem generellen Schema:

1. Formuliere die Nullhypothese H_o .
2. Formuliere die Alternativhypothese H_A und entscheide über die Seitigkeit des Tests.
3. Suche den geeigneten Test für das zu untersuchende Problem.
4. Prüfe die Voraussetzungen des vorgesehenen Verfahrens.
5. Wähle ein Signifikanzniveau (i.d. Regel 5%).
6. Berechne aus den Stichprobenwerten den Wert t_{exp} (z_{exp} , χ_{exp}^2 , F_{exp}) der Zufallsvariablen (T , Z , χ^2 , F), die getestet werden soll.
7. Suche aus der tabellierten Verteilungsfunktion der Zufallsvariablen (t -, z -, χ^2 - oder F -Verteilung) in Abhängigkeit von der Anzahl der Freiheitsgrade FG und der Seitigkeit des Tests einen Grenzwert (z.B. $t_{\alpha, FG}$), der mit der gewählten Irrtumswahrscheinlichkeit α überschritten wird.
8. Verwerfe die H_o , sofern der berechnete Wert t_{exp} der Zufallsvariablen größer als der Grenzwert $t_{\alpha, FG}$ und somit zu "unwahrscheinlich" erscheint.

⁹Beachte: die Notation für die Macht des Tests ist leider nicht einheitlich, Manchmal wird $(1 - \beta)$ für die Macht verwendet. Die hier verwendete Notation richtet sich nach [16].

Sofern eine Entscheidung zwischen zwei konkreten Hypothesen zu treffen ist, sollte die Macht des Testes berechnet werden. Nur so kann entschieden werden, ob die Akzeptanz der H_o tatsächlich auf sicheren Beinen steht.

5.2 Parametrische Tests

5.2.1 t-Test

zum Vergleich eines Mittelwerte mit einem theoretischen Wert

Fragestellung:

Weicht der experimentell gefundene Mittelwert \bar{x} der Stichprobe signifikant von einem angenommenen Mittelwert der Grundgesamtheit μ ab?

Voraussetzungen:

Die Grundgesamtheit, aus der die Stichprobe stammt, sei normalverteilt mit dem unbekanntem Mittelwert μ . Die gemessenen Daten sind mindestens intervallskaliert.

Rechenweg:

1. Berechne $t_{exp} = \frac{|\bar{x} - \mu|}{s} \sqrt{n}$.
2. Lese $t_{\alpha, FG}$ aus der t -Tabelle ab.
3. Verwerfe die H_o (" $\bar{x} = \mu$ "), wenn $t_{exp} \geq t_{\alpha, FG}$.

5.2.2 t-Test

zum Vergleich der Mittelwerte unabhängiger Stichproben

Fragestellung:

Sind die Mittelwerte \bar{x} und \bar{y} zweier Stichproben X und Y signifikant verschieden?

Voraussetzungen:

Die Grundgesamtheiten, aus denen die Stichproben stammen, sind beide normalverteilt. Die gemessenen Daten sind mindestens intervallskaliert. Die Varianzen sind unbekannt, aber gleich.

Rechenweg:

1. Berechne $t_{exp} = \frac{|\bar{x} - \bar{y}|}{s_o} \cdot \sqrt{\frac{n_x n_y}{n_x + n_y}}$
mit $s_o = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}$.
2. Lese $t_{\alpha, FG}$ aus der t -Tabelle ab, wobei die Zahl der Freiheitsgrade FG durch $n_x + n_y - 2$ berechnet wird.

3. Verwerfe die H_0 (“ $\mu_x = \mu_y$ ”), wenn $t_{exp} \geq t_{\alpha,FG}$.

5.2.3 t-Test

zum Vergleich der Mittelwerte verbundener Stichproben

Fragestellung:

Sind die Mittelwerte \bar{x} und \bar{y} zweier verbundener Stichproben X und Y signifikant verschieden?

Voraussetzungen:

Die Stichproben seien verbunden, die Meßwerte mindestens intervallskaliert. Die n Differenzen seien normalverteilt mit dem unbekanntem Mittelwert $\Delta\mu$.

Rechenweg:

1. Berechne $t_{exp} = \frac{|\bar{d}|}{s_d} \cdot \sqrt{n}$

mit der mittleren Differenz $\bar{d} = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)$

und der Stabw. der Differenzen $s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n [\bar{d} - (x_i - y_i)]^2}$.

2. Lese $t_{\alpha,FG}$ mit $FG = n - 1$ aus der t -Tabelle ab.
3. Verwerfe die H_0 (“ $\Delta\mu = 0$ ”), wenn $t_{exp} \geq t_{\alpha,FG}$.

5.2.4 t-Test

zur Prüfung des Korrelationskoeffizienten

Fragestellung:

Ist der aus n Wertepaaren ermittelte Korrelationskoeffizient r signifikant verschieden von Null?

Voraussetzungen:

Die Stichproben X und Y entstammen normalverteilten Grundgesamtheiten.

Rechenweg:

1. Berechne $t_{exp} = \frac{|r|}{\sqrt{1-r^2}} \cdot \sqrt{n-2}$.

2. Lese $t_{\alpha,FG}$ mit $FG = n - 2$ aus der t -Tabelle ab.
3. Verwerfe die H_0 (“ $\rho = 0$ ”), wenn $t_{exp} \geq t_{\alpha,FG}$.

5.2.5 χ^2 -Anpassungstest zum Vergleich von beobachteten mit erwarteten Häufigkeiten

Fragestellung:

Weichen n beobachtete Häufigkeiten f_i einer Stichprobe signifikant von erwarteten Häufigkeiten ϕ_i einer vermuteten Verteilung ab?

Voraussetzungen:

Es genügen nominalskalierte Daten. Für stetige Variablen müssen Klassen gebildet werden. Bei bivariaten Datensätzen kann mit den Häufigkeiten f_{ij} bzw. ϕ_{ij} gerechnet werden.

Rechenweg:

1. Berechne zu den n beobachteten Häufigkeiten f_i die zugehörigen theoretischen (erwarteten) Häufigkeiten ϕ_i und berechne die Summe der normierten quadratischen Differenzen

$$\chi_{exp}^2 = \sum_{i=1}^n \frac{(f_i - \phi_i)^2}{\phi_i} = \sum_{i=1}^n \frac{f_i^2}{\phi_i} - N \quad .$$

mit $N =$ Stichprobenumfang (Anzahl aller Messungen).

2. Lese $\chi_{\alpha, FG}^2$ mit $FG = n - 1 - a$ aus der χ^2 -Tabelle ab, wobei a die Anzahl Parameter angibt, die aus der Stichprobe geschätzt werden mußten, um die erwarteten Häufigkeiten zu berechnen.
3. Verwerfe die H_0 ("Die beobachtete und die theoretische Häufigkeitsverteilung sind gleich"), wenn $\chi_{exp}^2 \geq \chi_{\alpha, FG}^2$.

5.2.6 Kolmogorov-Smirnov-Anpassungstest zum Vergleich von beobachteten mit erwarteten Verteilungen

Fragestellung:

Entstammt eine beobachtete Verteilungsfunktion $\tilde{F}(x_i)$ einer theoretischen Wahrscheinlichkeitsverteilung $F(x)$?

Voraussetzungen:

Die Verteilungsfunktion der Grundgesamtheit ist stetig, die Daten mindestens intervallskaliert.

Rechenweg:

1. Berechne zu den Werten $\tilde{F}(x_i)$ der beobachteten Verteilungsfunktion die zugehörigen theoretischen Werte $F(x_i)$.
2. Bestimme die innerhalb des Definitionsbereiches der Verteilung auftretende Maximalabweichung $\Delta_{exp} = \sup |\tilde{F}(x) - F(x)|$.
3. Entnehme der Kolmogorov-Smirnov-Verteilung für das Signifikanzniveau α und den Stichprobenumfang n den Wert $\Delta_{\alpha, n}$.

4. Verwerfe die H_0 (“Die beobachtete Verteilung entstammt der theoretischen Wahrscheinlichkeitsverteilung”), wenn $\Delta_{exp} \geq \Delta_{\alpha,n}$.

5.3 Nichtparametrische Tests

Parametrische Tests setzen grundsätzlich Annahmen über die Verteilung der Grundgesamtheit voraus, oft z.B. normalverteilte Daten. Die Prüfung dieser Annahmen wiederum ist manchmal – je nach Datenbeschaffenheit – nur mit einer großen Unsicherheit möglich.

Verteilungsfreie oder *nichtparametrische* Tests sind für Situationen gefunden worden, in denen diese Annahmen unsicher sind oder offensichtlich nicht zutreffen. Man spricht in diesem Zusammenhang auch von *robusten* Verfahren.

Die *Vorteile* verteilungsfreier Tests sind

- meist leichter zu verstehen und oft einfach zu berechnen.
- können in Situationen verwendet werden, wo klassische parametrische Tests nicht angewendet werden dürfen.
- besonders wenn Daten niedrig (sprich nicht intervall-) skaliert sind, sind parameterfreie Tests angebracht.

Die *Nachteile* sind

- sie ignorieren oft den tatsächlichen numerischen Wert eines Parameters, d.h. sie verschwenden Information (Effizienzkriterium).
- Die *Macht* der parameterfreien Tests ist kleiner, H_0 wird also eher irrtümlicherweise beibehalten als bei parametrischen Tests.

In der Literatur werden eine Vielzahl von verteilungsfreien Verfahren beschrieben. Im folgenden werden einige wenige parameterfreie Tests für Vergleiche von Mittelwerten vorgestellt. Für einen besseren Einblick in diese Verfahren und für eine Vielzahl von weiteren existierenden Methoden sei auf die spezifische Literatur verwiesen ([4, 11, 24]).

5.3.1 Vorzeichentest (Two Sample Sign Test) zum Vergleich zweier Mittelwerte verbundener Stichproben

Der “*Vorzeichentest (Two Sample Sign Test)*” für Paardifferenzen benutzt nicht die numerischen Werte der Daten, sondern die Vorzeichen *plus* und *minus* als Testgrundlage. Dies ist nicht gerade besonders effizient. Der Test hat allerdings den Vorteil, daß er bereits bei der Datensichtung “mit dem bloßen Auge” durchgeführt werden kann.

Fragestellung:

Sind die Werte x_i und y_i zweier verbundener Stichproben X und Y signifikant verschieden?

Voraussetzungen:

Die Stichproben seien verbunden, die Meßwerte mindestens ordinalskaliert.

Rechenweg:

1. Weise jedem Datenpaar, für das unterschiedliche Werte vorliegen, ein “+” oder ein “-” Zeichen zu. Addiere die “+” und “-” Zeichen, um die Zahl n der in die Berechnung eingehenden Datenpaare zu ermitteln. Die Zahl der Pluswerte sei k .
2. Berechne den Wert

$$z_{exp} = \left| \left(\frac{k}{n} - \frac{1}{2} \right) \sqrt{4n} \right| .$$
3. Lese für das Signifikanzniveau α und Zweiseitigkeit des Tests den Wert $z_{(\alpha/2)}$ aus Tabelle der Standardnormalverteilung ab.
4. Verwerfe die H_0 (“ $\frac{k}{n} = \frac{1}{2}$ ”), wenn $z_{exp} \geq z_{\alpha/2}$.

Literatur:

[13], S. 420–426.

5.3.2 Wilcoxon-Test zum Vergleich zweier Mittelwerte verbundener Stichproben

Der *Wilcoxon-Test für Paardifferenzen* entspricht in der Fragestellung dem Vorzeichenstest, ist allerdings effizienter, da er die Größe der Differenzen berücksichtigt.

Fragestellung:

Sind die Mediane zweier verbundenen Stichproben X und Y signifikant verschieden?

Voraussetzungen:

Die beiden Grundgesamtheiten sollen stetige Verteilungen von gleicher Form haben. Die Stichproben seien verbunden und die Daten mindestens ordinalskaliert.

Rechenweg:

1. Berechne die Meßwertdifferenzen $d_i = x_i - y_i$. Die Differenzen $d_i = 0$ bleiben unberücksichtigt, es bleiben n Differenzen $d_i \neq 0$ zu betrachten.
2. Bringe die n Meßwertdifferenzen d_i entsprechend ihrer Absolutbeträge $|d_i|$ in eine aufsteigende Rangfolge mit den Rängen $R(d_i)$.
3. Berechne die Summe W^+ über die Rangzahlen $R(d_i)$ aller positiven Meßwertdifferenzen $d_i > 0$ und entsprechend die Summe W^- der $R(d_i)$ aller negativen Differenzen $d_i < 0$.
4. Die *kleinere* der beiden Größen wird als W_{exp} bezeichnet.
5. Verwerfe die H_0 (“*Die Mediane sind gleich*”), wenn $W_{exp} \leq W_{(n,\alpha)}$, wobei $W_{(n,\alpha)}$ aus der “*W-Tabelle*” entnommen wird.

Literatur:

[14], S. 103.

5.3.3 *U*-Test zum Vergleich zweier Mittelwerte unverbundener Stichproben

Der *U*-Test von *Mann und Whitney* (Rangsummen-Test) vergleicht die Mittelwerte zweier Grundgesamtheiten mit *unterschiedlichen* Varianzen (der *t*-Test kann nicht verwendet werden).

Fragestellung:

Sind die Mediane zweier unabhängiger Stichproben X_1 und X_2 verschieden?

Voraussetzungen:

Die Grundgesamtheiten sind stetige Verteilungen von gleicher Form. Die Stichproben sind unabhängig, die Daten mindestens ordinalskaliert.

Rechenweg:

1. Bringe die $(n_1 + n_2)$ Stichprobenwerte in eine *gemeinsame* Reihenfolge und berechne die Summen R_1 und R_2 der Rangzahlen der Stichproben X_1 und X_2 .
2. Berechne U_1 und U_2 entsprechend

$$U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - R_2$$

Probe: $U_1 + U_2 = n_1 \cdot n_2$.

3. Die *kleinere* der beiden Größen wird als U_{exp} bezeichnet.
4. Verwerfe die H_0 ("Die Mediane sind gleich"), wenn $U_{exp} \leq U_{(n_1, n_2, \alpha)}$, wobei $U_{(n_1, n_2, \alpha)}$ aus der "*U*-Tabelle" entnommen wird.

Literatur:

[14], S. 101; [22], S. 381.

6 Analysen

6.1 Korrelationsanalyse

- **Typisierung korrelativer Zusammenhänge**

Formale Korrelation

|
ja

|
nein

⇓

Inhomogenitätskorrelation

|
ja

|
nein

⇓

Gemeinsamkeitskorrelation

|
ja

|
nein

⇓

Kausale Korrelation

- **Gesamte, erklärte und unerklärte Streuung**

Die gesamte Quadratsumme Q_y der gemessenen y -Werte

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

läßt sich in zwei Anteile zerlegen:

- Der Quadratsummen-Anteil **um** die Regressionsgerade ist

$$Q_e^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

mit y_i = Meßwert an der Stelle x_i , und \hat{y}_i = aus der empirischen Regressionsgeraden berechneter Wert an der Stelle x_i .

- Der Anteil $Q_{\hat{y}}^2$, der sich aus der Differenz der vorhergesagten Werte \hat{y}_i vom Gesamtmittel \bar{y} ergibt, ist

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

Diesen letzten Anteil nennen wir die *Erklärte Quadratsumme*. Es zeigt sich nun, daß sich die Gesamtstreuung gerade als *Summe* der erklärten und der unerklärten Streuung ergibt!

- **Güte der Anpassung: Bestimmtheitsmaß r^2**

Um ein allgemeines Maß für die Güte der Anpassung einer Regressionskurve (die Gerade stellt lediglich einen Spezialfall dar) anzugeben, liegt es nahe, dafür das Verhältnis zwischen erklärter und gesamter Varianz zu benutzen.

Die so definierte Größe heißt *Bestimmtheitsmaß* B

$$B = r^2 = \frac{Q_{\hat{y}}^2}{Q_y^2}.$$

Im Fall der linearen Regression ist B stets gerade das Quadrat des Korrelationskoeffizienten r , und wird dann als r^2 angegeben.

• **Test der Signifikanz der Korrelation;** ($H_0: \rho = 0$)

Der Korrelationskoeffizient der Stichprobe kann als (zufallsbehaftete) Realisierung eines Korrelationskoeffizienten ρ der Grundgesamtheit angesehen werden. Es ist dann zu prüfen, ob der beobachtete Zusammenhang zwischen den Variablen tatsächlich existiert.

Um dies testen zu können, muß die Voraussetzung erfüllt sein, daß *beide* Variablen Zufallsvariablen sind und normalverteilte Grundgesamtheiten besitzen, und daß die Anzahl der Meßwerte nicht klein ist (Die Betrachtungsweise ist also anders als bei der Regressionsanalyse, s.u.). *Ein Test ohne Beachtung dieser Voraussetzung ist wertlos.* Der Test selbst ist (wie immer) sehr einfach:

Die Testvariable $T = r \sqrt{\frac{n-2}{1-r^2}}$ ist t -verteilt mit $n-2$ Freiheitsgraden.

Fragestellung:

Ist der Korrelationskoeffizient der Grundgesamtheit ρ signifikant verschieden von Null?

Voraussetzungen:

Die beiden Grundgesamtheiten sollen stetig und normalverteilt sein.

Rechenweg:

1. Berechne $t_{exp} = r \sqrt{\frac{n-2}{1-r^2}}$.
2. Verwerfe die H_0 (" $\rho = 0$ "), wenn $t_{exp} \geq t_{\alpha/2, n-2}$, wobei der tabellierte Wert der einseitigen t -Tabelle entnommen wird (Bei Ablesen aus der zweiseitigen t -Tabelle muß natürlich der Wert $t_{\alpha, n-2}$ genommen werden).

6.2 Regressionsanalyse

Zugrundeliegendes lineares Modell

1. Die Werte der unabhängigen Variablen sind fest: x ist keine Zufallsvariable
2. Für jedes x ist der zugehörige Y -Wert eine normalverteilte Zufallsvariable mit dem Mittelwert $\mu_{Y|x} = \alpha + \beta x$, wobei α und β konstant sind¹⁰. Die Gerade $\mu_{Y|x} = \alpha + \beta x$ heißt *Regressionsgerade der Grundgesamtheit*.
3. Die Varianz der Y -Werte Verteilung der Grundgesamtheit ist an jeder Stelle x gleich groß (*Homoskedastizität*). Wir bezeichnen sie mit σ_e^2 .

Auf der Basis dieses Modells umfaßt die statistische Regressionsanalyse folgende Probleme:

1. Berechnung von Punktschätzungen für die Parameter α , β und $\mu_{Y|x}$.
2. Abschätzung der Varianz σ_e^2 .
3. Berechnung von Konfidenzintervallen für die Parameter α , β und $\mu_{Y|x}$.
4. Testen von Hypothesen zu diesen Parametern.

Lösungen der Probleme

ad 1) Punktschätzungen:

Es kann gezeigt werden, daß die nach der Methode der kleinsten Quadrate berechnete Regressionsgerade der Stichprobe eine erwartungstreue Schätzung der Populations-Regressionsgeraden ist. Die Punktschätzungen für α , β und $\mu_{Y|x}$ sind durch a , b und $\hat{y}(x)$ gegeben.

ad 2) Varianzabschätzung:

Die an jeder Stelle x gleiche Varianz σ_e^2 der Zufallsvariablen Y wird durch die Streuung s_e^2 der gemessenen y -Werte **um** die Regressionsgerade abgeschätzt:

$$s_e^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y})^2$$

(Es wird durch $n-2$ geteilt, weil wir durch die Schätzung der Parameter A und B im obigen Ausdruck zwei Freiheitsgrade verloren haben. Der Index e kommt von "Error")

In der Praxis wird s_e^2 durch folgenden Ausdruck berechnet:

$$s_e^2 = \frac{1}{n-2} \left[\sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i \right]$$

Die Standardabweichung s_e heißt *Standardfehler der Regressionsgeraden*¹¹.

¹⁰Im folgenden wird α sowohl als Koeffizient der Regressionsgerade als auch zur Angabe der Irrtumswahrscheinlichkeit benutzt. Dies ist von der Nomenklatur her nicht sauber, sollte aber zu keinen Verwechslungen führen.

¹¹In EXCEL: STFEHLERYX(y;x)

ad 3) **Konfidenzintervalle**

Die Berechnung von Konfidenzintervallen und die Tests beruhen auf der Kenntnis der Verteilungen und Varianzen der einzelnen Parameter. Da wir die Varianzen der Grundgesamtheit nicht kennen, müssen wir sie aus Stichprobenwerten abschätzen. Die zugehörige Testverteilung ist deshalb Students t -Verteilung.

Aus der Schätzung von s_e^2 lassen sich die Varianzschätzungen für die Wahrscheinlichkeitsverteilungen der Parameter α , β und des prognostizierten Wertes $Y(x_0)$ ableiten:

$$s_a^2 = \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] s_e^2$$

$$s_b^2 = \frac{s_e^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$s_{y_o}^2 = \left[\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] s_e^2$$

HINWEIS: Die Varianz der Vorhersage der Zufallsvariablen Y ist für verschiedene x -Werte unterschiedlich groß; Sie ist an der Stelle $x_o = \bar{x}$ am kleinsten, und wird nach beiden Seiten zunehmend (quadratisch) größer !

Für die Konfidenzintervalle der Parameter bzw. eines geschätzten Wertes auf der Regressionsgerade gilt

$$\begin{pmatrix} \text{Parameter-} \\ \text{Punkt-} \\ \text{Schätzung} \end{pmatrix} \pm \begin{pmatrix} \text{Zu } \gamma \text{ gehöriger} \\ \text{Wert der} \\ \text{t-Verteilung} \end{pmatrix} \cdot \begin{pmatrix} \text{Geschätzte} \\ \text{Standard-} \\ \text{Abweichung} \end{pmatrix}$$

Daraus folgt mit $\gamma = 1 - \alpha$:

- Ein $(1 - \alpha) \cdot 100$ Prozent Intervall für α ist

$$a \pm t_{n-2, \alpha/2} \cdot s_a$$

- Ein $(1 - \alpha) \cdot 100$ Prozent Intervall für β ist

$$b \pm t_{n-2, \alpha/2} \cdot s_b$$

- Ein $(1 - \alpha) \cdot 100$ Prozent Intervall für $\mu_{Y|x_o}$ ist

$$a + bx_0 \pm t_{n-2, \alpha/2} \cdot s_{y_o}$$

ad 4) **Test der Geradensteigung B auf einen bestimmten Wert B_o**

Die Testvariable $T = \frac{b - B_o}{s_b}$ ist t -verteilt mit $n - 2$ Freiheitsgraden:

Fragestellung:

Ist die Geradensteigung B der Grundgesamtheit signifikant verschieden von einer Steigung B_o ?

Voraussetzungen:

Die Grundgesamtheiten entsprechen dem linearen Modell I (s.o.).

Rechenweg:

1. Berechne $t_{exp} = \frac{b - B_o}{s_b}$.
2. Verwerfe die $H_o : ("B = B_o'')$, wenn $t_{exp} \geq t_{n-2, \alpha/2}$.

Abschließende Bemerkungen

Welche Analyse ist an welche Voraussetzungen gebunden ?

Korrelationskoeffizienten und Regressionskurven lassen sich an beliebigen Daten berechnen. Bei ihrer Interpretation, d.h. bei der Durchführung der Regressions- und Korrelationsanalyse, muß man jedoch unbedingt die damit verknüpften Modellvorstellungen beachten!

Bei der behandelten *Regressionsanalyse* wird die unabhängige Variable als fest vorgegeben betrachtet, während die abhängige normalverteilt ist. Ist X ebenfalls eine Zufallsvariable, so ist das modifizierte Modell II anzuwenden (s. Literatur). Wird die Regressionsgleichung zu Vorhersagen benutzt, so ist dies - wie bei allen statistischen Verfahren - nur innerhalb des Wertebereichs zulässig, der durch Stichprobenwerte abgedeckt ist. Streuen die Daten in beträchtlichem Maß, so sollten die Konfidenzbereiche für Y berechnet werden.

Ist bei einer *Signifikanzprüfung des Korrelationskoeffizienten* nicht sicher, ob die Stichprobendaten normalverteilten Grundgesamtheiten entstammen, so ist ein entsprechender Test durchzuführen. Nicht normalverteilte Daten können manchmal zur Normalverteilung transformiert werden; in diesem Fall ist mit den transformierten Daten die Korrelationsanalyse durchzuführen. Sollte eine Normalverteilung auch nicht angenähert erreichbar sein, so muß mit verteilungsfreien Verfahren gearbeitet werden (z.B. Rangkorrelation).

Weiterhin sollte erkannt werden, daß ein *wichtiger* und ein *signifikanter* Effekt nicht immer dasselbe ist: bei genügend großen Stichproben können auch schwache Effekte (also x -Werte, die die y -Werte nur schwach bestimmen) als signifikant nachgewiesen werden. Umgekehrt gilt bei kleinen Stichproben, daß selbst ein relativ hoher Wert von r manchmal nicht als signifikant eingestuft wird. In diesem Fall sollte der Stichprobenumfang erhöht werden, um den Effekt statistisch zu belegen.

Abschließend muß noch betont werden, daß eine signifikante Korrelation grundsätzlich nicht bedeutet, daß eine Ursache-Wirkung Beziehung besteht. Es ist nur ein Hinweis darauf, daß eine solche Verknüpfung bestehen kann. Dies zu Prüfen ist aber die Sache des untersuchenden Wissenschaftlers und kann von der Statistik nicht übernommen werden!

6.3 Varianzanalyse

- **Grundgedanke**

- **Frage:** Hat die Variation eines (Versuchs)parameters einen signifikanten Einfluß auf das Ergebnis ?
- **Prinzip:** Unterscheidung in feste und zufällige Effekte (z.B. Zufällige Meßfehler — Bewußte Variation eines Versuchsparameters)
- **Vorgehensweise:**
Einteilung der Daten in m Stichproben mit den Stichprobenumfängen n_i . (Für jede *Faktorstufe* eine Stichprobe).
- Zerlegung der Gesamtvarianz nach Streuungsursachen:

$$\text{Varianz innerhalb der SP } s^2 = \overline{s}_{i,innerh.}^2 = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (x_{i,j} - \overline{x}_i)^2}{\sum_{i=1}^m (n_i - 1)}$$

$$\text{Varianz zwischen den SP } s_{zwischen}^2 = \frac{1}{m - 1} \sum_{i=1}^m (\overline{x}_i - \overline{\overline{x}})^2$$

(\overline{x}_i sind die SP-Mittelwerte, $\overline{\overline{x}}$ ist der Gesamtmittelwert.)

- Prüfung der Varianzen auf Gleichheit durch F -Test

- **Testdurchführung:**

H_0 : Die Mittelwerte $\mu_1, \mu_2, \dots, \mu_m$ sind alle gleich

H_A : Wenigstens einer der Mittelwerte ist verschieden

Test mit der Signifikanz α durch Vergleich der Varianz *zwischen* den Gruppen mit der *internen* Varianz: Das Verhältnis muß größer als ein tabellierter F -Wert sein.

$$f_{exp} = \overline{n} \frac{s_{zwischen}}{s_{innerh.}}$$

Verwerfe H_0 wenn $f_{exp} \geq F_{(m-1, \sum (n_i-1), \alpha)}$ (Bem.: Für Werte < 1 wird H_0 in jedem Fall beibehalten. Einseitiger Test).

- **Test-Voraussetzungen**

- Die m Stichproben stammen aus m normalverteilten Grundgesamtheiten.
- Die m Grundgesamtheiten haben diesselbe Varianz; dabei kann σ^2 unbekannt sein.
- Die m Stichproben seien unabhängig.

- **Hinweis**

Bei der Planung eines Versuches sollte man darauf achten, für jede Faktorstufe die *gleiche Anzahl von Wiederholungen* zu haben ($\overline{n} = n_1 = n_2 = \dots = n_m$). Eine solche Versuchsanordnung heißt *balanciert* und hat für statistische Auswertungen viele Vorteile.

7 Chemometrische Anwendungen

7.1 Präzision, Richtigkeit und Genauigkeit von Meßverfahren

Präzision

Die Präzision (precision) eines Verfahrens kann durch den zufälligen Meßfehler beschrieben werden. Ist die Varianz bzw. die Standardabweichung als Maß für den zufälligen Fehler sehr klein geworden, dann hat das Verfahren eine hohe Präzision erreicht. Als Maß für die Präzision kann eine inverse Funktion der Standardabweichung z.B. $1/\sigma$ verwendet werden. Je größer der Funktionswert ist, desto präziser arbeitet das Meßverfahren.

Richtigkeit (accuracy of the mean; bias)

Bei der Untersuchung von Richtigkeit eines Meßverfahrens steht die Frage über den systematischen Fehler des Verfahrens im Vordergrund. Im Unterschied zur Präzision, die direkt im Zusammenhang mit der Streuung von Meßergebnissen steht, ist die Richtigkeit eines Verfahrens danach zu beurteilen, wie weit der Mittelwert der Meßergebnisse vom "wahren" Wert entfernt ist. Dieser wahre Wert ist als theoretische Größe immer unbekannt. Er kann beispielsweise als Naturkonstante gegeben sein, empfohlen sein, oder als hypothetischer Wert vorgegeben sein. Demnach ist die Richtigkeit vorerst nicht quantifizierbar. Die Richtigkeit eines Meßverfahrens kann indirekt überprüft werden durch:

- Parallelbestimmungen mit einem anderen Meßverfahren
- Zusatzversuche (in der Analytik durch Hinzufügen bekannter Substanzmengen zur Ursprungssubstanz)
- Mischversuche (in der Analytik durch Mischen einer zu untersuchenden; Substanz in verschiedenen Konzentrationen oder Mengen)
- Ringversuche

Genauigkeit

Unter Genauigkeit eines Meßverfahrens verstehen *Funk et al.* [10] die *gesamte Abweichung* eines Meßverfahrens. Da sich die Richtigkeit eines Meßverfahrens nur am systematischen Fehler orientiert, und die Präzision eines Verfahrens am zufälligen Fehler, ist die Frage offen, wie Meßverfahren insgesamt zu bewerten sind.

Einfach ist die Frage zu beantworten, wenn beide Meßverfahren die gleiche Richtigkeit besitzen. In diesem Fall wird das Verfahren mit der größeren Präzision gewählt. Umgekehrt wird bei zwei Verfahren mit der gleichen Präzision das Verfahren mit der größeren Richtigkeit zu wählen sein.

Schwieriger ist die Frage zu beurteilen, wenn ein Meßverfahren eine hohe Präzision und eine geringe Richtigkeit hat und das andere eine geringe Präzision aber eine hohe Richtigkeit. In diesem Fall kann die Wahl des Verfahrens von der jeweiligen Anwendung abhängen. Generell ist zu berücksichtigen, daß ein Verfahren mit hoher Streuung oft nur relativ schwer weiter präzisiert werden kann. Hingegen besteht bei einem systematischen Fehler häufig die

Möglichkeit der Justierung. In diesem Fall sollte der Präzision das größere Gewicht gegeben werden.

Empfindlichkeit, Sensitivität

Unter der *Empfindlichkeit* oder *Sensitivität* eines Analysenverfahrens versteht man die Änderung des Signalwerts pro Änderung des Gehalts, d.h. die Steigung der Eichgeraden. Die Empfindlichkeit geht ein in die Definition der Verfahrensstandardabweichung (Verhältnis von Streuung zu Empfindlichkeit), und damit in die Berechnung der Nachweisgrenze.

7.2 Statistische Beurteilung von Analyseverfahren

(Nach [8], S. 64 ff.) Die überwiegende Zahl der in der Wasseranalytik benutzten Analyseverfahren sind eichbedürftige Relativverfahren, bei denen ein gemessenes Signal – der Informationswert y – über eine Eichfunktion in eine Konzentration x umgerechnet werden muß. Die Nachweisgrenze (NG) ist ein wichtiges Kennzeichen der Leistungsfähigkeit eines Analyseverfahrens. Sie ist gekennzeichnet durch den kleinsten Wert des Gehaltes einer Substanz in einer Analysenprobe, für den die vorliegende Analysenmethode Signalwerte liefert, die sich mit 95% Wahrscheinlichkeit von solchen Signalwerten unterscheiden, die der Gehalt "Null" in der Analysenprobe liefert.

Eine Substanz ist dann nachgewiesen, wenn der gefundene Wert ihres Gehaltes im Untersuchungsmaterial größer als die Nachweisgrenze ist. Liegt der gefundene Wert unterhalb der Nachweisgrenze, aber noch unterhalb der Bestimmungsgrenze, so kann noch kein exakter Zahlenwert als Analyseergebnis angegeben, sondern bestenfalls die Größenordnung des Gehalts abgeschätzt werden [9].

Während man also bei Messung eines Wertes oberhalb der Nachweisgrenze davon ausgehen kann, daß sehr wahrscheinlich tatsächlich etwas in der Probe ist, führen umgekehrt sehr niedrige Gehalte in der Probe an der Grenze der Nachweisgrenze oft (ca. 50% der Fälle) dazu, daß das Signal als nicht signifikant verschieden vom Blindwert interpretiert wird. Daraus folgt die sinnvolle Definition der *Erfassungsgrenze*. Diese stellt ein höheres Konzentrationsmaß, ab dem man "on the save side" ist, d.h. ab dem das Fehlerrisiko, eine vorhandene Substanz nicht zu detektieren, ebenfalls sehr gering wird. Auf die Gegebenheiten der Qualitätssicherung im Bereich der Rückstandsanalytik abgestimmt ist schließlich der früher oft verwendete Begriff der *Bestimmungsgrenze*, der allerdings uneinheitlich gehandhabt wird.

Zur Bestimmung dieser statistischen Maße ist es nötig, die Empfindlichkeit und die Streuung des Meßverfahrens zu quantifizieren, sowie die Voraussetzungen der regressionsanalytischen Kalkulationen zu prüfen.

7.2.1 Prüfung der Varianzhomogenität von Kalibrierfunktionen

(Vereinfachtes Verfahren, aus [10], S. 32 ff.) Zur Überprüfung der Varianzhomogenität wird vorgeschlagen, jeweils zehn Standardproben der niedrigsten x_1 , und der höchsten Konzentration x_{10} des Arbeitsbereiches getrennt zu analysieren, wobei jeweils zehn Informationswerte $y_{i,j}$ erhalten werden ($n_i = 10$). Für beide Gruppen werden die Mittelwerte \bar{y}_1 und

\bar{y}_{10} sowie die Standardabweichungen s_1 und s_{10} berechnet. Mit Hilfe des einfachen Varianztests (F -Test) werden die Streuungen der Informationswerte an den Grenzen des Arbeitsbereiches auf signifikante Unterschiede hin untersucht.

1. Berechnung der Prüfgröße (F_{exp})

$$F_{exp} = \frac{s_{10}^2}{s_1^2}$$

(es wird davon ausgegangen, daß $s_{10} > s_1$; ansonsten werden Zähler und Nenner getauscht.)

2. Ablesen des tabellierten Wertes F_{tab} aus der F -Tabelle für $P = 99\%$, $FG_1 = n_1 - 1$, $FG_2 = n_{10} - 1$
3. Entscheidung: Wenn $F_{exp} > F_{tab}$, so ist der Unterschied zwischen den Varianzen signifikant.

Bei einem signifikanten Unterschied sollte der vorläufige Arbeitsbereich soweit eingengt werden, bis die Bedingung der Varianzhomogenität erfüllt ist. Ansonsten muß mit einer gewichteten Regression gearbeitet werden.

7.2.2 Prüfung der Linearität der Eichgerade

(aus [10], S. 34 ff.) Um bei der Prüfung auf Linearität auf die Durchführung von *Mehrfachanalysen* der Eichproben verzichten zu können, wird ein Anpassungstest nach Mandel [17] empfohlen. Dieser testet die durch Wahl einer anderen Ausgleichsfunktion (hier: Regressionsfunktion 2. Grades) erhaltene *Verringerung der Restvarianz* auf Signifikanz.

Testdurchführung

- Aus den vorliegenden Konzentrations- und Informationswerten zu mindestens $N = 10$ unterschiedlichen Eichproben werden berechnet:
 - a) die lineare Eichfunktion $y = a + bx$ samt Residualstandardabweichung s_{e1}
 - b) die quadratische Eichfunktion $y = a + bx + cx^2$ samt Residualstandardabweichung s_{e2} .
- Aus den Residualstandardabweichungen und s_{e1} und s_{e2} wird die Differenz Δs der Varianzen berechnet:

$$\Delta s^2 = (N - 2) \cdot s_{e1}^2 - (N - 3) \cdot s_{e2}^2$$

- Die Differenz wird im F -Test auf signifikanz geprüft:
 1. Berechnung der Prüfgröße (F_{exp})

$$F_{exp} = \frac{\Delta s^2}{s_{e2}^2}$$

2. Ablesen des tabellierten Wertes F_{tab} aus der F -Tabelle für $P = 99\%$, $FG_1 = 1$, $FG_2 = N - 3$

3. Entscheidung: Wenn $F_{exp} > F_{tab}$, so ist der Unterschied zwischen den Varianzen signifikant.

Ist der Unterschied nicht signifikant, so ist die Eichgerade linear.

Ist der Unterschied dagegen signifikant, so stellt die Anpassung der Parabel tatsächlich ein besseres Modell für die Daten dar: die Eichgerade ist signifikant nichtlinear. Erfahrungsgemäß kann in diesem Fall durch Einengung des Arbeitsbereichs eine Verbesserung erzielt werden. Es sollte, wenn dies gewünscht wird, in einem engeren Arbeitsbereich eine neue Eichung durchgeführt und der Linearitätstest wiederholt werden, anderenfalls muß mit einer nicht linearen Eichgerade gearbeitet werden (siehe [10], Kap. 3.4).

ANMERKUNG: Dieser Linearitätstest hat gegenüber dem einfachen Vergleich (F -Test) der beiden Restvarianzen s_{e1}^2 und s_{e2}^2 den Vorteil, daß die Prüfgröße *nicht* durch die unterschiedlichen Freiheitsgrade beeinflusst wird, was bei einer niedrigen Anzahl von Einzeldaten stark ins Gewicht fallen würde.

7.2.3 Vertrauensbereiche um die Eichgerade

In Kapitel 6.3 (Regressionsanalyse) wurden einige wichtige Größen definiert, die zur Berechnung der Nachweisgrenze benötigt werden: Aus der Streuung der Meßwerte um die Regressionsgerade, s_e , läßt sich die Varianz $s_{\hat{y}_i}^2$ eines einzelnen, über die Eichgerade vorhergesagten, Informationswerts y_i errechnen, und darauf basierend ein Konfidenzintervall $\hat{y}_i \pm \Delta \hat{y}_i$. Die Gleichungen hierzu sind:

Standardfehler der Regressionsgeraden:

$$s_e^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Streuung des vorhergesagten Informationswerts:

$$s_{\hat{y}_i}^2 = s_e^2 \cdot \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

Prognoseintervall eines vorhergesagten Informationswerts:

$$\hat{y}_i \pm \Delta \hat{y}_i = \hat{y}_i \pm t_{FG, \alpha/2} \cdot s_{\hat{y}_i}$$

mit $FG = n - 2$, wobei n die Anzahl der Meßpunkte beim Aufstellen der Eichgerade sind.

Erfolgt nun eine Messung der (unbekannten) Konzentration x_j (nicht notwendigerweise identisch zu einer der Eichmessungen x_i , und (aus mathematischer Sicht) nicht einmal notwendigerweise innerhalb des Wertebereichs der Eichung), so ergibt sich für die Varianz $s^2(\hat{y}_j)$ des zugehörigen Informationswerts

$$s^2(\hat{y}_j) = s_e^2 \cdot \left[1 + \frac{1}{n} + \frac{(x_j - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

Durch wiederholte Messung (n_a Messungen für eine Probe) läßt sich das Konfidenzband etwas verkleinern:

$$s^2(\hat{y}_j) = s_e^2 \cdot \left[\frac{1}{n_a} + \frac{1}{n} + \frac{(x_j - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

Trägt man die Konfidenzwerte für alle Werte x grafisch auf, so erhält man die leicht gekrümmten Konfidenzbänder (prediction intervals) um die Eichfunktion.

Für den Analyseschritt ist es nun nötig, aus den Konfidenzbändern für den Informationswert durch Invertierung Konfidenzintervalle für eine unbekannte Konzentration x_a zu berechnen.

Die Invertierung ergibt für den Vertrauensbereich von x bei vorgelegtem Signalwert y [8]:

$$\bar{x}_a \pm \Delta \bar{x}_a = \bar{x}_a \pm \frac{s_e \cdot t_{FG, \alpha/2}}{b} \sqrt{\frac{1}{n_a} + \frac{1}{n} + \frac{(\bar{x}_a - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

mit b = Steigung der Eichgeraden, und \bar{x} = Mittelwert der Konzentrationen bei der Eichung.

HINWEIS: Der Ausdruck $\frac{s_e}{b}$ wird übrigens *Verfahrensstandardabweichung* genannt und stellt eine wichtige Größe zur Charakterisierung der Güte des Meßverfahrens dar.

7.2.4 Nachweisgrenze

Ausgehend von der Eichgerade

$$y = a + bx$$

können Nachweisgrenze, Erfassungsgrenze und Bestimmungsgrenze ausgehend von einem kritischen Wert y_{crit} errechnet werden. Diese Werte sind von besonderem Interesse im Bereich der Spurenanalytik. Die Verfahren zur Errechnung der Kennwerte können hierbei unterteilt werden in solche, die darauf basieren, daß mehrere Blindwertmessungen durchgeführt, und statistisch ausgewertet werden, und solche, die zusätzlich die Informationen aus der Eichgeraden verwenden.

Nachweisgrenze: Methode der Blindwertstreuung

Werden n_b Versuche zur Bestimmung des Blindwerts durchgeführt, so lassen sich ein mittlerer Blindwert \hat{y}_{blank} und die zugehörige Standardabweichung s_{blank} berechnen. Der kritische Informationswert y_{crit} wird dann als obere Grenze des Vertrauensintervalls für \hat{y}_{blank} aus den Blindwertmessungen aufgefaßt:

$$y_{crit} = \hat{y}_{blank} + s_{blank} \cdot t_{FG, \alpha/2} \sqrt{\frac{1}{n_b}}$$

mit $FG = n_b - 1$.

Ist das Signal einer Analyse über diesem kritischen Wert, so läßt sich mit einer Irrtumswahrscheinlichkeit α sagen, daß es durch eine Konzentration in der Probe hervorgerufen wurde. Aufgrund dieser *qualitativen Entscheidung* kann nun leicht ausgerechnet werden, welcher Konzentration $x = NG$ dieser Informationswert entspricht:

$$NG = \frac{s_{blank}}{b} \cdot t_{FG, \alpha/2} \sqrt{\frac{1}{n_b}}$$

Entsprechend einem Vorschlag der DIN 32645 [1] kann der numerische Wert der Nachweisgrenze noch weiter abgesenkt werden, wenn angenommen wird, daß für die Analyse einer

Probe dieselbe Standardabweichung wie für die Analyse der Blindwerte besteht:

$$NG = \frac{s_{blank}}{b} \cdot t_{FG,\alpha/2} \sqrt{\frac{1}{n_b} + \frac{1}{n_a}}$$

Nachweisgrenze: Eichkurvenverfahren

Bei dieser Methode wird zur Berechnung des kritischen Werts die obere Konfidenzgrenze des Achsenabschnitts der Eichgerade benutzt [1]:

$$y_{crit} = a + s_e \cdot t_{FG,\alpha/2} \sqrt{\frac{1}{n} + \frac{1}{n_a} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

mit n = Anzahl der Messungen zur Aufstellung der Eichgeraden, n_a = Anzahl der Messungen der zu untersuchenden Probe, \bar{x}^2 = Quadrat des Mittelwerts der Eichkonzentrationen, und $FG = n - 2$. Einsetzen von y_{crit} in die Gleichung der Eichgeraden ergibt

$$NG = \frac{s_e}{b} \cdot t_{FG,\alpha/2} \sqrt{\frac{1}{n} + \frac{1}{n_a} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

7.2.5 Erfassungsgrenze

Für Konzentrationen an der Nachweisgrenze besteht eine 50%-Wahrscheinlichkeit, daß gemessene Werte im Streubereich des Blindwertes liegen. In anderen Worten: Das Risiko, einen Blindwert irrtümlich als Signal zu akzeptieren, ist zwar gering (namentlich = α), aber die Wahrscheinlichkeit für einen Fehler 2. Art, nämlich eine tatsächlich vorhandene Konzentration als nicht nachweisbar zu erklären, ist relativ hoch. Eine verlässliche Erfassung von Konzentrationen ist also erst möglich, wenn auch die Fehlerwahrscheinlichkeit β verkleinert wird. Für $\alpha = \beta$ ergibt sich in erster Näherung eine Verdoppelung der Nachweisgrenze. Dieser Wert wird mit dem Terminus *Erfassungsgrenze* [1] bezeichnet:

$$EG = 2 \cdot NG$$

7.2.6 Bestimmungsgrenze

Die Bestimmungsgrenze ist ein traditioneller Begriff, der uneinheitlich gehandhabt wurde. Nach [10] ergibt sich die Bestimmungsgrenze *BG* aus der Forderung, daß die obere Vertrauensbereichsgrenze des Blindwerts der unteren Vertrauensbereichsgrenze der quantifizierbaren Analysewerte entspricht. Dies würde der oben definierten Erfassungsgrenze entsprechen. Nach einem Vorschlag von Long und Winefordner (1983; zit. in [8]) wird die Bestimmungsgrenze analog zur Nachweisgrenze errechnet, wobei jedoch ein t -Wert von 10 (statt $t_{FG,\alpha/2} \approx 2$) eingesetzt wird.

“Die Bestimmungsgrenze ist eine pragmatisch festgelegte Größe (d.h. sie ist ohne Bezug zu Sicherheit und Risiko). Sie gibt die untere Gehaltsgrenze für einen maximal tolerierbaren relativen Zufallsfehler an. Bestimmungen unterhalb dieser Grenze sind mit einem unakzeptabel hohen Zufallsfehler belastet.”

7.3 Ringversuche

(Siehe hierzu HARTUNG, S. 342 ff.) Mit Hilfe von Ringversuchen können Meß- und Analysenmethoden auf systematische und zufällige Fehler hin untersucht werden.

Bei analytischen Fragestellungen in Ringversuchen geht man aus von einem Probenmaterial mit genau vorgegebenem Stoffgehalt¹².

Proben des Referenzmaterials werden an die teilnehmenden Laboratorien versandt und dort auf die vorgegebene Substanz hin quantitativ analysiert. Geht man davon aus, daß keine Meßfehler auftreten, so würden alle Laboratorien zu den gleichen Ergebnissen gelangen. Sind nun bei den meisten Teilnehmern die Analysen nahezu gleich und treten nur bei einigen wenigen stärkere Abweichungen auf, so wird man diese Abweichungen genauer auf systematische Fehler hin untersuchen. Ein Ringversuch dieser Art wird auch interlaboratorielle Vergleich genannt. Ringversuche können ebenfalls auf Vergleiche innerhalb von Laboratorien angewendet werden (intralaboratorielle Vergleich). Beispielsweise können die zu untersuchenden Analysen von verschiedenen Mitarbeitern vorgenommen werden. Ebenso denkbar ist die Durchführung der Analysen von verschiedenen Mitarbeitern an verschiedenen Tagen. Die Durchführung von Analysen in verschiedenen Laboratorien von verschiedenen Mitarbeitern an verschiedenen Tagen kann ebenfalls ausgewertet werden. In den letzten beiden Fällen spricht man von einem hierarchischen Aufbau des Ringversuchs.

¹²In den USA beispielsweise werden zertifizierte Referenzmaterialien mit genau eingestellten Stoffgehalten in den unterschiedlichsten Trägermaterialien bzw. Matrices vom National Bureau of Standards angeboten.

Literatur

- [1] DIN 32645. *Teil1: Chemische Analytik; Nachweis-, Erfassungs- und Bestimmungsgrenze; Ermittlung unter Wiederholbedingungen; Begriffe, Verfahren, Auswertung* 5. 1994.
- [2] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.
- [3] V. Barnett and T. Lewis. *Outliers in Statistical Data*. Wiley, Chichester, 2 edition, 1985.
- [4] J. Bortz, G.A. Lienert, and K. Boehnke. *Verteilungsfreie Methoden in der Biostatistik*. Springer, Berlin, 2 edition, 1990.
- [5] C. Chatfield. *Problem Solving. A Statistician's Guide*. Chapman and Hall, London, 1994.
- [6] V.T. Chow, D.R. Maidment, and L.W. Mays. *Applied Hydrology*. McGraw-Hill, New York, 1988.
- [7] K.L. Chung. *Elementary Probability Theory with Stochastic Processes*. Springer, New York, 3 edition, 1979.
- [8] W. Einax, H. W. Zwanziger, and S. Geiß. *Chemometrics in Environmental Analysis*. VCH, Weinheim, 1997.
- [9] Deutsche Forschungsgemeinschaft. *Methodensammlung zur Rückstandsanalytik von Pflanzenschutzmitteln*. VCH, Weinheim, 11. Lieferung, 1991.
- [10] W. Funk, V. Dammann, C. Vonderheid, and G. Oehlmann. *Statistische Methoden in der Wasseranalytik*. VCH, Weinheim, 1985.
- [11] M. Hollander and D.A. Wolfe. *Nonparametric Statistical Methods*. Wiley, New York, 1973.
- [12] D. Huff. *How to lie with Statistics*. Norton, New York, 1954.
- [13] R. Khazanie. *Elementary Statistics in a World of Applications*. Scott, Foresman and Company, Glenview, Illinois, 1979.
- [14] W. Köhler, G. Schachtel, and P. Voleske. *Biometrie. Einführung in die Statistik für Biologen und Agrarwissenschaftler*. Springer, Berlin, 1984.
- [15] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg Studium, Aufbaukurs Mathematik, Braunschweig, 3 edition, 1991.
- [16] E. Kreyszig. *Statistische Methoden und ihre Anwendungen*. Vandenhoeck und Rupprecht, Göttingen, 7 edition, 1979.
- [17] J. Mandel. *The statistical analysis of experimental data*. Intersc. Publ., J. Wiley & Sons, New York, 1964.
- [18] M. Nagel, K. D. Wernecke, and W. Fleischer. *Coputergestützte Datenanalyse*. Verlag Technik, Berlin, München, Mit MS-Dos Diskette für Statistikprogramm ISP, 1997.

- [19] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes*. Cambridge University Press, Cambridge, 2 edition, 1992.
- [20] W. Rechenberg. *Fresenius Z. Analyt. Chem.*, 311:590, 1982.
- [21] L. Ries. *Chemometrie*. Europa Fachhochschule Fresenius, Idstein, Skript, Ausgabe 4, September 1997.
- [22] L. Sachs. *Angewandte Statistik*. Springer, Berlin, 7 edition, 1992.
- [23] C.-D. Schönwiese. *Praktische Statistik für Meteorologen und Geowissenschaftler*. Bornträger, Stuttgart, 1985.
- [24] S. Siegel. *Nichtparametrische Statistische Methoden*. Fachbuchhandlung für Psychologie, Verlagsbuchhandlung, Frankfurt, 1976.
- [25] E.R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut, 1985.