

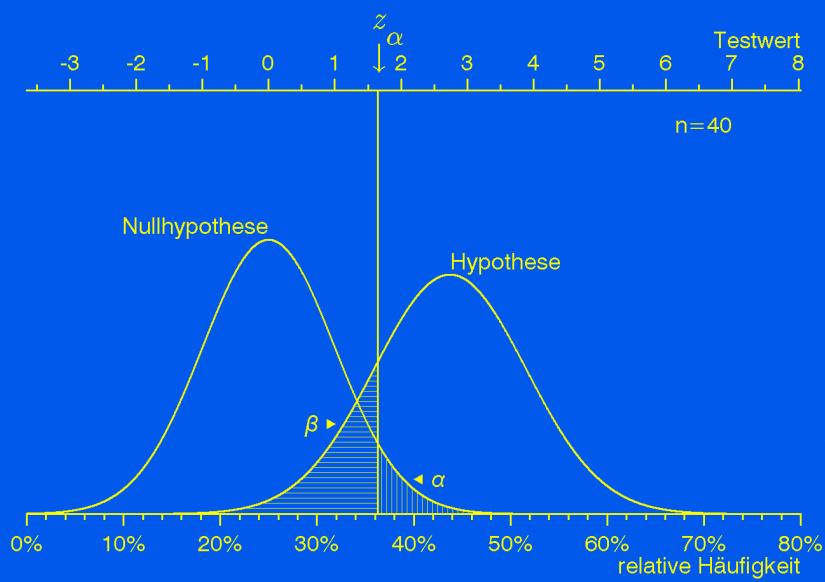
# Mathematik - Arbeitspapiere Nr. 59

## Teil B: Hochschuldidaktisches Material

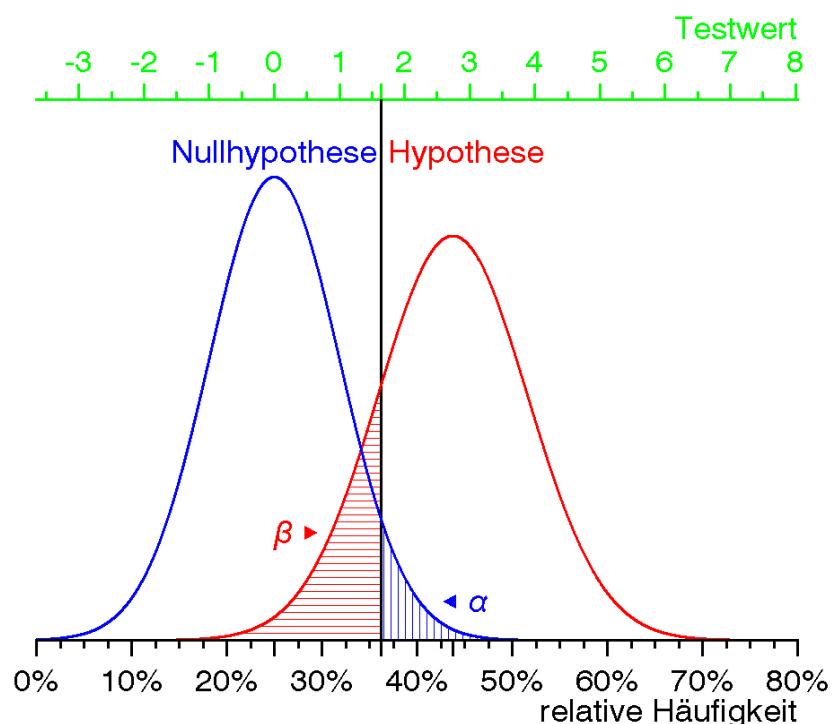
Skript zur

# Statistik in den Naturwissenschaften

Gerhard Osius



Skript zur  
**Statistik in den Naturwissenschaften**  
**Gerhard Osius**



**Mathematik Arbeitspapiere Nr. 59**  
Teil B: Hochschuldidaktisches Material  
Fachbereich Mathematik/Informatik  
Universität Bremen  
April 2007

*Anderson's law:<sup>1</sup>*

*There is no problem, no matter how complex, which, upon careful analysis, does not become more complex.*

## Vorwort

Dieses Skript ist im Laufe mehrerer Jahre entstanden und liegt jetzt erstmals in einer relativ vollständigen Form vor. Es ist primär als Begleit- und Nachschlagewerk zur Lehrveranstaltung „Statistik in den Naturwissenschaften“ konzipiert. Das Ziel dieser Veranstaltung ist es, eine Einführung in statistischen Methoden und deren Grundlagen zu geben, die dann weiterführende eigene Literatur-Studien ermöglicht. Deshalb sollen hier nicht möglichst viele, sondern exemplarisch einige grundlegende Verfahren ausführlich erläutert und angewandt werden. Auf strenge Herleitungen der Verfahren wird bewußt verzichtet, aber auf präzise Formulierungen wird Wert gelegt und für (mathematisch) Interessierte werden auch Begründungen erwähnt.

Die effiziente Anwendung statistischer Verfahren auf konkrete Daten erfordert heutzutage einen Computer mit entsprechender Software. Ein wissenschaftlich-technischer Taschenrechner (in Kombination mit statistischen Tabellen) bietet insbesondere bei umfangreicheren Daten zu wenig Komfort und (wenn überhaupt) auch nur eingeschränkte Möglichkeiten für graphische Darstellungen. Für die hier behandelten Verfahren ist allerdings keine spezielle Statistik-Software erforderlich, sondern lediglich ein Tabellen-Kalkulationsprogramm. Da sich hierunter das Programm Microsoft Excel als Standard etabliert hat, werden die Tipps für die statistische Auswertung hier nur für dieses Programm genauer ausgeführt (Grundkenntnisse hierzu werden im Rahmen der Übungen zur Veranstaltung vermittelt).

Das Skript gliedert sich in drei große Teile sowie einen Anhang mit statistischen Tabellen, einem umfangreicheren Datensatz (StatLab-Auswahl 1985) und einer Zusammenstellung der Tipps für Microsoft Excel.

Der erste Teil ist der beschreibenden Statistik gewidmet. Hier werden erst die elementaren Techniken zur Datenbeschreibung behandelt: Tabellen, Diagramme und Maßzahlen (Mittelwert, Median, Standardabweichung). Danach wird auf den Zusammenhang zweier Merkmale eingegangen, und der lineare Zusammenhang ausführlicher erläutert. Hierbei wird (wegen ihrer Bedeutung) auch auf die gewichtete lineare Regression eingegangen. Anschließend werden zunächst solche nicht-linearen Zusammenhänge behandelt, die sich unter Verwendung von Transformationen auf lineare Zusammenhänge zurückführen lassen. Als Erweiterung des linearen Zusammenhangs wird noch der quadratische Zusammenhang betrachtet.

---

<sup>1</sup>vom Science-Fiction-Autor Poul Anderson, zitiert aus Hodges et al (1975), StatLab, Box 8.

Im zweiten Teil werden die wichtigsten stochastischen Grundbegriffe erläutert, die der deduktiven Statistik zugrunde liegen. Zuerst werden Wahrscheinlichkeiten und Zufallsvariablen eingeführt, wobei von (geringen) Vorkenntnissen aus der Schule ausgegangen wird. Als konkrete Verteilungen werden die Gleichverteilung, die Normalverteilung, die Binomialverteilung und die Poissonverteilung ausführlicher behandelt (die erst im Teil 3 benötigten Chiquaret-, t- und F-Verteilungen werden dort vorgestellt). Danach wird die Idee des Schätzens von Parametern am Beispiel von Erwartungswert und Standardabweichung erläutert.

Der dritte Teil ist der deduktiven (schließenden) Statistik gewidmet und behandelt Schätzungen mit Konfidenzgrenzen und statistische Tests sowie damit zusammenhängende statistische Analysen. Zuerst wird am Beispiel einer einzelnen Wahrscheinlichkeit (für ein interessierendes Ziel-Ereignis) die Idee von Konfidenzgrenzen und Tests ausführlich erläutert. Anschließend werden Vergleiche zweier Wahrscheinlichkeiten und danach Tests für ein und zwei Erwartungswerte behandelt. Hierbei werden überwiegend asymptotische Konfidenzgrenzen und Tests verwendet, die bei nicht zu kleinem Stichprobenumfang stets anwendbar sind — im Gegensatz zu den teilweise auch behandelten exakten Verfahren, die spezielle Verteilungsannahmen (wie z.B. Normalverteilung) voraussetzen. Danach wird die lineare (ungewichtete und gewichtete) Regressionsanalyse aus der Sicht der schließenden Statistik erneut behandelt. Weiter werden Anpassungstests zum Vergleich von beobachteten mit erwarteten Häufigkeiten erläutert und als Anwendung hiervon werden zweidimensionale Kontingenztafeln analysiert. Bei allen behandelten Tests wird großer Wert auf die Analyse der Testschärfe und des Fehlerrisikos 2. Art inclusive der darauf aufbauenden Versuchsplanung (Bestimmung des erforderlichen Stichprobenumfangs) gelegt.

Die statistischen Verfahren werden immer in mehreren Schritten vorgestellt, die in der Vorlesung teilweise parallel dargestellt werden, aber im Skript typischerweise in der Reihenfolge behandelt werden:

- einführendes Beispiel zur Motivation,
- allgemeines Ziel und Durchführung (mit illustrienden Graphiken),
- Zusammenfassung der Rechenschritte in einer Box,
- Analyse des einführenden Beispiels (und ggf. weiterer).

Beim (dringend empfohlenen) Nachvollziehen der Rechen-Beispiele ist zu beachten, daß die Zwischen- und Endergebnisse hier sinnvoll gerundet angegeben sind, aber in darauf aufbauenden Berechnungen trotzdem mit den nicht-gerundeten Werten (d.h. mit voller Genauigkeit des Computers) weitergerechnet wird. Deshalb können sich beim eigenen Überprüfen der hiesigen Rechnungen geringfügige Unterschiede ergeben, allerdings sollten die Abweichungen erst in der letzten oder (bei genaueren Angaben) in den letzten beiden hier angegebenen Dezimalstellen auftreten.

Der Kurs hat den Umfang von zwei Semesterwochenstunden und wird durch Übungen gleichen Umfangs begleitet. Da die Veranstaltung primär für Studierende im Fach Biologie vorgesehen ist, sind auch die Anwendungsbeispiele schwerpunktmäßig aus diesem Bereich. Bei den Beispielen im Skript (und den Übungen) handelt es sich weitgehend um konkrete und realistische Anwendungen (teilweise in stark vereinfachter Darstellung). Die jeweils wechselnden Übungsaufgaben sind nicht ins Skript aufgenommen, aber die aktuelle Serie findet man im Internet (vgl. unten).

Die vorliegende 4. Auflage wurde vollständig neu (im Satzsystem Latex) von Heidi Eckl-Reichelt geschrieben. Joachim Schalthöfer hat das endgültige Layout erstellt und insbesondere die bisherigen Graphiken konvertiert bzw. neu angefertigt. Beiden möchte ich an dieser Stelle herzlich danken. Inhaltlich unterscheidet sich diese Auflage von der 3. Auflage (Juli 2004) nur durch geringfügige Umformulierungen und Korrekturen. Ein wesentlicher Vorzug der Neuaufgabe ist, daß man in der digitalen PDF-Version jetzt nach Stichwörtern suchen kann (was vorher nicht möglich war).

Erfahrungsgemäß finden sich trotz Korrekturlesens vorwiegend in den neueren Passagen (aber nicht nur dort) einige Druckfehler. Bevor man daher am eigenen Verständnis zweifelt, sollte man auch einen Fehler im Skript in Erwägung ziehen. Für Hinweise auf Druckfehler oder andere Kommentare bin ich dankbar (am besten per e-mail: siehe unten).

Bremen, im September 2005

*Gerhard Osius*

PS Die aktuelle Fassung des Skriptes (sowie Übungsaufgaben und weiteres Material zur Veranstaltung) findet man unter:

[http://www.math.uni-bremen.de/~osis/download/lehre/Statistik\\_NW/](http://www.math.uni-bremen.de/~osis/download/lehre/Statistik_NW/).

Die Zugangsdaten für diese Internet-Seite sowie das Kennwort für das Skript können Studierende per e-Mail erfragen:

[osis@math.uni-bremen.de](mailto:osis@math.uni-bremen.de) .

# Inhaltsverzeichnis

<b>1</b>	<b>Beschreibende Statistik</b>	<b>1</b>
1.1	Daten und Merkmale . . . . .	1
1.2	Häufigkeiten und Diagramme . . . . .	3
1.3	Maßzahlen: Median, Mittelwert, Standardabweichung . . . . .	6
1.3.1	Lagemaße: Mittelwert und Median . . . . .	6
1.3.2	Streuungsmaße: Standardabweichung, Varianz und Spannweite . . . . .	8
1.3.3	Berechnung von Mittelwert und Standardabweichung . . . . .	9
1.3.4	Anwendung: StatLab-Auswahl 1985 . . . . .	11
1.4	Zusammenhang zweier Merkmale . . . . .	14
1.4.1	Zusammenhang zweier qualitativer Merkmale . . . . .	14
1.4.2	Zusammenhang zwischen einem <i>qualitativen</i> und einem <i>quantitativen</i> Merkmal . . . . .	16
1.5	Linearer Zusammenhang zweier quantitativer Merkmale . . . . .	17
1.5.1	Die Regressionsgerade . . . . .	17
1.5.2	Streuung um die Regressionsgerade, Bestimmtheitsmaß . . . . .	21
1.5.3	Der Korrelationskoeffizient . . . . .	22
1.5.4	Berechnung: Regressionsgerade, Bestimmtheitsmaß und Korrelationskoeffizient . . . . .	27
1.5.5	Anwendung: StatLab-Auswahl 1985 . . . . .	28
1.5.6	Gewichtung von Beobachtungen . . . . .	30
1.5.7	Anwendung . . . . .	33
1.6	Funktionaler Zusammenhang und linearisierende Transformationen . . . . .	35
1.6.1	Regressionsfunktion und Bestimmtheitsmaß . . . . .	36
1.6.2	Die logarithmische Transformation . . . . .	38
1.6.3	Logarithmischer Zusammenhang . . . . .	40
1.6.4	Exponentieller Zusammenhang . . . . .	42
1.6.5	Potenz-Zusammenhang . . . . .	44
1.6.6	Linearisierende Transformationen . . . . .	46
1.6.7	Logit-Transformation und logistischer Zusammenhang . . . . .	48
1.6.8	Periodischer Zusammenhang . . . . .	52
1.7	Quadratischer Zusammenhang . . . . .	58
1.7.1	Schätzung mit der Methode der kleinsten Quadrate . . . . .	59
1.7.2	Quadratischer vs. linearer Zusammenhang . . . . .	63
1.7.3	Quadratischer Zusammenhang bei transformierten Variablen . . . . .	65

<b>2 Grundbegriffe der Stochastik</b>	<b>68</b>
2.1 Wahrscheinlichkeiten und Zufallsvariablen . . . . .	69
2.1.1 Definitionen und Beispiele . . . . .	69
2.1.2 Eigenschaften von Wahrscheinlichkeiten . . . . .	75
2.1.3 Bedingte Wahrscheinlichkeit . . . . .	79
2.1.4 Unabhängigkeit zweier Zufallsvariablen . . . . .	80
2.1.5 Unabhängigkeit <i>mehrerer</i> Zufallsvariablen . . . . .	82
2.2 Diskrete Zufallsvariablen . . . . .	84
2.2.1 Die diskrete Gleichverteilung ( <i>Laplace</i> -Verteilung) . . . . .	85
2.3 Die Binomial-Verteilung . . . . .	87
2.3.1 Definition der Binomial-Verteilung . . . . .	88
2.3.2 Binomial-Wahrscheinlichkeiten . . . . .	89
2.3.3 Anwendung: Wahlumfragen . . . . .	91
2.3.4 Fakultät und Binomialkoeffizient . . . . .	93
2.4 Stetige Zufallsvariablen mit Dichten . . . . .	95
2.5 Die Normalverteilung . . . . .	100
2.5.1 Die Dichte der Normalverteilung . . . . .	100
2.5.2 Die Standard-Normalverteilung . . . . .	102
2.5.3 Berechnung von Wahrscheinlichkeiten bei Normalverteilung . . . . .	102
2.5.4 Eigenschaften der Normalverteilung . . . . .	103
2.5.5 Auftreten der Normalverteilung . . . . .	104
2.5.6 Die Normal-Approximation der Binomial-Verteilung . . . . .	106
2.6 Erwartungswert, Varianz und Standardabweichung von Zufallsvariablen . . . . .	108
2.6.1 Der Erwartungswert einer <i>diskreten</i> Zufallsvariablen . . . . .	109
2.6.2 Der Erwartungswert einer <i>stetigen</i> Zufallsvariablen . . . . .	110
2.6.3 Varianz und Standardabweichung einer Zufallsvariablen . . . . .	110
2.6.4 Eigenschaften von Erwartungswert, Varianz und Standardabweichung . . . . .	111
2.6.5 Schätzung von Erwartungswert und Standardabweichung . . . . .	114
2.7 Die Poisson-Verteilung . . . . .	119
<b>3 Deduktive Statistik</b>	<b>123</b>
3.1 Schätzung einer Wahrscheinlichkeit mit Konfidenzgrenzen . . . . .	124
3.1.1 Schätzen einer <i>Wahrscheinlichkeit</i> . . . . .	124
3.1.2 <i>Einseitige exakte</i> Konfidenzgrenzen für eine <i>Wahrscheinlichkeit</i> (nach <i>Pearson-Clopper</i> ) . . . . .	126
3.1.2 a Die <i>exakte untere einseitige</i> Konfidenzgrenze . . . . .	127
3.1.2 b Die <i>exakte obere einseitige</i> Konfidenzgrenze . . . . .	128

3.1.2 c	<i>Die exakten zweiseitigen Konfidenzgrenzen</i>	130
3.1.2 d	<i>Die Berechnung der exakten Konfidenzgrenzen</i>	131
3.1.3	<i>Asymptotische (approximative) Konfidenzgrenzen für eine Wahrscheinlichkeit</i>	133
3.1.3 a	<i>Normale Konfidenzgrenzen</i>	134
3.1.3 b	<i>Große Konfidenzgrenzen für große Umfänge</i>	137
3.1.4	<i>Konfidenz-Box 1:</i>	140
3.2	<i>Testen einer Wahrscheinlichkeit</i>	141
3.2.1	<i>Einführung in statistische Tests am Beispiel</i>	142
3.2.2	<i>Ein- und zweiseitige Tests für eine Wahrscheinlichkeit</i>	150
3.2.3	<i>Fehlerrisiko 2. Art und Testschärfe (Power)</i>	153
3.2.4	<i>Versuchsplanung: erforderlicher Stichprobenumfang</i>	156
3.3	<i>Vergleich zweier Wahrscheinlichkeiten</i>	158
3.3.1	<i>Schätzung des Unterschieds mit Konfidenzgrenzen</i>	159
3.3.2	<i>Testen von Hypothesen</i>	161
3.3.3	<i>Test- und Konfidenz-Box 2:</i>	164
3.3.4	<i>Fehlerrisiko 2. Art und Testschärfe (Power)</i>	165
3.3.5	<i>Versuchsplanung: erforderlicher Stichprobenumfang</i>	166
3.3.6	<i>Anwendungen</i>	168
3.4	<i>Konfidenzgrenzen und Tests für einen Erwartungswert</i>	171
3.4.1	<i>Konfidenzgrenzen für den Erwartungswert</i>	171
3.4.2	<i>Tests für den Erwartungswert</i>	174
3.4.3	<i>Fehlerrisiko 2. Art und Testschärfe (Power)</i>	180
3.4.4	<i>Versuchsplanung: erforderlicher Stichprobenumfang</i>	181
3.4.5	<i>Die Student'sche <math>t</math>-Verteilung</i>	183
3.5	<i>Vergleich zweier Erwartungswerte</i>	185
3.5.1	<i>Schätzung des Unterschieds mit Konfidenzgrenzen</i>	185
3.5.2	<i>Testen von Hypothesen</i>	189
3.5.3	<i>Fehlerrisiko 2. Art und Testschärfe (Power)</i>	193
3.5.4	<i>Versuchsplanung erforderlicher Stichprobenumfang</i>	194
3.6	<i>Lineare Regressionsanalyse</i>	197
3.6.1	<i>Ungewichtete Regression bei homogenen Varianzen</i>	198
3.6.1 a	<i>Schätzung der Parameter und Modellüberprüfung</i>	198
3.6.1 b	<i>Testen des Anstiegs</i>	200
3.6.1 c	<i>Fehlerrisiko 2. Art, Testschärfe und Versuchsplanung</i>	204
3.6.1 d	<i>Prognosen</i>	208
3.6.1 e	<i>Anwendungen</i>	209

3.6.2	<i>Gewichtete</i> Regressionsanalyse bei <i>proportionalen</i> Varianzen . . . . .	215
3.6.2 a	Anwendungen . . . . .	219
3.7	Anpassungstests:	
	Chiquaret- und Likelihood-Quotienten-Test . . . . .	222
3.7.1	Einfache Nullhypothesen . . . . .	224
3.7.2	Nullhypothesen mit <i>unbekannten</i> Parametern . . . . .	227
3.7.3	Fehlerrisiko 2. Art und Versuchsplanung . . . . .	230
3.7.3 a	Chiquaret-Test für <i>einfache</i> Nullhypothesen . . . . .	230
3.7.3 b	Chiquaret-Test für Nullhypothesen <i>mit Parametern</i> . . . . .	232
3.7.3 c	Likelihood-Quotienten-Test . . . . .	234
3.7.4	Die Chiquaret-Verteilung . . . . .	239
3.8	Zweidimensionale Kontingenztafeln . . . . .	242
3.8.1	Vergleich der Verteilung eines Merkmals in verschiedenen Gruppen	243
3.8.2	Test auf Unabhängigkeit zweier Merkmale . . . . .	252

## Anhang

<b>A Literatur-Auswahl</b>	<b>259</b>
<b>B Statistik-Funktionen von Microsoft Excel</b>	<b>260</b>
<b>C StatLab-Daten</b>	<b>261</b>
<b>T Statistische Tabellen</b>	<b>264</b>
Verteilungsfunktion der Standard-Normalverteilung . . . . .	265
Quantile der Standard-Normalverteilung und der <i>t</i> -Verteilung . . . . .	267
Quantile der Chiquaret-Verteilung . . . . .	269
<b>Boxenverzeichnis</b>	<b>277</b>
<b>Index</b>	<b>278</b>

# 1 Beschreibende Statistik

Die Aufgabe der beschreibenden Statistik besteht darin, eine Menge von Daten (Datensatz) in möglichst übersichtlicher Form darzustellen. Bei größeren Datenmengen steht hierbei eine Reduktion (Zusammenfassung) der Daten im Hinblick auf wesentliche Aspekte im Vordergrund.

Für die beschreibende Darstellung von Daten verwendet man

- *Tabellen* (z.B. Häufigkeitstabellen)
- *Graphiken* (z.B. Häufigkeitsdiagramme, Punktwolken)
- *Maßzahlen* (z.B. Mittelwert, Standardabweichung, Korrelation).

## 1.1 Daten und Merkmale

Die darzustellenden Daten werden typischerweise im Rahmen einer *Studie* (z.B. über die Entwicklung von Neugeborenen), eines *Experiments* (z.B. über das Wachstum einer Pflanze) oder durch *Beobachtungen* (z.B. Monitoring der Ozon-Konzentration in der Luft) erhoben. Für jedes *Untersuchungsobjekt* (z.B. Individuum einer Studie, Pflanze in einem Experiment oder Beobachtungszeitpunkt) werden dabei ein oder mehrere *Merkmale* erfasst, wobei anzustreben ist, daß alle interessierenden Merkmale vollständig, d.h. an *jedem* Untersuchungsobjekt, erhoben werden.

Als konkretes Beispiel eines umfangreichen Datensatzes dienen die *StatLab-Daten*. Die Daten wurden im Rahmen einer medizinischen Studie von Neugeborenen („Kaiser Foundation Health Plan“) in Oakland (Kalifornien) unter Leitung von J. Yerushalmy in den Jahren 1961-1972 erhoben. Für jedes Kind (und für seine Eltern) wurden zahlreiche Daten zu zwei Zeitpunkten registriert: bei der Geburt (in den Jahren 1961-1963) und ca. 10 Jahre danach (zur Kontrolle). Die vollständigen Daten von je 1296 Mädchen und Jungen sind in dem Lehrbuch von Hodges, Krech und Crutchfield (1975) „*StatLab: an empirical introduction to statistics*“ enthalten. Wir betrachten hier eine (zufällig erzeugte) Auswahl von je 50 Mädchen und Jungen aus der gesamten Studie und dabei auch nur einige besonders interessierende Merkmale (vgl. Anhang **A2**).

### *Arten von Merkmalen (Variablen)*

Die interessierenden Merkmale (auch *Variablen* genannt) können unterschiedlicher Natur sein und man unterscheidet (auch im Hinblick auf die anzuwendenden Methoden) verschiedene Arten.

Ein Merkmal ist *qualitativ*, wenn es verschiedene Ausprägungen besitzt, die sich *nicht* inhaltlich durch *Zahlen* repräsentieren lassen. Beispiele hierfür sind die folgenden Merkmale (mit ihren Ausprägungen): der Familienstand (ledig, verheiratet, verwitwet) und die AB0-Blutgruppe (0, A, B und AB).

Ein qualitatives Merkmal mit nur *zwei* Ausprägungen heißt auch *alternativ* und ist typischerweise durch das Vorhandensein oder Fehlen eines Charakteristikums gegeben (etwa: Raucher - Nichtraucher oder weiblich - männlich).

Ein *ordinales* (qualitatives) Merkmal liegt vor, wenn sich seine Ausprägungen *sinnvoll anordnen* lassen, z.B. die Bewertung eines Zustands durch die Kategorien: schlecht - mittel - gut.

Obwohl sich die Ausprägungen qualitativer Merkmale nicht inhaltlich durch Zahlen darstellen lassen, werden sie trotzdem oft *formal* durch eine Zahl (typischerweise eine laufende Nummer) *codiert* bzw. *klassifiziert*, z.B. bei den vier Blutgruppen durch  $1 = \text{Null}$ ,  $2 = A$ ,  $3 = B$ ,  $4 = AB$ . Eine solche Codierung dient ausschließlich der Identifikation der Ausprägung durch Zahlen und stellt keinen *inhaltlichen* Zusammenhang zwischen Ausprägung und Zahlen-Code dar. Insbesondere ist es unsinnig Mittelwerte solcher Codes zu bilden, z.B. ist eine *mittlere Blutgruppe* von 2,25 für eine Population eine sinnlose Angabe.

Im Gegensatz zu den qualitativen Merkmalen sind *quantitative* Merkmale solche, deren Ausprägungen auf natürliche Weise durch *Zahlen* beschrieben oder gemessen werden. Man unterscheidet hier noch zwischen *diskreten* und *stetigen* (kontinuierlichen) Merkmalen. Ein *diskretes* Merkmal nimmt nur isolierte Werte an, typischerweise ganze Zahlen wie z.B. die Anzahl von Mutanten einer Zellkultur oder die konsumierten Zigaretten pro Tag. Demgegenüber kann ein *stetiges* Merkmal kontinuierliche Werte aus einem Zahlenintervall annehmen (z.B. das Gewicht in kg oder die Größe in cm), die in der Praxis dann in *gerundeter* Form vorliegen.

### Rundung:

Unter Berücksichtigung der Meßgenauigkeit (oder aus Gründen der Übersicht) werden Meßwerte oft *gerundet* angegeben (*Zur Erinnerung*: Die Ziffern 0 – 4 werden *abgerundet* und 5 – 9 *aufgerundet*). Eine gerundete Angabe ist immer so zu interpretieren, daß der genaue Wert in einem entsprechenden *Intervall* liegt, z.B.

$$\text{Gewicht} = 75 \text{ [kg]} \quad \text{bedeutet} \quad 74,5 \leq \text{Gewicht} < 75,5.$$

Hierbei ist das Symbol “ $\leq$ ” die mathematische Abkürzung für “*kleiner oder gleich*” und “ $<$ ” die für “*echt kleiner*”. (Als Eselbrücke: die “kleine” Spitze zeigt zum kleineren Wert und die “große” Öffnung zum größeren Wert).

Manchmal wird auch nur abgerundet, z.B. beim Alter in *vollendeten Lebensjahren*:

$$\text{Alter} = 30 \text{ Jahre} \quad \text{bedeutet} \quad 30 \leq \text{Alter} < 31.$$

Durch Rundung werden stetige Merkmale in diskrete überführt, wie z.B. das Alter in vollendeten Jahren. Daher ist der Unterschied zwischen stetigen und diskreten Merkmalen in der Praxis eher fließend, aber dennoch von theoretischer Bedeutung.

## 1.2 Häufigkeiten und Diagramme

Tritt ein Merkmal im Datensatz in  $K$  verschiedenen Ausprägungen  $A_1, \dots, A_K$  auf, so interessiert man sich in erster Linie für die *Häufigkeiten* dieser Ausprägungen  $A_k$ .

Die absolute *Häufigkeit*  $n_k$  von  $A_k$  ist die Anzahl aller Untersuchungsobjekte mit dieser Ausprägung. Die Summe aller absoluten Häufigkeiten ist die Gesamtzahl  $n$  aller Untersuchungsobjekte, die auch als *Stichprobenumfang* bezeichnet wird:

$$n = n_1 + n_2 + \dots + n_K = \sum_{k=1}^K n_k = \sum_k n_k \quad (\text{Stichprobenumfang})$$

Die relative *Häufigkeit (Frequenz)* von  $A_k$  ist der Anteil

$$f_k = \frac{n_k}{n}.$$

Er wird meist in Prozent angegeben, d.h. als  $100 \cdot f_k\%$ . Es gilt (Kontrolle)

$$\begin{array}{lll} 0 \leq f_k \leq 1 & \text{bzw.} & 0 \leq 100 \cdot f_k \leq 100, \\ \sum_{k=1}^K f_k = 1 & \text{bzw.} & \sum_{k=1}^K 100 \cdot f_k = 100. \end{array}$$

### Tabellierung:

Die absoluten und relativen Häufigkeiten werden in einer *Häufigkeitstabelle* für das Merkmal zusammengefasst. Eine solche Tabelle (vgl. Tabelle 1) enthält typischerweise je eine Spalte für die laufende Nr.  $k$ , die Ausprägung  $A_k$  und die Häufigkeiten  $n_k$  und  $f_k$ . Zur Kontrolle sollten auch die Summen über alle  $n_k$  bzw.  $f_k$  angegeben werden.

### Graphische Darstellungen:

Zur Veranschaulichung von Häufigkeiten werden diese durch Flächen dargestellt, deren Größen den Häufigkeiten entsprechen. Typische Darstellungen sind (vgl. Abb. 1):

- *Säulen-Diagramm (Histogramm)*

Die Häufigkeiten entsprechen den Flächen von Säulen und (bei gleicher Säulenbreite) somit auch der *Höhe* der Säulen.

- *Kreis-Diagramm (Torten-Diagramm)*

Aufteilung eines Kreises in Segmente, wobei der Winkel  $w_k$  (und damit auch die zugehörige Fläche) des  $A_k$  repräsentierenden Kreissegments der relativen Häufigkeit  $f_k$  entspricht, d.h.

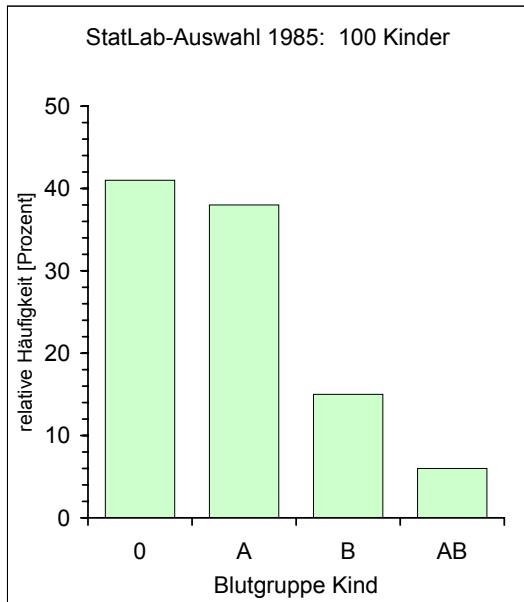
$$w_k = f_k \cdot 360^\circ.$$

**Beispiel: AB0-Blutgruppe in der StatLab-Auswahl 1985**Tabelle 1: Häufigkeitstabelle für die  $n = 100$  Kinder

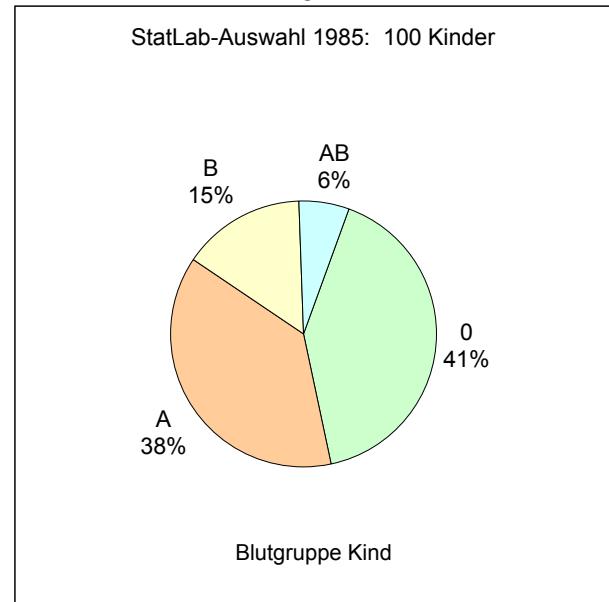
Nr $k$	Merkmal/Klasse $A_k$	Häufigkeit absolut $n_k$	Häufigkeit relativ $f_k$
1	0	41	41,0%
2	A	38	38,0%
3	B	15	15,0%
4	AB	6	6,0%
$\sum$		100	100,0%

Abb. 1: Darstellung der Blutgruppenverteilung der  $n = 100$  Kinder

Säulen-Diagramm (Histogramm)



Kreis-Diagramm

**Klassifizierung eines Merkmals**

Liegen sehr viele verschiedene Ausprägungen eines Merkmals vor (typischerweise bei stetigen Merkmalen), so liefert die Häufigkeitstabelle keine sinnvolle Zusammenfassung mehr, und die Diagramme werden unübersichtlich. Hier ist meist eine Reduktion sinnvoll, indem man das ursprüngliche Merkmal neu klassifiziert, z.B. durch starke Rundung bei einem stetigen Merkmal. Die so entstandenen, neuen Klassen  $k = 1, \dots, K$  entsprechen dann den Ausprägungen  $A_1, \dots, A_K$  des *klassifizierten* Merkmals. Die Anzahl  $K$  der Klassen soll dabei nicht zu groß sein, als grobe Faustregel kann gelten:

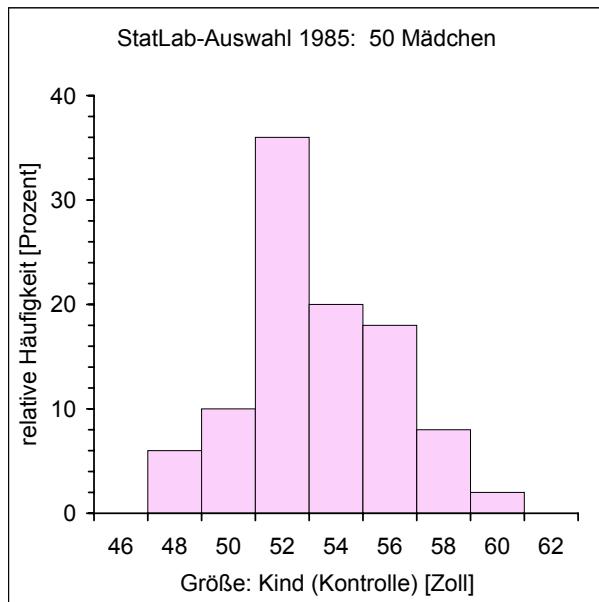
$$(\text{Anzahl } K \text{ der Klassen}) \quad \leq \quad \frac{1}{5} (\text{Anzahl } n \text{ der Untersuchungsobjekte})$$

**Beispiel: Größe [Zoll] der 50 Mädchen zum Kontrollzeitpunkt in der StatLab-Auswahl 1985**

Tabelle 2: Häufigkeitstabelle der klassifizierten Größe

Nr k	Merkmal von (incl.) [Zoll]	bis (excl.) [Zoll]	Klassen- Mitte [Zoll]	Häufigkeit absolut $n_k$	Häufigkeit relativ $f_k$
1	47	-	49	48	3
2	49	-	51	50	5
3	51	-	53	52	18
4	53	-	55	54	10
5	55	-	57	56	9
6	57	-	59	58	4
7	59	-	61	60	1
$\Sigma$		-		50	100%

Abb. 2: Säulen-Diagramm (Histogramm) der Größen-Verteilung der  $n = 50$  Mädchen



## ☒ Tipps für Microsoft Excel

**Diagramme:** Die oben genannten (und zahlreiche weitere) *Grafiken* lassen sich unter Verwendung des *Diagramm-Assistenten* von *Excel* einfach erstellen.

**Klassifizierung:** Eine *Häufigkeitstabelle* für ein *qualitatives* Merkmal (z.B. die Blutgruppe) erzeugt man aus den Originaldaten mit dem Befehl **ZÄHLENWENN**. Und für ein *quantitatives* Merkmal (z.B. die Körpergröße) lässt sich eine *Häufigkeitstabelle* mit dem Befehl **HÄUFIGKEIT** in einer *Matrix-Formel* erstellen (zur *Eingabe* von *Matrix-Formeln* konsultiere man die *Excel-Hilfe*).

## 1.3 Maßzahlen: Median, Mittelwert, Standardabweichung

Der Ausgangspunkt für die folgenden Betrachtungen ist ein quantitatives Merkmal  $x$ . Unser Ziel ist es, die beobachteten  $x$ -Werte  $x_1, \dots, x_n$  (*Stichprobe vom Umfang n*) durch aussagekräftige *Maßzahlen* zusammenzufassen (Reduktion der Daten auf wesentliche Aspekte). Hierfür geben wir *Lage-* und *Streuungsmaße* an, die folgende Fragen beantworten:

Wo liegen die Werte (im Mittel)?

Wie stark streuen die Werte?

Die Maßzahlen werden zunächst inhaltlich erläutert (1.3.1-2), dann wird auf ihre Berechnung eingegangen (1.3.3) und abschließend wird ein Anwendungsbeispiel durchgerechnet (1.3.4).

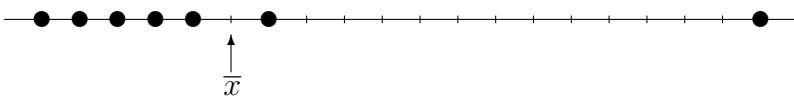
### 1.3.1 Lagemaße: Mittelwert und Median

Zur Charakterisierung der *Lage* der Werte  $x_1, \dots, x_n$  sucht man einen *mittleren Wert*. Die Präzisierung eines solchen mittleren Wertes kann auf verschiedene Weise erfolgen, von denen wir nur die beiden wichtigsten besprechen und vergleichen: das *arithmetische Mittel (Mittelwert)* und den *Median*.

Der *Mittelwert*  $\bar{x}$  der Werte  $x_1, \dots, x_n$  ist definiert als ihr arithmetisches Mittel,

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_i x_i \quad (\text{Mittelwert, arithmetisches Mittel}) .$$

Geometrisch entspricht der Mittelwert dem *Schwerpunkt* der  $x$ -Werte,



d.h. die Summe aller Abweichungen vom Mittelwert ist Null:

$$\sum_i (x_i - \bar{x}) = 0.$$

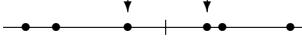
Der *Median*  $x_{50\%}$  (zentraler Wert) teilt die Werte  $x_1, \dots, x_n$  derart, daß *gleichviel* Werte *unterhalb* wie *oberhalb* des Medians liegen. Zur Bestimmung des Medians werden die Werte zunächst *aufsteigend* nach ihrer Größe *sortiert* und man erhält die sogenannte *Rangliste* der Werte:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)} \quad (\text{Rangliste der } x\text{-Werte}) .$$

Insbesondere ist also  $x_{(1)}$  der *kleinste* und  $x_{(n)}$  der *größte* Wert. Im Gegensatz zu  $x_{(1)}$  und  $x_{(n)}$  ist  $x_1$  der erste und  $x_n$  der letzte Wert in der *gegebenen* Reihenfolge der  $x$ -Werte in der Stichprobe.

Der Median  $x_{50\%}$  zerlegt die Rangliste derart, daß *gleichviele* Werte unterhalb wie oberhalb liegen, d.h.

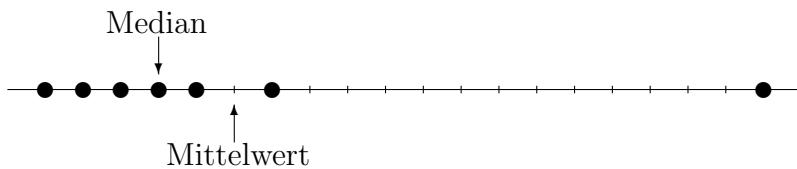
- für *ungerades*  $n$  ist:  $x_{50\%} = x_{(i)}$  mit  $i = \frac{1}{2}(n+1)$   
  
 $(n=5)$

- für *gerades*  $n$  ist:  $x_{50\%} = (x_{(i)} + x_{(i+1)})/2$  mit  $i = \frac{1}{2}n$   
  
 $(n=6)$

### Median vs. Mittelwert

Der Mittelwert hängt von allen  $x$ -Werten ab, und wird von extremen *Ausreißern* (nach oben oder unten) stark beeinflußt. Der Median ist dagegen *robust* gegen solche Ausreißer, weil er nur von den *zentralen*  $x$ -Werten abhängt.

#### Beispiel 1:



Verschiebt man den größten Wert weiter nach rechts, wird nur der Mittelwert größer, aber der Median ändert sich nicht.

#### Beispiel 2:

Bei einem Prozeß in den USA machen drei Gutachter folgende Vorschläge über die Höhe des Schadensersatzanspruchs:

Gutachter	Betrag $x[\$]$
1	3 000
2	30 000
3	3 000 000

Es ist hier von entscheidender Bedeutung, ob das Gericht den Mittelwert  $\bar{x} = 1\,011\,000$  oder den Median  $x_{50\%} = 30\,000\$\text{}$  der Gutachtervorschläge wählt.

**Beispiel 3:**

Bei der Angabe eines „mittleren Preises“ für ein Produkt ist es sinnvoll, den *Median* der Einzelpreise verschiedener Anbieter zu verwenden, wie es z.B. in der Zeitschrift *test* der *Stiftung Warentest* üblich ist. Hierdurch werden Verzerrungen sowohl durch „Dumpingpreise“ (in Ballungsgebieten mit viel Konkurrenz) als auch durch „Wucherpreise“ (in ländlichen Gebieten ohne Konkurrenz) vermieden.

### 1.3.2 Streuungsmaße: Standardabweichung, Varianz und Spannweite

Der umgangssprachliche Begriff der Streuung von Zahlen (oder Punkten) wird in der Statistik durch unterschiedliche Streuungsmaße präzisiert. Als Maßzahl für die Streuung der Werte  $x_1, \dots, x_n$  wird vorwiegend die *Standardabweichung* bzw. deren Quadrat, die *Varianz*, und zusätzlich auch die *Spannweite* verwendet.

**Standardabweichung und Varianz**

Ausgangspunkt für die Berechnung der Standardabweichung sind die *Abweichungen* ( $x_i - \bar{x}$ ) der  $x$ -Werte von ihrem Mittelwert. Da man auf das *Vorzeichen* dieser Abweichung keinen Wert legt (d.h. ob  $x_i$  unter- oder oberhalb von  $\bar{x}$  liegt), wird die *quadratische Abweichung*  $(x_i - \bar{x})^2$  verwendet. Man könnte auch den *Absolutbetrag*  $|x_i - \bar{x}|$  nehmen, aber das hat sich nicht durchgesetzt, weil die Berechnungen mit dem *Quadrat* einfacher sind.

Die (*empirische*) *Varianz* der  $x$ -Werte ist nun die *mittlere quadratische Abweichung vom Mittelwert*:

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

Man beachte, daß im Nenner die um 1 *reduzierte* Anzahl  $n-1$  aller  $x$ -Werte und nicht der Stichprobenumfang  $n$  steht. Eine genaue Begründung hierfür wird erst später (im Abschnitt **2.6.5**) gegeben. Grob gesprochen liegt es daran, daß man zur Bestimmung von  $s_x^2$  bereits *einen* aus den  $x$ -Werten *berechneten* Parameter (den Mittelwert  $\bar{x}$ ) verwendet.

Da die Varianz  $s_x^2$  als *Dimension* das Quadrat der Dimension der  $x$ -Werte hat (wenn  $x$  eine Länge ist, so ist die Varianz eine Fläche), verwendet man ihre Wurzel als Streuungsmaß, die sogenannte (*empirische*) *Standardabweichung*  $s_x$ ,

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

Je größer  $s_x^2$  bzw.  $s_x$  ist, desto stärker streuen die  $x$ -Werte. Für  $s_x = 0$  sind alle  $x$ -Werte gleich (keine Streuung). Im Bereich  $\bar{x} \pm 2s_x$  liegen immer mindestens 75% aller  $x$ -Werte, oft jedoch bis zu 95% (z.B. bei der Normalverteilung, vgl. 2.4).

*Warnung:* Im Gegensatz zur obigen Definition werden gelegentlich auch die entsprechenden Streuungsmaße verwendet, bei denen man durch  $n$  statt durch  $n-1$  dividiert (was wir hier *nicht* wollen). Bei Taschenrechnern oder Rechnerprogrammen sind meist beide Versionen (mit  $n-1$  und mit  $n$ ) implementiert und man sollte sich vergewissern, daß man die obige Version verwendet.

### Spannweite

Ein völlig anderes und leicht zu berechnendes Streuungsmaß ist die Spannweite, definiert als Differenz vom größten Wert  $x_{(n)}$  zum kleinsten Wert  $x_{(1)}$  in der Rangliste der  $x$ -Werte (vgl. 1.3.1).

$$\text{Spannweite} = (\text{größter Wert}) - (\text{kleinster Wert}) = x_{(n)} - x_{(1)}.$$

Für die Berechnung wird nicht die gesamte Rangliste, sondern nur das Maximum  $x_{(n)}$  und das Minimum  $x_{(1)}$  der  $x$ -Werte benötigt. Diese extremen Werte sind meist ohnehin von Interesse, z.B. für die Skalierung von graphischen Darstellungen.

### 1.3.3 Berechnung von Mittelwert und Standardabweichung

Die Varianz  $s_x^2$  der  $x$ -Werte  $x_1, \dots, x_n$  kann auch berechnet werden ohne hierzu die einzelnen Abweichungsquadrate  $(x_i - \bar{x})^2$  vom Mittelwert  $\bar{x}$  bestimmen zu müssen. Durch algebraische Umformungen läßt sich die Varianz durch die Summe  $\sum x$  der  $x$ -Werte und die Summe  $\sum x^2$  der quadrierten  $x$ -Werte darstellen. Die zugehörigen Schritte zur Berechnung von Mittelwert, Varianz und Standardabweichung sind in der *Rechen-Box 1* zusammengefaßt.

<b>Rechen-Box 1 für Daten einer Variablen <math>x</math></b>	
<i>Stichprobenumfang</i>	$n$
<i>x-Werte:</i>	$x_1, x_2, \dots, x_n$
<i>Summen:</i>	
<i>x-Werte</i>	$\sum x = x_1 + x_2 + \dots + x_n$
<i>x-Quadrate</i>	$\sum x^2 = x_1^2 + x_2^2 + \dots + x_n^2$
<i>Mittelwert:</i>	$\bar{x} = \frac{1}{n}(\sum x)$
<i>Summe der quadratischen Abweichungen:</i>	
	$\begin{aligned} Sxx &= \sum_i (x_i - \bar{x})^2 = (\sum x^2) - \frac{1}{n}(\sum x)^2 \\ &= (\sum x^2) - n\bar{x}^2 \end{aligned}$
<i>(empirische) Varianz:</i>	$s_x^2 = \frac{1}{n-1} Sxx$
<i>(empirische) Standardabweichung:</i>	$s_x = \sqrt{\frac{1}{n-1} Sxx}$

**Skalenverschiebung:**

Zur Vermeidung großer Zahlen (beim Quadrieren der  $x$ -Werte) kann man die  $x$ -Werte um einen Wert  $c$  verschieben und mit den verschobenen Werten  $u = x - c$  rechnen. Die *Mittelwerte* sind dann entsprechend verschoben,

$$\bar{u} = \bar{x} - c \quad \text{bzw.} \quad \bar{x} = \bar{u} + c,$$

und die *Standardabweichungen* der  $x$ - und  $u$ -Werte stimmen sogar überein:

$$s_x = s_u \quad \text{und} \quad Sxx = Suu.$$

**Lineare Transformation:**

Transformiert man die  $x$ -Werte (z.B. beim Übergang auf eine andere Maß-Einheit) linear zu

$$u = a + bx,$$

so transformiert sich der Mittelwert entsprechend, d.h.

$$\bar{u} = a + b\bar{x}$$

und die Standardabweichung der  $u$ -Werte ergibt sich zu:

$$s_u = bs_x \quad \text{und} \quad Suu = b^2 Sxx.$$

*Beispiele:* Umrechnung der Länge von Zoll bzw. Inch ( $x$ ) in cm ( $u$ ):

$$u = 2,54x .$$

Umrechnung der Temperatur von °Celsius ( $x$ ) in °Fahrenheit ( $u$ )

$$u = 32 + 1,8x .$$

Umrechnung eines Umsatzes ( $x$ ) auf den Rückgang ( $u$ ) des Umsatzes im Bezug auf den zur Kostendeckung erforderlichen Umsatz  $a$ :

$$u = a - x .$$

### **Regeln zur Rechengenauigkeit**

Generell sollen alle Berechnungen mit *maximaler* Genauigkeit erfolgen, d.h. weder die Eingangsdaten ( $x$ -Werte) noch die *Zwischenergebnisse*  $\sum x$ ,  $\sum x^2$  und  $Sxx$  dürfen gerundet werden. Lediglich die *Endergebnisse* sind sinnvoll zu runden.

Konkret ist die Summe  $\sum x$  mit *derselben* Anzahl von Nachkommastellen anzugeben wie die  $x$ -Werte, aber die Summe  $\sum x^2$  und  $Sxx$  mit der *doppelten Anzahl* von Nachkommastellen. Wenn man mit einem (Taschen-)Rechner arbeitet, sollte man die Zwischenergebnisse mit voller Genauigkeit im Rechner *speichern* und mit diesen (genauen) Werten weiterrechnen.

Als Endergebniss werden *Mittelwert* und *Standardabweichung* meist mit einer Nachkommastelle *mehr* als die  $x$ -Werte angegeben, bei extrem großen Umfängen (ab  $n \geq 100$ ) sogar mit *zusätzlichen* Nachkommastellen.

#### **1.3.4 Anwendung: StatLab-Auswahl 1985**

##### **Beispiel: Körpergröße der Mädchen zum Kontrollzeitpunkt**

Zur Bestimmung des Medians und der Spannweite sind die Körpergrößen [Zoll] der  $n = 50$  Mädchen zunächst aufsteigend zu sortieren und die entsprechende Rangliste ist in Spalte 2 der Tabelle 1 (auszugsweise) wiedergegeben. Die Tabelle enthält weiter die quadrierten Größen  $x^2$  sowie die zur Berechnung von Mittelwert und Standardabweichung erforderlichen Summen  $\sum x$  und  $\sum x^2$ . Da die Quadrate  $x^2$  relativ große Zahlen sind, wurde die Berechnung von Mittelwert und Standardabweichung zur Demonstration *zusätzlich* auch unter Verwendung der Verschiebung (Transformation)  $u = x - 50$  durchgeführt, die hier zu kleineren Quadraten  $u^2$  führt. Es ist klar, daß man normalerweise *entweder* mit den  $x$ -Werten *oder* mit den verschobenen  $u$ -Werten arbeitet, aber *nicht* mit *beiden*. Man beachte, daß die Rangordnung der  $u$ -Werte mit derjenigen der  $x$ -Werte übereinstimmt.

Tabelle 1: Rangliste und transformierte Werte  $u = x - 50$ .

Rang	$x$	$x^2$	$u = x - 50$	$u^2$	Nr.
1	47,4	2246,76	-2,6	6,76	44
2	48,7	2371,69	-1,3	1,69	9
3	48,9	2391,21	-1,1	1,21	8
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
25	52,9	2798,41	2,9	:	22
26	52,9	2798,41	2,9	8,41	49
:	:	:	:	:	:
:	:	:	:	:	:
:	:	:	:	:	:
48	57,5	3306,25	7,5	56,25	45
49	57,9	3352,41	7,9	62,41	32
50	60,0	3600,00	10,0	100,00	50
$\sum$	2661,3	142002,59	161,3	872,59	

Aus der Tabelle 1 ergeben sich mit den Formeln aus der *Rechen-Box 1* die folgenden Maßzahlen:

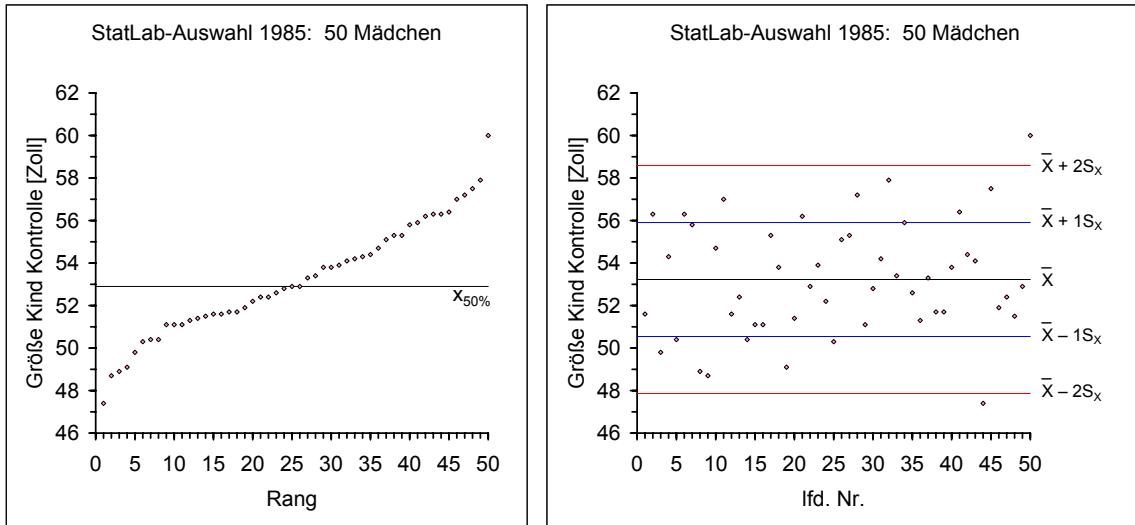
$$\text{Spannweite: } x_{(50)} - x_{(1)} = 12,6; \quad n = 50$$

$$\text{Median: } x_{50\%} = 52,9$$

$$\text{Mittelwert: } \bar{x} = 53,23; \quad \bar{u} = 3,23;$$

$$\text{Standardabweichung: } s_x = 2,68; \quad s_u = 2,68; \quad S_{xx} = S_{uu} = 352,2362.$$

Man beachte, daß der Median und der Mittelwert sich hier zwar nur geringfügig unterscheiden, aber nicht *exakt* übereinstimmen. Zur Veranschaulichung sind die Meßwerte in Abbildung 1 zusammen mit dem Median, dem Mittelwert und den Bereichen  $\bar{x} \pm 1s_x$  sowie  $\bar{x} \pm 2s_x$  dargestellt.



**Abb. 1.** Links: Die geordneten  $x$ -Werte mit ihrem Median  $x_{50\%}$ .

Rechts: Die ungeordneten  $x$ -Werte mit ihrem Mittelwert  $\bar{x}$  und den durch die Standardabweichung  $s_x$  gegebenen Bereichen  $\bar{x} \pm 1s_x$  sowie  $\bar{x} \pm 2s_x$ . Man erkennt deutlich, daß die Standardabweichung ein Maß für die *Streuung* ist: der größte Teil der  $x$ -Werte „streut“ im Bereich  $\bar{x} \pm 1s_x$  und außerhalb des Bereichs  $\bar{x} \pm 2s_x$  liegen nur sehr wenige Werte.

### ☒ Tipps für Microsoft Excel

**Funktionen:** Die folgenden *Statistik-Funktionen* stehen in *Excel* direkt zur Verfügung (mit Erläuterung in Klammern, sofern erforderlich):

MITTELWERT

MEDIAN

STABW

(= Standardabweichung)

VARIANZ

MIN

(= kleinster Wert)

MAX

(= größter Wert).

**Rangliste:** Man erhält sie mit dem Befehl *Sortieren* im *Excel*-Menü *Daten*.

**Diagramme:** Eine Darstellung (wie Abb. 1) erstellt man mit dem *Diagramm-Assistenten* als *Standardtyp Punkt (XY)*. Die waagerechten Hilfslinien in Abb. 1 werden dabei *nicht* dargestellt, lassen sich aber als „Geraden“ einzeichnen (vgl. auch 1.5.7).

**Formate:** Zur Übersicht sollte man Zahlen in *Excel formatieren*, indem man im Menü *Format - Zellen* für jedes Feld (oder ganze Bereiche) eine sinnvolle Anzahl von *Nachkommastellen* einstellt, wie z.B. in den *Regeln zur Rechengenauigkeit* in 1.3.1. Die *Formatierung* hat *keinen* Einfluß auf die interne *Rechengenauigkeit* von *Excel*.

## 1.4 Zusammenhang zweier Merkmale

Ein Datensatz enthält typischerweise für jedes Untersuchungsobjekt nicht nur Angaben über ein *einziges* sondern über *mehrere Merkmale*. Nachdem man sich einen Überblick über jedes *einzelne* Merkmal verschafft hat (durch Häufigkeiten, Diagramme oder Maßzahlen) interessiert sich man sich über mögliche Zusammenhänge zwischen *zwei* oder *mehreren* Merkmalen. Wir werden hier nur je *zwei* Merkmale gleichzeitig betrachten und diese suggestiv mit  $X$  und  $Y$  bezeichnen. Die Methoden zur Darstellung und Beschreibung des Zusammenhangs zwischen  $X$  und  $Y$  hängen nun davon ab, ob die Merkmale *qualitativ* oder *quantitativ* sind. Wenn eines der beiden Merkmale *qualitativ* ist, so lassen sich die bisherigen Methoden für ein Merkmal sinngemäß übertragen. Sind jedoch beide Merkmale *quantitativ*, so ergeben sich völlig neue Verfahren, auf die wir erst in den folgenden Abschnitten 1.5 und 1.6 eingehen.

### 1.4.1 Zusammenhang zweier qualitativer Merkmale

Wir betrachten zuerst den Fall mit *zwei qualitativen* Merkmalen  $X$  und  $Y$ . Das Merkmal  $X$  besitze die  $I$  verschiedenen Ausprägungen  $A_1, \dots, A_I$  und  $Y$  die  $J$  verschiedenen Ausprägungen  $B_1, \dots, B_J$ . Zur Beschreibung des Zusammenhangs beider Merkmale kann man die Verteilung aller Merkmalskombinationen  $(A_i, B_j)$  in der Stichprobe vom Umfang  $n$  durch eine *Kreuztabelle* der beobachteten *Anzahlen*  $n_{ij}$  oder der *relativen Häufigkeiten*  $f_{ij} = \frac{1}{n} n_{ij}$  beschreiben (vgl. Tabelle 1). Zur Veranschaulichung kann man entsprechende (zwei oder dreidimensionale) Säulendiagramme erstellen.

*Tabelle 1: StatLab-Auswahl 1985*  
Kreuztabelle von Geschlecht ( $X$ ) und AB0 =-Blutgruppe ( $Y$ )

Geschlecht	0	$A$	$B$	$AB$	Summe
	(1)	(2)	(3)	(4)	
weiblich (1)	23%	20%	4%	3%	50%
männlich (2)	18%	18%	11%	3%	50%
insgesamt	41%	38%	15%	6%	100%

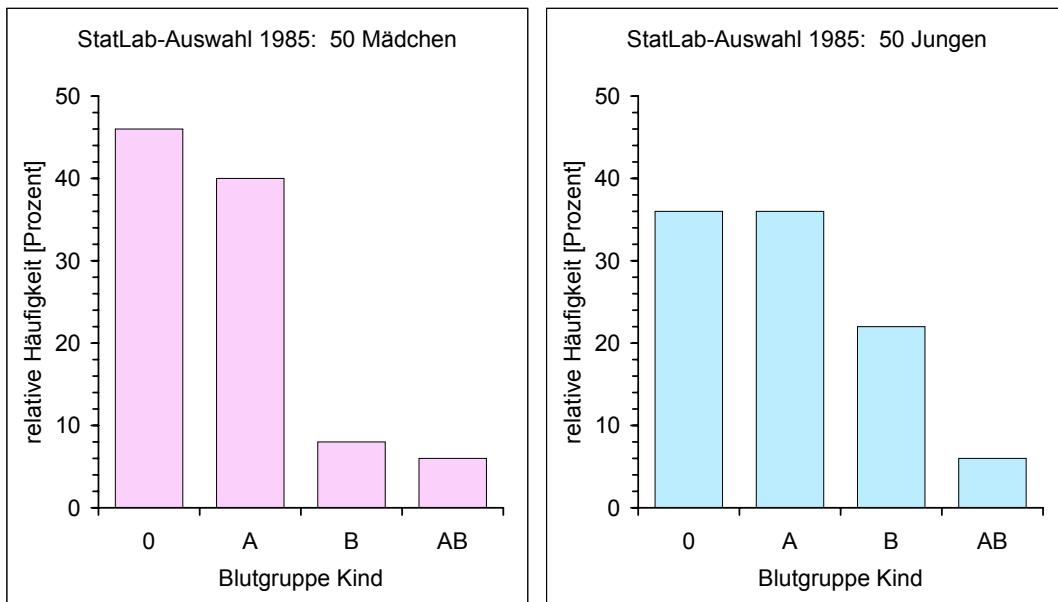
Will man die Verteilung von  $Y$  innerhalb der verschiedenen Ausprägungen von  $X$  *miteinander vergleichen*, so kann man die relativen Häufigkeiten *aller* Ausprägungen  $y_j$  *innerhalb einer festen* Ausprägung  $x_i$  von  $X$  bestimmen und diese in einer Tabelle zusammenfassen (vgl. Tabelle 2) und graphisch darstellen (vgl. Abbildung 1). Umgekehrt kann man auch die entsprechende Tabelle der Verteilung von  $X$  in den verschiedenen  $Y$ -Gruppen aufstellen (vgl. Tabelle 3).

**Beispiel: StatLab-Auswahl 1985**Tabelle 2: Verteilung der AB0-Blutgruppe ( $Y$ ) nach Geschlecht ( $X$ )

Geschlecht	0	A	B	AB	Summe
weiblich	46%	40%	8%	6%	100%
männlich	36%	36%	22%	6%	100%
insgesamt	41%	38%	15%	6%	100%

**Abb.1:** Verteilung der AB0-Blutgruppe ( $Y$ ) nach Geschlecht ( $X$ )

Es gibt einen deutlichen Unterschied zwischen den Geschlechtern bei Blutgruppe  $B$  (vgl. auch Tabelle 3).

Tabelle 3: Verteilung des Geschlechts ( $X$ ) nach AB0-Blutgruppe ( $Y$ )

Geschlecht	0	A	B	AB	Summe
	(1)	(2)	(3)	(4)	
weiblich (1)	56,1%	52,6%	26,7%	50,0%	50%
männlich (2)	49,3%	47,4%	73,3%	50,0%	50%
insgesamt	100%	100%	100%	100%	100%

### 1.4.2 Zusammenhang zwischen einem *qualitativen* und einem *quantitativen* Merkmal

Wir betrachten jetzt den Fall, daß  $X$  ein *quantitatives* Merkmal (mit entsprechenden  $x$ -Werten) und  $Y$  ein *qualitatives* Merkmal ist mit  $J$  verschiedenen Ausprägungen  $B_1, \dots, B_J$ . Zur Darstellung des Zusammenhangs beider Merkmale zerlegt man den Datensatz zunächst entsprechend den Ausprägungen von  $Y$  in  $J$  Teildatensätze (*Y-Gruppen*), in denen  $Y$  jeweils *dieselbe* Ausprägung hat. In jeder *Y-Gruppe* kann man dann die Verteilung der  $x$ -Werte durch Tabellen, Graphiken oder Maßzahlen beschreiben und dabei *Vergleiche* zwischen den *Y-Gruppen* anstellen.

Im einfachsten Fall ist  $Y$  ein *alternatives* Merkmal (z.B. das Geschlecht) und es gibt dann nur *zwei* *Y-Gruppen* (z.B. die 50 Mädchen und 50 Jungen der Statlab-Auswahl). Ein Vergleich der  $x$ -Mittelwerte (oder Mediane) beider *Y-Gruppen* zeigt dann, ob das Merkmal  $Y$  einen Einfluß auf die mittlere *Lage* der  $x$ -Werte und somit auf das Merkmal  $X$  hat. Entsprechend kann man auch die *Streuung* der  $x$ -Werte in beiden Gruppen miteinander vergleichen. Im Fall mit *mehr* als zwei *Y-Gruppen* (z.B. die vier *ABO-Blutgruppen*) kann man je *zwei* *Y-Gruppen* auf diese Weise miteinander vergleichen.

#### **Beispiel: StatLab-Auswahl 1985**

#### **Körpergröße $X$ zum Kontrollzeitpunkt und Geschlecht $Y$**

Zur Untersuchung des Zusammenhangs von Körpergröße  $X$  (zum Kontrollzeitpunkt) und Geschlecht  $Y$  werden die Mittelwerte und Standardabweichungen der Körpergröße für beide Geschlechter (*Y-Gruppen*) separat bestimmt. Aus Tabelle 4 ergibt sich, daß sich die mittleren Körpergrößen beider Geschlechter nur gering unterscheiden, aber die Standardabweichung bei den Mädchen um fast 20% größer ist als bei den Jungen. Bei diesen Daten zeigt sich also ein *geschlechtspezifischer Unterschied* bzgl. der *Streuung* der Körpergröße, der aber nicht sehr stark ausgeprägt ist.

*Tabelle 4:* Mittelwert und Standardabweichung der Körpergröße  $X$  nach Geschlecht ( $Y$ ) getrennt und insgesamt

Geschlecht	Anzahl $n$	Mittelwert $\bar{x}$ [Zoll]	Standardabweichung $s_x$ [Zoll]
weiblich	50	53,23	2,68
männlich	50	53,12	2,27
insgesamt	100	53,17	2,48

## 1.5 Linearer Zusammenhang zweier quantitativer Merkmale

Wir wollen uns jetzt mit dem *Zusammenhang* zweier Merkmale  $X$  und  $Y$  beschäftigen, wenn *beide* Merkmale *quantitativ* sind. Für jedes Untersuchungsobjekt fassen wir den  $x$ - und  $y$ -Wert zu einem geordneten *Paar*  $(x, y)$  zusammen. Eine Stichprobe vom Umfang  $n$  besteht dann aus den Wertepaaren  $(x_1, y_1), \dots, (x_n, y_n)$  aller  $n$  Untersuchungsobjekte.

Um sich einen ersten Überblick über die Daten zu verschaffen, wird man die  $(x, y)$ -Paare als sogenannte *Punktwolke* in einem Koordinatensystem darstellen. Eine solche Graphik heißt auch  $(x, y)$ -*Plot*. Zur besseren Übersicht verwendet man hierbei geeignete *Abschnitte* der  $x$ - bzw.  $y$ -Achse, die den jeweiligen *Nullpunkt* typischerweise nur dann enthalten, wenn dieser nicht zu weit von den beobachteten Werten entfernt ist.

Unser Ziel ist es zu untersuchen, ob trotz der Streuung der Punkte ein systematischer Zusammenhang zwischen den  $x$ - und den  $y$ -Werten erkennbar ist. Genauer interessieren wir uns dafür, ob sich die  $y$ -Werte aus den  $x$ -Werten unter Verwendung einer geeigneten Funktion  $y = f(x)$  *vorhersagen* lassen. Hierbei werden wir nicht erwarten, daß der aus dem Wert  $x_i$  prognostizierte Funktionswert  $f(x_i)$  *exakt* mit dem beobachteten Wert  $y_i$  übereinstimmt, aber letzterer sollte “im Mittel” um den Funktionswert  $f(x_i)$  streuen. Wir wollen uns in diesem Abschnitt zunächst mit *linearen Trends* beschäftigen, d.h. *lineare* Funktionen  $y = f(x) = a + bx$  betrachten. Im nächsten Abschnitt werden wir dann weitere Funktionen durch geeignete Transformationen auf lineare Funktionen zurückführen.

Man beachte, daß die beiden Merkmale bei dieser Suche nach einer Funktion  $y = f(x)$  nicht mehr gleichberechtigt sind:  $x$  wird jetzt als *Einflußvariable* und  $y$  als *Zielvariable* angesehen. In konkreten Anwendungen ist es oft von vornherein klar, welche der beiden betrachteten Merkmale die *Einflußgröße* und welches die *Zielgröße* ist. Prinzipiell kann man aber die Fragestellung auch umkehren und nach einer Funktion  $x = g(y)$  suchen. Hierdurch ergeben sich aber methodisch keine neuen Aspekte, weil man durch einen Tausch beider Merkmale (Umbenennung von  $X$  zu  $Y$  und umgekehrt) wieder bei der ursprünglichen Fragestellung landet.

### 1.5.1 Die Regressionsgerade

Die einfachste Form der Abhängigkeit der *Zielvariablen*  $y$  von der *Einflußvariablen*  $x$  ist durch eine *lineare* Funktion gegeben:

$$y = f(x) = a + bx \quad (\text{lineare Funktion}).$$

Geometrisch stellt diese Funktion eine *Gerade* dar. Dabei ist  $a = f(0)$  der *Achsenabschnitt* (Schnittpunkt der Geraden mit der  $y$ -Achse, die durch den Nullpunkt der  $x$ -Achse gegeben ist). Und der *Anstieg*  $b$  quantifiziert den *Zuwachs* von  $y$ , wenn  $x$  um eine Einheit wächst:

$f(x + 1) = f(x) + b$ . Allgemeiner führt eine *additive* Änderung der Einflußvariablen auch zu einer *additiven* Änderung der Zielvariablen

$$f(x + \Delta) = f(x) + b\Delta.$$

Wenn der Anstieg *positiv* ist, d.h.  $b > 0$ , so liegt eine *ansteigende* Gerade vor, d.h. mit wachsenden  $x$ -Werten wachsen auch die  $y$ -Werte und man spricht von einem *positiven linearen Trend (Zusammenhang)*. Entsprechend liegt bei *negativem* Anstieg (*fallende Gerade*) ein *negativer linearen Trend (Zusammenhang)* vor: bei wachsenden  $x$ -Werten fallen die  $y$ -Werte. Der Fall  $b = 0$  (*kein Anstieg*) entspricht einer zur  $x$ -Achse *parallelen* Geraden  $y = a$ , und hier liegt *kein Trend (Zusammenhang)* vor, weil  $y$  *konstant* gleich  $a$  ist und somit nicht von  $x$  abhängt.

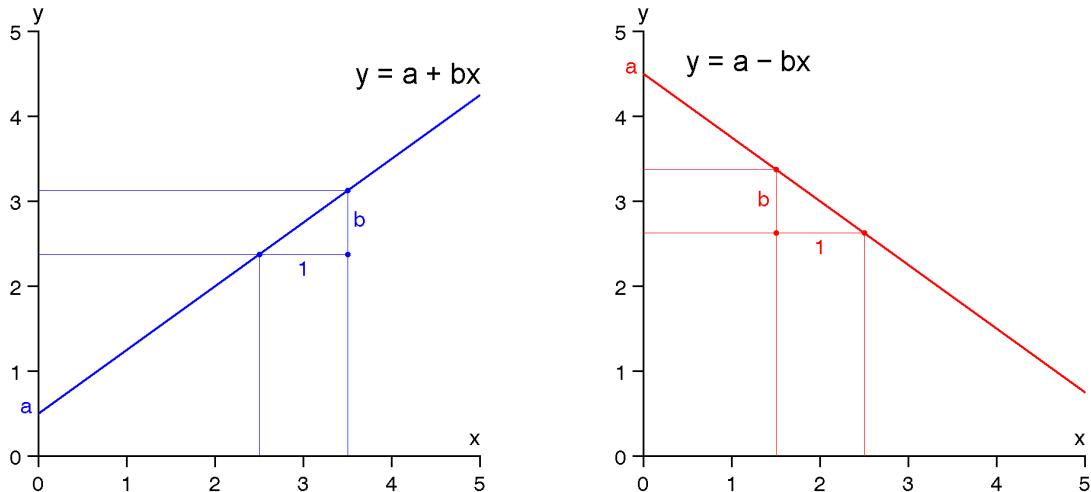


Abb. 1: Gerade mit *positivem* Anstieg (links) und *negativem* Anstieg (rechts).

Da sich für jede Wahl der Parameter  $a$  und  $b$  eine andere Gerade ergibt, stellt sich die Frage, welche dieser Geraden die beobachteten  $(x, y)$ -Punktewolke am besten beschreibt. Die Präzisierung der Formulierung „am besten“ erfolgt nun durch die Methode der kleinsten Quadrate, die uns „optimale“ Werte dieser Parameter  $a$  und  $b$  für eine Punktewolke liefert.

### **Die Methode der kleinsten Quadrate:**

Unser Ziel ist es, für jeden *beobachteten*  $x$ -Wert  $x_i$  den  $y$ -Wert durch den Funktionswert  $f(x_i) = a + bx_i$  vorherzusagen, wobei dieser möglichst wenig vom *beobachteten*  $y$ -Wert  $y_i$  abweichen soll. Folglich soll der *Vorhersage-Fehler*

$$y_i - f(x_i) = y_i - (a + bx_i)$$

einen möglichst kleinen *Absolutbetrag* haben bzw. sein *Quadrat* soll möglichst gering sein. Die nach der Methode der kleinsten Quadrate optimalen Werte für  $a$  und  $b$  werden nun so bestimmt (vgl. Abb. 2), daß die folgende *Summe quadratischer* Vorhersage-Fehler *minimal* wird

$$SQ(a, b) = \sum_i (y_i - (a + bx_i))^2 \quad \longrightarrow \quad \text{Minimum.}$$

Diese Quadratsumme nimmt für die folgenden Werte ihr Minimum an:

$$\begin{aligned} \hat{b} &= \frac{Sxy}{Sxx} & \text{mit} & \begin{aligned} Sxy &= \sum_i (x_i - \bar{x})(y_i - \bar{y}) \\ Sxx &= \sum_i (x_i - \bar{x})^2 \end{aligned} \\ \hat{a} &= \bar{y} - \hat{b}\bar{x} & \text{bzw.} & \bar{y} = \hat{a} + \hat{b}\bar{x}. \end{aligned}$$

Die zugehörige *Regressionsgerade* (*Trend-Gerade*)  $y = \hat{a} + \hat{b}x$  geht immer durch den *Mittelpunkt*  $(\bar{x}, \bar{y})$  der  $(x, y)$ -Werte (vgl. Abb. 4). Sie kann stets berechnet werden, falls mindestens *zwei verschiedene*  $x$ -Werte vorliegen, d.h. falls  $Sxx > 0$  ist.

### **Interpretation:**

Die Regressionsgerade  $y = \hat{a} + \hat{b}x$  ist die (nach der Methode der kleinsten Quadrate) *optimale* Gerade durch die beobachteten Punkte. Sie beschreibt *im Mittel* die *lineare* Abhängigkeit des Merkmals  $y$  vom Merkmal  $x$  und wird für *Prognosen* (Vorhersage von  $y$  aus  $x$ ) benutzt.

*Warnung:* Vertauscht man  $x$  mit  $y$  und fragt nach der linearen Abhängigkeit des Merkmals  $X$  vom Merkmal  $Y$ , so ergibt sich eine Regressionsgerade  $x = \hat{c} + \hat{d}y$ , die im allgemeinen *nicht mit obiger Geraden*  $y = \hat{a} + \hat{b}x$  übereinstimmt (vgl. Abb. 3).

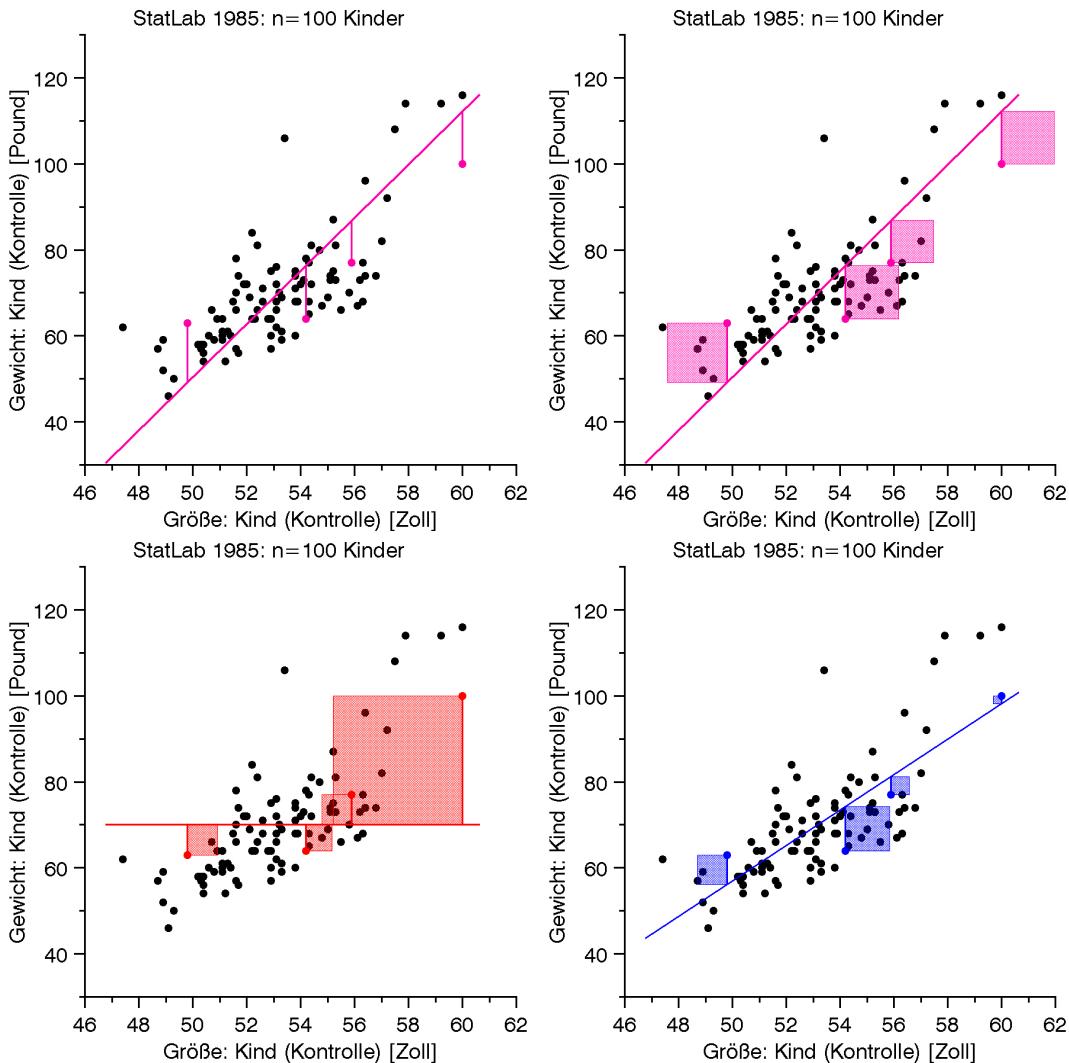
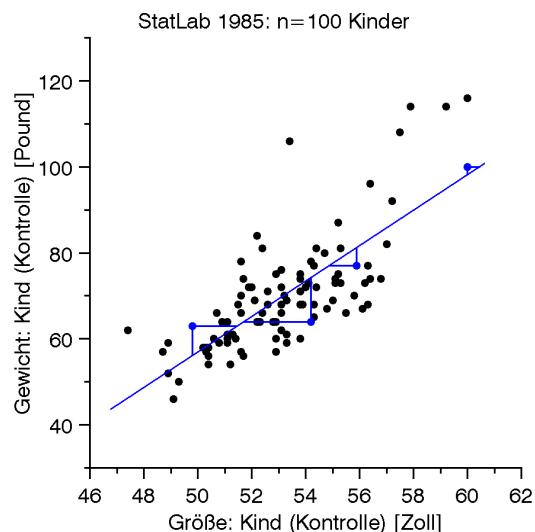


Abb. 2: Vorhersage-Fehler (als Strecken) und ihre Quadrate (als Flächen) für einige Punkte bei drei verschiedenen Geraden. Unten links ist die Mittelwertsgerade  $y = \bar{y}$  (d.h.  $a = \bar{y}, b = 0$ ) und rechts die Regressionsgerade  $y = \hat{a} + \hat{b}x$  dargestellt.

Abb. 3:  
Vorhersage-Fehler der  $y$ -Werte und die davon (im allgemeinen) verschiedenen Vorhersage-Fehler der  $x$ -Werte einiger Punkte.



### 1.5.2 Streuung um die Regressionsgerade, Bestimmtheitsmaß

Obwohl die Regressionsgerade den linearen Trend für die beobachteten Punkte beschreibt, bleibt noch offen, wie *gut* die Regressionsgerade die Streuung der  $y$ -Werte *erklärt*. Oder präziser gefragt: Wie groß ist die *Streuung der  $y$ -Werte um die Regressionsgerade* im Verhältnis *zur Streuung um den  $y$ -Mittelwert*?

Die *Streuung der  $y$ -Werte um die Regressionsgerade* (vgl. Abb. 2 unten rechts) entspricht dem *Minimum* der Summe quadratischer Vorhersage-Fehler

$$SQ(\hat{a}, \hat{b}) = \sum_i (y_i - (\hat{a} + \hat{b}x_i))^2.$$

Und die *Streuung der  $y$ -Werte um den Mittelwert  $\bar{y}$*  ist (vgl. Abb. 2 unten links)

$$\begin{aligned} Syy &= \sum_i (y_i - \bar{y})^2 \\ &= SQ(\bar{y}, 0) \geq SQ(\hat{a}, \hat{b}), \end{aligned}$$

wobei sich letzte Ungleichung daraus ergibt, daß  $SQ(\hat{a}, \hat{b})$  das *Minimum* von  $SQ(a, b)$  bzgl.  $a$  und  $b$  ist. Die Differenz dieser Streuungen ist die *absolute Streuungsreduktion*

$$Syy - SQ(\hat{a}, \hat{b}) \geq 0.$$

Der *relative Anteil der Streuungs-Reduktion an der Gesamtstreuung  $Syy$*  heißt das *Bestimmtheitsmaß (der Regressionsgeraden)*:

$$B_{xy} = \frac{Syy - SQ(\hat{a}, \hat{b})}{Syy} = 1 - \frac{SQ(\hat{a}, \hat{b})}{Syy} \quad (\text{Bestimmtheitsmaß})$$

Damit ist das Bestimmtheitsmaß  $B_{xy}$  ein dimensionsloses Maß für den Grad (die Stärke) des linearen Zusammenhangs  $y = a + bx$ . Es wird meist in Prozent angegeben. Ist z.B. (vgl. Abb. 5)  $B_{xy} = 80\%$ , so beträgt die Streuung um die Regressionsgerade nur 20% (d.h.  $100\% - 80\%$ ) *der Streuung um den  $y$ -Mittelwert* und man sagt dann auch suggestiv, daß die Regressionsgerade 80% *der Streuung der  $y$ -Werte um ihren Mittelwert* erklärt. Je *größer* das Bestimmtheitsmaß  $B_{xy}$  ist, desto *besser* wird die Streuung der  $y$ -Werte durch die Regressionsgerade *erklärt*.

#### **Eigenschaften des Bestimmtheitsmaßes $B_{xy}$ einer Regressionsgeraden**

1.  $B_{xy}$  ist nur für  $Sxx > 0$  und  $Syy > 0$  definiert, d.h. wenn *nicht alle  $x$ -Werte übereinstimmen und nicht alle  $y$ -Werte übereinstimmen*.
2. Das Bestimmtheitsmaß liegt stets zwischen 0 und 1 (inklusive):

$$0 \leq B_{xy} \leq 1 \quad \text{bzw. } 0\% \leq B_{xy} \leq 100\%.$$

3.  $B_{xy} = 1$  bzw.  $B_{xy} = 100\%$  gilt genau dann, wenn die Streuung  $SQ(\hat{a}, \hat{b})$  um die Regressionsgerade *gleich Null* ist, d.h. wenn *alle* Punkte  $(x_i, y_i)$  *exakt* auf der Regressionsgeraden  $y = \hat{a} + \hat{b}x$  liegen.
  4.  $B_{xy} = 0$  bzw.  $B_{xy} = 0\%$  gilt genau dann, wenn die Regressionserade  $\bar{y} = \hat{a} + \hat{b}x$  *exakt* mit der parallel zur  $x$ -Achse verlaufenden „Mittelwerts-Geraden“  $y = \bar{y}$  übereinstimmt, d.h. es ist  $\hat{b} = 0$  und  $\hat{a} = \bar{y}$ .
  5. Bei der *Beurteilung* von  $B_{xy}$  ist die *Anzahl*  $n$  der Punkte mit zu berücksichtigen: z.B. liegen  $n = 2$  Punkte *stets* auf der Regressionsgeraden, d.h.  $B_{xy} = 1$ .  
Als *Faustregel* (die in **3.6.1 b** begründet wird) kann gelten:

Für  $B_{xy} > \frac{4}{n+2}$  liegt eine bedeutsame (signifikante) Streuungsreduktion vor.

- ## 6. Alternative Darstellung zur Berechnung:

$$B_{xy} = \frac{(Sxy)^2}{Sxx \cdot Syy}.$$

7. Das Bestimmtheitsmaß ist *symmetrisch* in beiden Merkmalen:

$$B_{xy} = B_{yx}.$$

8. Das Bestimmtheitsmaß ist *invariant* gegenüber *linearen* Umskalierungen der  $x$ -Werte zu  $u = c_1 + d_1x$  und der  $y$ -Werte zu  $v = c_2 + d_2y$  mit  $d_1, d_2 \neq 0$ :

$$B_{uv} = B_{xy}.$$

### 1.5.3 Der Korrelationskoeffizient

Als Maß für den *Grad des linearen Zusammenhangs* der Merkmale  $X$  und  $Y$  wird meistens nicht das *Bestimmtheitsmaß*  $B_{xy}$  sondern der sogenannte *Korrelationskoeffizient (von Pearson)* angegeben:

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} \quad (\text{Korrelationskoeffizient}).$$

Eine Interpretation des Korrelationskoeffizienten ergibt sich daraus, daß der Korrelationskoeffizient die *Wurzel aus dem Bestimmtheitsmaß* ist, wobei sein *Vorzeichen* mit dem des Anstiegs  $\hat{b}$  der Regressionsgeraden übereinstimmt:

$$r_{xy} = \begin{cases} +\sqrt{B_{xy}} & \text{falls } \hat{b} \geq 0 \\ -\sqrt{B_{xy}} & \text{falls } \hat{b} \leq 0 \end{cases}, \quad B_{xy} = r_{xy}^2.$$

Man beachte, daß im Fall  $\hat{b} = 0$  (kein Anstieg der Regressionsgeraden) bzw.  $S_{xy} = 0$  beide Zusammenhangsmaße den Wert Null annehmen:  $r_{xy} = 0$  und  $B_{xy} = 0$ .

Der Korrelationskoeffizient ist, ebenso wie das Bestimmtheitsmaß, eine *dimensionslose Zahl*. In der Praxis ist es sinnvoll, *beide Maße* anzugeben.

### **Korrelationskoeffizient vs. Bestimmtheitsmaß**

Ein Vorteil des Korrelationskoeffizienten ist, daß sein *Vorzeichen* gleichzeitig die *Richtung* des linearen Trends angibt: *positives* bzw. *negatives* Vorzeichen bedeuten einen *positiven* bzw. *negativen* linearen Trend. Demgegenüber läßt sich das Bestimmtheitsmaß direkt als (*prozentualer*) Anteil der durch die Regressionsgerade *erklärten Streuung* interpretieren. Darüberhinaus läßt sich ein Bestimmtheitsmaß auch analog für *nicht-lineare Funktionen*  $y = f(x)$  definieren (vgl. 1.6.1), wo kein dem Korrelationskoeffizient entsprechendes Maß zur Verfügung steht.

### **Eigenschaften des Korrelationskoeffizienten $r_{xy}$**

1.  $r_{xy}$  ist nur für  $S_{xx} > 0$  und  $S_{yy} > 0$  definiert, d.h. wenn *nicht alle x-Werte übereinstimmen und nicht alle y-Werte übereinstimmen*.
2. Der Korrelationskoeffizient liegt stets zwischen  $-1$  und  $+1$  (inklusive):

$$-1 \leq r_{xy} \leq +1 \quad (\text{Normierung}).$$

3.  $r_{xy} = +1$  gilt genau dann, wenn *alle* Punkte  $(x_i, y_i)$  auf der Regressionsgeraden  $y = \hat{a} + \hat{b}x$  liegen und diese *steigend* ist, d.h.  $\hat{b} > 0$ .

$r_{xy} = -1$  gilt genau dann, wenn *alle* Punkte  $(x_i, y_i)$  auf der Regressionsgeraden  $y = \hat{a} + \hat{b}x$  liegen und diese *fallend* ist, d.h.  $\hat{b} < 0$ .

4.  $r_{xy} = 0$  gilt genau dann, wenn weder ein steigender noch ein fallender *linearer* Trend in der Punktfolge  $(x_i, y_i)$  vorhanden ist, d.h. die Regressionsgerade ist  $y = \bar{y}$  und liegt parallel zur  $x$ -Achse ( $\hat{b} = 0$ ).

5. Bei der Beurteilung von  $r_{xy}$  ist die *Anzahl n* der Punkte mit zu berücksichtigen, z.B. liegen  $n = 2$  Punkte *stets* auf der Regressionsgeraden, d.h.  $|r_{xy}| = 1$ .

Als *Faustregel* (die in 3.6.1 b begründet wird) kann gelten:

Für  $|r_{xy}| > \frac{2}{\sqrt{n+2}}$  liegt ein bedeutsamer (signifikanter) linearer Trend vor.

6. Alternative Darstellung zur Berechnung:

$$r_{xy} = \hat{b} \sqrt{\frac{S_{xx}}{S_{yy}}} = \hat{b} \frac{s_x}{s_y} \quad .$$

7. Der Korrelationskoeffizient ist *symmetrisch* in beiden Merkmalen:

$$r_{xy} = r_{yx}.$$

8. Bis auf sein Vorzeichen ist der Korrelationskoeffizient *invariant* gegenüber *linearen* Umskalierungen der  $x$ -Werte zu  $u = c_1 + d_1x$  und der  $y$ -Werte zu  $v = c_2 + d_2y$  mit  $d_1, d_2 \neq 0$ :

$$r_{uv} = r_{xy} \quad \text{falls } d_1 \text{ und } d_2 \text{ beide positiv oder beide negativ sind.}$$

$$r_{uv} = -r_{xy} \quad \text{falls von } d_1, d_2 \text{ ein Wert positiv und der andere negativ ist.}$$

Die Abbildungen 4–5 zeigen verschiedene Punktwolken mit ihren Regressionsgeraden und Angaben zum Bestimmtheitsmaß und Korrelationskoeffizient.

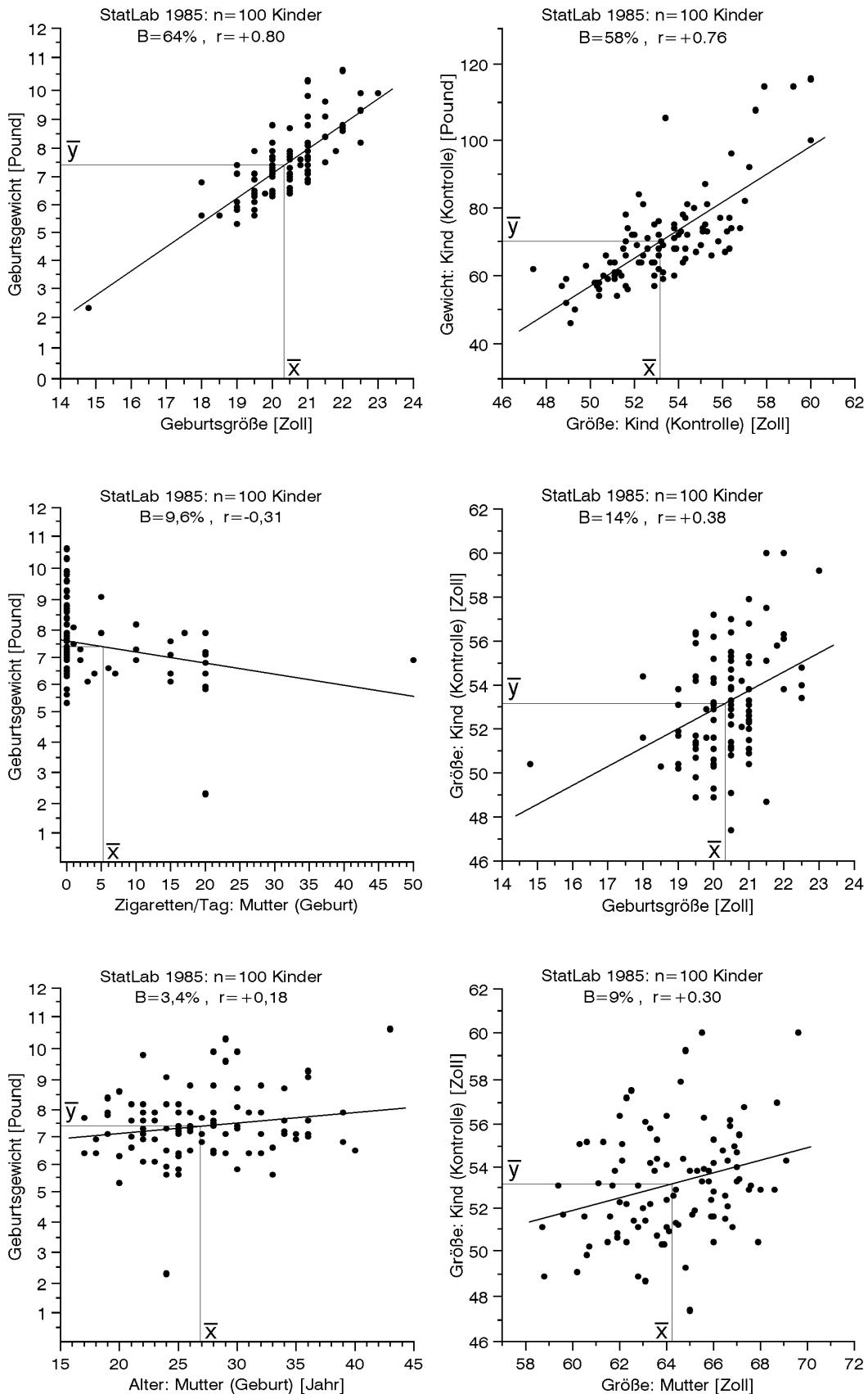


Abb. 4: Punktwolken verschiedener Merkmale der StatLab-Auswahl 1985 mit Regressionsgerade, Bestimmtheitsmaß  $B$  und Korrelationskoeffizient  $r$ .

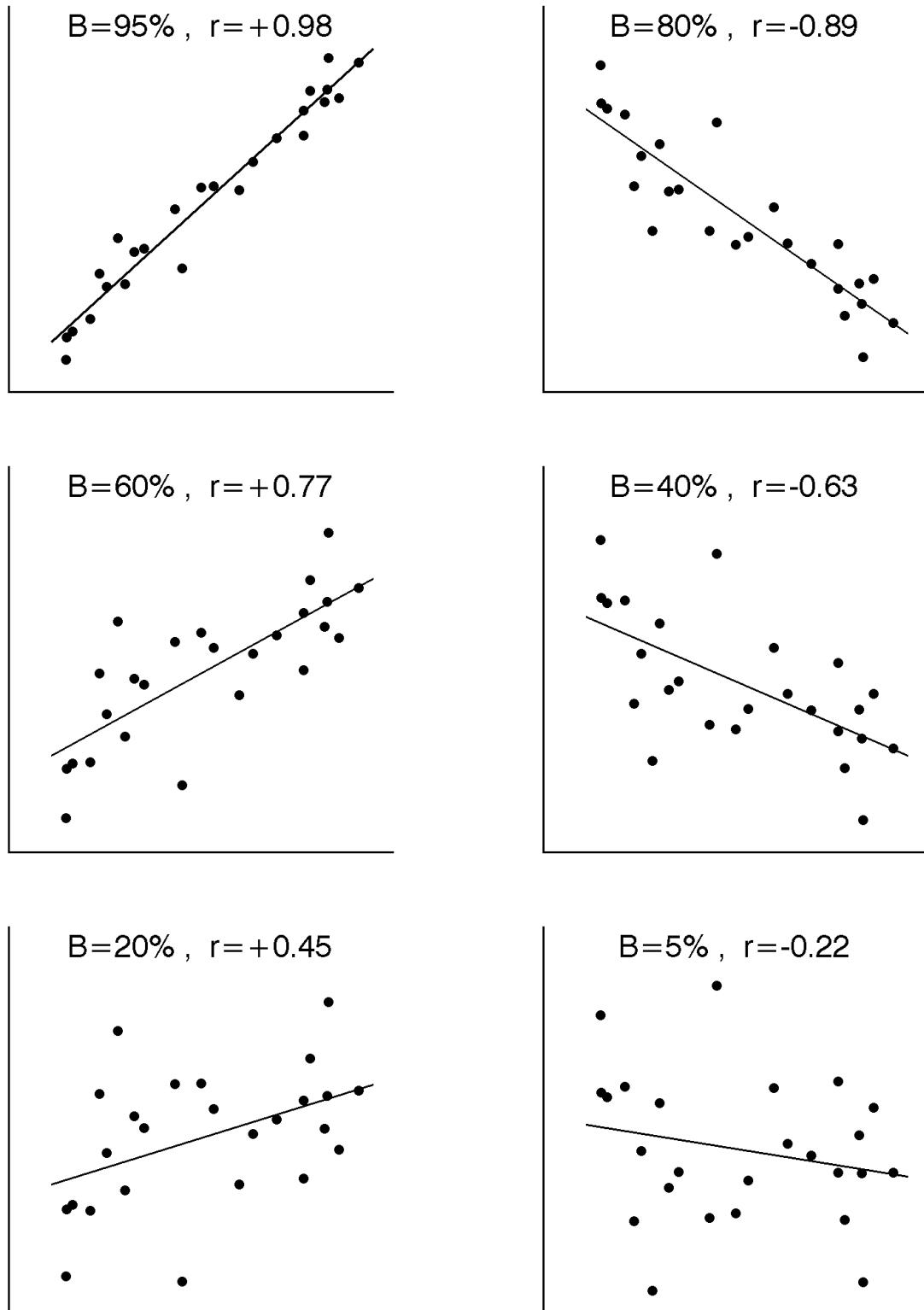


Abb. 5: Regressionsgerade, Bestimmtheitsmaß  $B$  und Korrelationskoeffizient  $r$  verschiedener (fiktiver) Punktwolken vom Umfang  $n = 25$ .

Die Beschriftung der Achsen ist weggelassen, weil die Skalierung hier keine Rolle spielt .

### 1.5.4 Berechnung: Regressionsgerade, Bestimmtheitsmaß und Korrelationskoeffizient

Durch algebraische Umformungen kann man zeigen, daß sich die Regressionsgerade, das Bestimmtheitsmaß und der Korrelationskoeffizient relativ einfach berechnen lassen aus den bereits bekannten Summen  $\sum x, \sum x^2$  für die  $x$ -Werte (vgl. 1.3.3), den entsprechenden Summen  $\sum y, \sum y^2$  für die  $y$ -Werte sowie der gemischten Summe  $\sum xy$  aller Produkte von  $x$ - mit  $y$ -Werten. Die zugehörigen Rechenschritte sind in der *Rechen-Box 2* zusammengefaßt.

<b><i>Rechen-Box 2 für Datenpaare (<math>x,y</math>)</i></b>									
<i>Stichprobenumfang:</i>	$n$								
<i>x-Werte:</i>	$x_1, x_2, \dots, x_n$								
<i>y-Werte:</i>	$y_1, y_2, \dots, y_n$								
<i>Summen:</i>									
<i>x-Werte:</i>	$\sum x = x_1 + x_2 + \dots + x_n$								
<i>y-Werte:</i>	$\sum y = y_1 + y_2 + \dots + y_n$								
<i>x-Quadrate:</i>	$\sum x^2 = x_1^2 + x_2^2 + \dots + x_n^2$								
<i>y-Quadrate:</i>	$\sum y^2 = y_1^2 + y_2^2 + \dots + y_n^2$								
<i>xy-Produkte:</i>	$\sum xy = x_1y_1 + x_2y_2 + \dots + x_ny_n$								
<i>Mittelwerte:</i>	$\bar{x} = \frac{1}{n}(\sum x),$	$\bar{y} = \frac{1}{n}(\sum y)$							
<i>Summen von Abweichungs-Quadraten und -Produkten:</i>									
$S_{xx} = \sum_i (x_i - \bar{x})^2$	$= (\sum x^2) - \frac{1}{n}(\sum x)^2$								
$S_{yy} = \sum_i (y_i - \bar{y})^2$	$= (\sum y^2) - \frac{1}{n}(\sum y)^2$								
$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y})$	$= (\sum xy) - \frac{1}{n}(\sum x)(\sum y)$								
<i>Regressionsgerade <math>y = \hat{a} + \hat{b}x</math>:</i>	$\hat{b} = \frac{S_{xy}}{S_{xx}}, \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$								
<i>Streuung um die Gerade:</i>	$SQ(\hat{a}, \hat{b})$	$= S_{yy} - \hat{b} \cdot S_{xy}$							
<i>Bestimmtheitsmaß:</i>	$B_{xy}$	$= \frac{(S_{xy})^2}{S_{xx} \cdot S_{yy}} = r_{xy}^2$							
<i>Korrelationskoeffizient:</i>	$r_{xy}$	$= \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$							

### Skalenverschiebung:

Zur Vermeidung großer Zahlen (beim Quadrieren) kann man die  $x$ - und  $y$ -Werte jeweils verschieben und mit den verschobenen Werten  $u = x - c$  und  $v = y - d$  rechnen. Dann gilt (vgl. auch 1.3.3)

$$S_{xx} = S_{uu}, \quad S_{yy} = S_{vv}, \quad S_{xy} = S_{uv}, \quad B_{xy} = B_{uv}, \quad r_{xy} = r_{uv}.$$

### Regeln zur Rechengenauigkeit

Neben den allgemeinen Regeln aus 1.3.3 ist hier zusätzlich zu beachten:

$\sum x$  bzw.  $\sum y$  hat dieselbe Anzahl Nachkommastellen wie  $x$  bzw.  $y$ .

$\sum x^2, S_{xx}$  bzw.  $\sum y^2, S_{yy}$  haben die doppelte Anzahl Nachkommastellen wie  $x$  bzw.  $y$ .

Die Zahl der Nachkommastellen von  $\sum xy, S_{xy}$  ist die Summe der jeweiligen Anzahlen von Nachkommastellen von  $x$  und  $y$ .

Das Bestimmtheitsmaß wird meist auf volle Prozente gerundet, weitere Nachkommastellen sind nur sinnvoll, wenn der Wert dicht bei 0% oder 100% liegt. Der Korrelationskoeffizient wird mit 2 (bis maximal 3) Nachkommastellen angegeben.

### 1.5.5 Anwendung: StatLab-Auswahl 1985

#### Geburtsgröße und Geburtsgewicht der Mädchen

Für die  $n = 50$  Mädchen soll die Abhängigkeit des Geburtsgewichts  $Y$  [Pound] von der Geburtsgröße  $X$  [Zoll] untersucht werden. Die zur Berechnung nach Rechen-Box 2 erforderlichen Größen sind (auszugsweise) in Tabelle 1 angegeben, aus der sich die folgenden Zwischen- und Endergebnisse bestimmen lassen:

$$\text{Summen: } S_{xx} = 67,0008; \quad S_{xy} = 59,7980; \quad S_{yy} = 82,6450$$

$$\text{Regressionsgerade } y = \hat{a} + \hat{b}x : \quad \hat{a} = -10,83; \quad \hat{b} = 0,8925$$

$$\text{Bestimmtheitsmaß, Korrelationskoeffizient: } B = 65\% \quad r = 0,80.$$

Die Punktwolke mit den Mittelwerten und der Regressionsgeraden ist in Abb. 6 dargestellt. Die Punktwolke zeigt einen klaren *positiven* Trend, und durch die Regressionsgerade wird  $B = 65\%$  der Gesamtstreuung des Geburtsgewichts durch die Geburtsgröße erklärt. Man beachte, daß sich durch eine Umskalierung der Größe [cm statt Zoll] und des Gewichts [kg statt Pound] zwar die Achsen-Skalierungen ändern, aber *nicht* das Bestimmtheitsmaß  $B$  und der Korrelationskoeffizient  $r$ .

Tabelle 1: Werte für Geburtsgröße  $x$  [Zoll] und Geburtsgewicht  $y$  [Pound]

Nr.	$x$	$y$	$x^2$	$y^2$	$x \cdot y$
1	19,8	6,4	392,04	40,96	126,72
2	19,5	6,3	380,25	39,69	122,85
3	19,5	7,1	380,25	50,41	138,45
4	20,0	7,3	400,00	53,29	146,00
5	19,0	7,1	361,00	50,41	134,90
:	:	:	:	:	:
45	21,5	9,1	462,25	82,81	195,65
46	19,0	5,9	361,00	34,81	112,10
47	20,0	7,1	400,00	50,41	142,00
48	21,0	8,2	441,00	67,24	172,20
49	20,0	7,4	400,00	54,76	148,00
50	21,5	7,5	462,25	56,25	161,25
$\Sigma$	1006,4	356,4	20323,82	2624,49	7235,43
Mittelwert	20,13	7,13			
Stand.abw.	1,17	1,30			

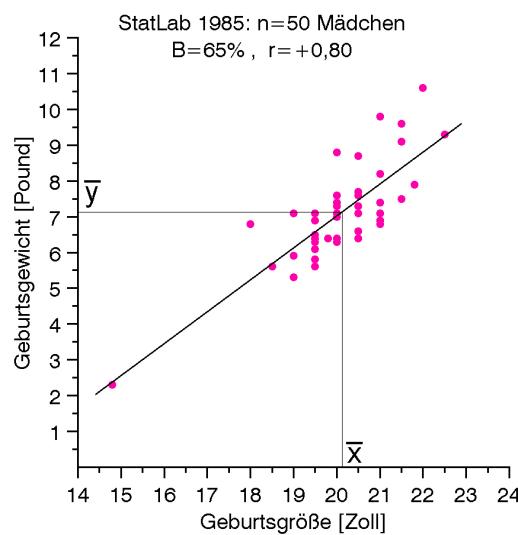


Abb. 6: Zusammenhang von Geburtsgröße und Geburtsgewicht für die Mädchen der StatLab-Auswahl 1985: Punkte, Mittelwerte, Regressionsgerade, Bestimmtheitsmaß  $B$  und Korrelationskoeffizient  $r$ .

## Tipps für *Microsoft Excel*

**Funktionen:** Zur Berechnung der (*ungewichteten*) Regressionsgeraden stehen folgende *Statistik-Funktionen* zur Verfügung (mit Erläuterung in Klammern):

STEIGUNG (= $\hat{b}$ )	ACHSENABSCHNITT (= $\hat{a}$ )
BESTIMMTHEITSMASS	KORREL (=Korrelationskoeffizient)

**Diagramme:** Punktfolgen (wie in Abb. 4–6) erstellt man mit dem *Diagramm-Assistenten* als *Standardtyp Punkt (XY)*. Die (*ungewichtete*) Regressionsgerade zeichnet man (nach Selektieren des Diagramms) mit dem Menüpunkt *Diagramm - Trendlinie hinzufügen* (sie kann auch „manuell“ als Gerade gezeichnet werden, vgl. 1.5.7).

**Tabelle:** Zum *Erlernen und besseren Verständnis* ist es sinnvoll, die Regressionsgerade in *Excel* zunächst wie im Beispiel mit einer *Tabelle* und der *Rechen-Box 2* zu bestimmen, wobei man obige *Excel-Funktionen* zur *Kontrolle* verwenden kann (vgl. hierzu auch die Tipps zur *gewichteten* Regression in 1.5.7).

### 1.5.6 Gewichtung von Beobachtungen

Bei der Bestimmung der Regressionsgeraden kann die Methode der kleinsten Quadrate dadurch verfeinert werden, daß man *nicht* alle Wertepaare  $(x_1, y_1), \dots, (x_n, y_n)$  *gleichberechtigt* berücksichtigt. Eine *unterschiedliche Gewichtung* einzelner Paare durch einen nicht-negativen *Gewichtsfaktor*  $w_i$  (engl: *weight*) ist immer dann angebracht, wenn die Präzision (Streuung) der  $y$ -Werte mit den  $x$ -Werten variiert und dies bei der Analyse berücksichtigt werden soll. Typische Situationen hierfür sind:

- Jeder Wert  $y_i$  ist bereits ein *Mittelwert* über eine Anzahl  $n_i$  von *Einzelwerten* (die aber nicht mehr *einzeln* vorliegen). Sofern nicht alle Anzahlen  $n_i$  übereinstimmen, sollte man sie als Gewichte  $w_i = n_i$  verwenden (vgl. das Beispiel in 1.5.7), um die unterschiedliche Präzision der  $y_i$  (bedingt durch die unterschiedliche Anzahlen  $n_i$ ) mit zu berücksichtigen.
- Die  $y$ -Werte sind *relative Häufigkeiten* der Form  $y_i = r_i/n_i$ . Dann sollte man die unterschiedliche Streuung der  $y$ -Werte dadurch berücksichtigen, daß man die Gewichte  $w_i = n_i/(y_i(1 - y_i))$  verwendet (dies ist die *geschätzte reziproke Varianz* der zugehörigen *relativen Häufigkeit*, vgl. Abschnitt 3.1.1). Hierbei darf  $y_i$  allerdings weder 0 noch 1 sein. Ein Beispiel findet man in 1.6.7.
- Wenn statt der  $y$ -Werte *transformierte* Werte  $v(y)$  verwendet werden (vgl. 1.6), so sollte man die unterschiedliche Streuung der  $v$ -Werte durch Gewichtsfaktoren (die sich aus der Transformation  $v(y)$  ergeben) mit berücksichtigen. Beispiele hierzu finden sich in 1.6.4.

Unter Berücksichtigung von Gewichtsfaktoren werden Parameter  $a$  und  $b$  der *gewichteten Regressionsgeraden* jetzt nach der Methode der *gewichteten kleinsten Quadrate* bestimmt durch Minimierung der *gewichteten Summe quadratischer* Vorhersage-Fehler

$$WSQ(a, b) = \sum_i w_i \cdot (y_i - (a + bx_i))^2 \quad \longrightarrow \quad \text{Minimum.}$$

Paare mit größerem Gewichtsfaktor werden dann stärker berücksichtigt, weil ihr Vorhersage-Fehler in der gewichteten Quadratsumme  $WSQ(a, b)$  stärker zu Buche schlägt. Wenn im Extremfall ein einzelnes Gewicht  $w_i$  gleich Null ist, so wird der zugehörige Punkt  $(x_i, y_i)$  in der Summe  $WSQ(a, b)$  überhaupt nicht berücksichtigt, und kann deshalb von vornherein weggelassen werden.

Die Koeffizienten  $\hat{a}$  und  $\hat{b}$  der so bestimmten gewichteten Regressionsgeraden  $y = \hat{a} + \hat{b}x$  ergeben sich analog dem ungewichteten Fall in **1.5.1**. Die erforderlichen Schritte für die Berechnung aller wichtigen Größen bei gewichteten Datenpaaren sind in der *Rechen-Box 3* zusammengestellt.

Ein Vergleich mit der *Rechen-Box 2* für den ungewichteten Fall zeigt, daß bei den Summen hier alle Summanden zusätzlich den Gewichtsfaktor enthalten und daß bei den weiteren Rechnungen (außer bei den Varianzen) die *Summe der Gewichte*  $\sum w$  statt des *Stichprobenumfangs*  $n$  verwendet wird.

Bei der Gewichtung kommt es nicht auf die *absolute* Größe der Gewichtsfaktoren  $w_i$  an, sondern nur auf deren *relative* Größe zueinander an. Multipliziert man alle Gewichte mit einem Faktor, so läßt sich dieser Faktor bei allen Summen in der *Rechen-Box 3* ausklammern („vor die Summe ziehen“) und kürzt sich bei der Berechnung der Mittelwerte, Varianzen, Regressionsgeraden, Bestimmtheitsmaß und Korrelationskoeffizient wieder heraus. Wenn insbesondere *alle Gewichtsfaktoren übereinstimmen*, so ergeben sich dieselben Ergebnisse wie im *ungewichteten* Fall (d.h. alle  $w_i = 1$ ) nach *Rechen-Box 2*.

In einer graphischen Darstellung kann man die Gewichtung mit berücksichtigen, indem man die *Punktgröße* in einem  $(x, y)$ -Plot proportional zum *Gewichtsfaktor* wählt (vgl. Abb. 7).

<b>Rechen-Box 3 für Datenpaare <math>(x,y)</math> mit Gewichtsfaktor <math>w</math></b>									
Stichprobenumfang:	$n$								
$x$ -Werte:	$x_1, x_2, \dots, x_n$								
$y$ -Werte:	$y_1, y_2, \dots, y_n$								
Gewichtsfaktoren:	$w_1, w_2, \dots, w_n$								
Summen:									
Gewichte:	$\sum w$	=	$w_1 + w_2 + \dots + w_n$						
$x$ -Werte:	$\sum wx$	=	$w_1x_1 + w_2x_2 + \dots + w_nx_n$						
$y$ -Werte:	$\sum wy$	=	$w_1y_1 + w_2y_2 + \dots + w_ny_n$						
$x$ -Quadrate:	$\sum wx^2$	=	$w_1x_1^2 + w_2x_2^2 + \dots + w_nx_n^2$						
$y$ -Quadrate:	$\sum wy^2$	=	$w_1y_1^2 + w_2y_2^2 + \dots + w_ny_n^2$						
$xy$ -Produkte:	$\sum wxy$	=	$w_1x_1y_1 + w_2x_2y_2 + \dots + w_nx_ny_n$						
gewichtete Mittelwerte:	$\bar{x} = \frac{1}{\sum w}(\sum wx),$			$\bar{y} = \frac{1}{\sum w}(\sum wy)$					
gewichtete Summen von Abweichungs-Quadraten und -Produkten:									
$Swxx$	$= \sum_i w_i \cdot (x_i - \bar{x})^2$	$= (\sum wx^2) - \frac{1}{\sum w}(\sum wx)^2$							
$Swyy$	$= \sum_i w_i \cdot (y_i - \bar{y})^2$	$= (\sum wy^2) - \frac{1}{\sum w}(\sum wy)^2$							
$Swxy$	$= \sum_i w_i \cdot (x_i - \bar{x})(y_i - \bar{y})$	$= (\sum wxy) - \frac{1}{\sum w}(\sum wx)(\sum wy)$							
gewichtete (empirische) Varianzen:	$s_x^2 = \frac{Swxx}{n-1}$			$s_y^2 = \frac{Swyy}{n-1}$					
gewichtete Regressionsgerade $y = \hat{a} + \hat{b}x$ :									
$\hat{b} = \frac{Swxy}{Swxx},$	$\hat{a} = \bar{y} - \hat{b}\bar{x}$								
Streuung um die Gerade:	$WSQ(\hat{a}, \hat{b})$	=	$Swyy - \hat{b} \cdot Swxy$						
Bestimmtheitsmaß (gewichtet):	$B_{xy}$	=	$\frac{(Swxy)^2}{Swxx \cdot Swyy}$	=	$r_{xy}^2$				
Korrelationskoeffizient (gewichtet):	$r_{xy}$	=	$\frac{Swxy}{\sqrt{Swxx \cdot Swyy}}$						

### 1.5.7 Anwendung

**Blutdruck und Alter** (vgl. Kinder-Osius-Timm 1982, Bsp. 7.34)

Bei 83 Männern wurden das Alter und der systolische Blutdruck bestimmt, wobei der Altersbereich von 15 bis 65 Jahren zur Vereinfachung in  $n = 10$  Fünf-Jahres-Klassen aufgeteilt ist, und nur noch das mittlere Alter  $x$  pro Klasse und der *Mittelwert*  $y$  des Blutdrucks zur Verfügung stehen (Tabelle 2). Da der Umfang  $w$  der Altersklassen (und damit auch die Präzision des mittleren Bludrucks  $y$ ) stärker variiert, wird er als *Gewichtsfaktor* bei der Berechnung der Regressionsgeraden berücksichtigt. Die für die *Rechen-Box* 3 erforderlichen Größen sind in Tabelle 2 enthalten und daraus ergeben sich die folgenden Zwischen- und Endergebnisse:

$$\text{Summen: } Swxx = 14215,0602; \quad Swxy = 11094,8976; \quad Swyy = 8942,2641$$

$$\text{Regressionsgerade } y = \hat{a} + \hat{b}x : \quad \hat{a} = 99,01; \quad \hat{b} = 0,7805$$

*Bestimmtheitsmaß, Korrelationskoeffizient:*

$$B = 97\%, \quad r = 0,98.$$

Die (gewichteten) Punkte sind mit den (gewichteten) Mittelwerten und der (gewichteten) Regressionsgeraden in Abb. 7 dargestellt. Die Punkte streuen nur gering um die Regressionsgerade, und bereits  $B = 97\%$  der Streuung des Blutdrucks wird durch das Alter erklärt.

Eine Faustregel besagt

$$\text{Blutdruck} = 100 + \text{Alter in Jahren.}$$

Dies entspricht einer Geraden mit  $a = 100$  und  $b = 1$ , die den Parametern  $\hat{a}$  und  $\hat{b}$  obiger Regressionsgerade entsprechen, wenn man diese auf *ganze Zahlen rundet*.

Tabelle 2: Werte für Alter  $x$  [Jahre] und Blutdruck  $y$  [mm Hg]

Nr.	Um- fang	Alter [Jahre]	Blutdruck [mm Hg]	(berechnete Werte)					
				$w$	$x$	$y$	$wx$	$wy$	$wx^2$
1	5	17,5	112,6	87,5	563,0	1531,25	63393,80	9852,50	
2	8	22,5	118,1	180,0	944,8	4050,00	111580,88	21258,00	
3	10	27,5	120,4	275,0	1204,0	7562,50	144961,60	33110,00	
4	7	32,5	122,3	227,5	856,1	7393,75	104701,03	27823,25	
5	11	37,5	127,9	412,5	1406,9	15468,75	179942,51	52758,75	
6	10	42,5	135,1	425,0	1351,0	18062,50	182520,10	57417,50	
7	8	47,5	133,6	380,0	1068,8	18050,00	142791,68	50768,00	
8	11	52,5	138,1	577,5	1519,1	30318,75	209787,71	79752,75	
9	8	57,5	144,1	460,0	1152,8	26450,00	166118,48	66286,00	
10	5	62,5	151,2	312,5	756,0	19531,25	114307,20	47250,00	
$\Sigma$	83	400,0	1303,4	3337,5	10822,5	148418,75	1420104,99	446276,75	

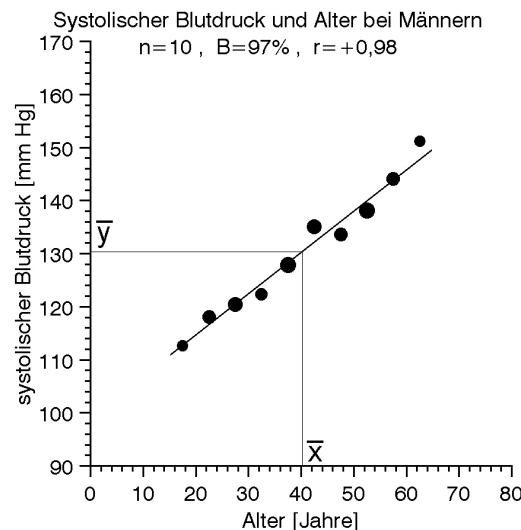


Abb. 7:

Zusammenhang von Blutdruck und Alter: gewichtete Punkte (die Größe entspricht dem Gewichtsfaktor  $w$ ), Mittelwerte, Regressionsgerade, Bestimmtheitsmaß  $B$  und Korrelationskoeffizient  $r$ .

### ☒ Tipps für Microsoft Excel

**Warnung:** Gewichtete Regressionsgeraden werden von *Excel nicht direkt* durch Befehle unterstützt, sondern müssen vom *Benutzer* bestimmt und gezeichnet werden. (Die Tipps in **1.5.5** beziehen sich nur auf *ungewichtete* Regressionsgeraden.)

**Berechnung:** Mit einer Tabelle (wie z.B. Tab. 2) bestimmt man zunächst alle *gewichteten* Summen  $\sum w, \dots, \sum wxy$  (vgl. Rechen-Box 3), aus denen man dann alle weiteren Größen berechnen kann.

Man kann hierbei sogar auf die Spalten mit den Produkten  $wx, \dots, wxy$  verzichten, weil sich die zugehörigen Spaltensummen  $\sum wx, \dots, \sum wxy$  auch direkt mit der *Excel-Funktion SUMMENPRODUKT* aus den Spalten für  $w, x$  und  $y$  berechnen lassen. Zum besseren Verständnis ist jedoch die vollständige Tabelle mit allen Spalten empfehlenswert, wobei man die Spaltensummen  $\sum wx, \dots, \sum wxy$  zusätzlich unter Verwendung der Funktion **SUMMENPRODUKT** kontrollieren kann.

**Diagramme:** Ein Diagramm mit *unterschiedlichen Punktgrößen* (entsprechend ihrer Gewichtung) lässt sich zwar mit dem Diagrammtyp *Blase* erzeugen, aber die Regressionsgerade kann dort *nicht* mit eingezeichnet werden (Abb. 7 ist auch nicht in Excel erstellt). Deshalb stellt man zuerst – wie bisher – die Punktwolke mit gleichgroßen Punkten dar. Die *gewichtete* Regressionsgerade wird dann hinzugefügt, indem man *zwei* Punkte auf der Geraden (in *Excel*) berechnet, wobei man typischerweise die zugehörigen Punkte für den *kleinsten* und *größten*  $x$ -Wert auf der  $x$ -Achse wählt. Dann werden diese beiden Geraden-Punkte als neue *Datenreihe* dem Diagramm hinzugefügt unter dem Menüpunkt *Diagramm - Datenquelle - Reihe - hinzufügen* (nach Selektieren des Diagramms). Die beiden Punkte werden dann wie folgt durch eine Gerade verbunden. Durch Rechts-Klick auf einer der beiden Punkte erscheint ein Kontext-Menü, in dem man *Datenreihe formatieren* auswählt. Im dann erscheinenden Fenster wählt man unter Muster im Feld Linie den Punkt *Automatisch* oder *Benutzerdefiniert* (wenn Farbe und Linientyp selbst gewählt werden sollen). Die beiden Endpunkte der Geraden sollte man „ausblenden“, indem man im Feld *Markierung* den Punkt *Ohne* aktiviert.

## 1.6 Funktionaler Zusammenhang und linearisierende Transformationen

Für die Abhängigkeit einer quantitativen Zielvariablen  $Y$  von einer quantitativen Einflußvariablen  $X$  haben wir bisher nur lineare Funktionen

$$y = f(x) = a + bx \quad (\text{linearer Zusammenhang})$$

behandelt. Wir wollen jetzt auch *nicht-lineare* Funktionen betrachten, z.B.

$$\begin{aligned} y &= f(x) = a + b \log(x) && (\text{logarithmischer Zusammenhang}), \\ y &= f(x) = Ae^{bx} && (\text{exponentieller Zusammenhang}), \\ y &= f(x) = Ax^b && (\text{Potenz-Zusammenhang}). \end{aligned}$$

Der *logarithmische* Zusammenhang zwischen  $y$  und  $x$  ist gleichzeitig ein *linearer* Zusammenhang zwischen  $y$  und dem *Logarithmus*  $u = \log(x)$  von  $x$ :

$$y = a + bu.$$

Damit haben wir den logarithmischen Zusammenhang unter Verwendung einer Transformation der  $x$ -Werte auf einen linearen Zusammenhang zwischen  $y$  und der transformierten Variablen  $u = \log(x)$  zurückgeführt.

Der *exponentielle* Zusammenhang läßt sich ebenfalls als ein *linearer* Zusammenhang zwischen dem natürlichen Logarithmus  $v = \ln(y)$  und  $x$  beschreiben

$$v = a + bx. \quad \text{mit} \quad a = \ln(A).$$

Und auch der *Potenz-Zusammenhang* ist ein *linearer* Zusammenhang zwischen den logarithmisch transformierten Variablen  $u = \log(x)$  und  $v = \log(y)$

$$v = a + bu. \quad \text{mit} \quad a = \log(A).$$

Unter Verwendung anderer Transformationen (statt des Logarithmus) lassen sich weitere nicht-lineare Zusammenhänge zwischen  $x$  und  $y$  auf lineare Zusammenhänge der transformierten Variablen zurückführen. Wir werden im folgenden einige typische solcher linearisierenden Transformationen mit Anwendungsbeispielen betrachten. Allerdings läßt sich leider nicht *jeder* nicht-lineare Zusammenhang durch geeignete Transformationen linearisieren. Wir werden deshalb (exemplarisch) auch den *quadratischen* Zusammenhang

$$y = f(x) = a + bx + cx^2 \quad (\text{quadratischer Zusammenhang})$$

behandeln, der sich *nicht* durch Transformationen von  $x$  und  $y$  linearisieren läßt.

Im Folgenden wird zunächst erläutert, wie man aus den beobachteten Wertepaaren  $(x_1, y_1), \dots, (x_n, y_n)$  analog zur Regressionsgeraden eine optimale *Regressionsfunktion* für die jeweilige Form des Zusammenhangs bestimmen kann.

### 1.6.1 Regressionsfunktion und Bestimmtheitsmaß

Unser Ausgangspunkt ist wieder eine Stichprobe mit den beobachteten Wertepaaren  $(x_1, y_1), \dots, (x_n, y_n)$  der beiden quantitativen Merkmale  $X$  und  $Y$ . Die Abhängigkeit des  $y$ -Wertes vom  $x$ -Wert soll „im Mittel“ durch eine Funktionsgleichung  $y = f(x)$  gegeben sein, d.h. bei vorgegebenem  $x$ -Wert beschreibt  $f(x)$  den *erwarteten (mittleren)*  $y$ -Wert. Für jede spezifische Art des Zusammenhangs (z.B. den linearen oder quadratischen Zusammenhang) ist die Funktion  $f(x)$  durch ihre *Parameter*  $a, b, \dots$  vollständig bestimmt. Analog zur Bestimmung der Regressionsgeraden (für den linearen Zusammenhang) wollen wir jetzt die optimalen Werte der Parameter  $a, b, \dots$  aus den beobachteten Daten bestimmen (*schätzen*).

#### **Die Methode der kleinsten Quadrate**

Unter allen Werten für die Parameter  $a, b, \dots$  der Funktion  $f(x)$  wollen wir - analog zur Regressionsgeraden - wieder diejenigen bestimmen, bei der die aus  $x_i$  prognostizierten Funktionswerte  $f(x_i)$  möglichst wenig von den beobachteten Werten  $y_i$  abweichen, d.h. der *Vorhersage-Fehler*  $y_i - f(x_i)$  soll möglichst gering sein. Hierzu werden die Parameter  $a, b, \dots$  der Funktion so gewählt (*geschätzt*), daß die zugehörige *Summe quadratischer Vorhersage-Fehler* *minimal* wird

$$SQ(a, b, \dots) = \sum_i (y_i - f(x_i))^2 \longrightarrow \text{Minimum.}$$

Für lineare Funktionen  $f(x) = a + bx$  erhält man so die Parameter  $\hat{a}$  und  $\hat{b}$  der *Regressionsgeraden*. Für *nicht-lineare* Funktionen ist die Lösung dieses Minimierungs-Problems mathematisch aufwendiger, und oft können die Lösungen  $\hat{a}, \hat{b}, \dots$  nur durch ein Iterations-Verfahren schrittweise approximiert werden. Die zugehörige (*geschätzte*) Funktion  $\hat{f}$  mit den so *geschätzten* Parametern  $\hat{a}, \hat{b}, \dots$  heißt auch die *Regressionsfunktion*. Sie läßt sich nur dann bestimmen, wenn die Anzahl der *verschiedenen*  $x$ -Werte mindestens so groß ist wie die Anzahl der zu schätzenden Parameter (z.B. beim quadratischen Zusammenhang wenn mindestens *drei verschiedene*  $x$ -Werte beobachtet wurden).

Wenn zusätzlich *Gewichtsfaktoren*  $w_i$  (vgl. 1.5.6) für jedes Paar  $(x_i, y_i)$  berücksichtigt werden sollen, so wird die Summe der *gewichteten* quadratischen Vorhersage-Fehler minimiert

$$WSQ(a, b, \dots) = \sum_i w_i [y_i - f(x_i)]^2 \longrightarrow \text{Minimum.}$$

Wir wollen im folgenden stets von der allgemeineren Situation mit *Gewichtsfaktoren*  $w_i$  ausgehen. Die *ungewichtete* Situation ist als Spezialfall mit den *konstanten* Gewichten  $w_i = 1$  (für alle  $i$ ) in der allgemeinen Situation enthalten, und wir werden auf entsprechende Vereinfachungen im ungewichteten Fall nur kurz hinweisen.

### **Streuung um die Regression-Funktion und Bestimmtheitsmaß**

Nachdem man die Regressionsfunktion bestimmt hat, taucht wieder die Frage auf, wie gut die Regressionsfunktion die Streuung der  $y$ -Werte erklärt. Oder präziser gefragt: Wie groß ist die *Streuung der  $y$ -Werte um die Regressionsfunktion* im Verhältnis zur *Streuung um den  $y$ -Mittelwert*?

Die (gewichtete) *Streuung der  $y$ -Werte um die Regressionsfunktion* entspricht wieder dem *Minimum* der gewichteten Summe quadratischer Vorhersage-Fehler

$$WSQ(\hat{a}, \hat{b}, \dots) = \sum_i w_i [y_i - \hat{f}(x_i)]^2.$$

Und die (gewichtete) *Streuung der  $y$ -Werte um den (gewichteten) Mittelwert  $\bar{y}$*

$$Swyy = \sum_i w_i [y_i - \bar{y}]^2$$

ist immer dann mindestens so groß wie die Streuung  $WSQ(\hat{a}, \hat{b})$  um die Regressionsfunktion, wenn die betrachteten Funktionenklasse durch spezielle Wahl der Parameter  $a, b, \dots$  auch alle *konstanten Funktionen* enthält. Bei den oben betrachteten Formen des Zusammenhangs liegt z.B. immer dann eine konstante Funktion vor, wenn mit Ausnahme von  $a$  alle restlichen Parameter  $b, \dots$  den Wert Null annehmen. Generell wird jede Form des Zusammenhangs sinnvollerweise auch den Spezialfall enthalten, daß  $y = f(x)$  konstant ist (d.h. eine parallel zur  $x$ -Achse verlaufende Gerade darstellt) und somit  $y$  überhaupt nicht von  $x$  abhängt. Wenn dies der Fall ist, dann definiert man (analog 1.5.2) den *relativen Anteil der Streuungs-Reduktion an der Gesamtstreuung* als das

*Bestimmtheitsmaß der Regressionsfunktion:*

$$B = \frac{Swyy - WSQ(\hat{a}, \hat{b}, \dots)}{Swyy} = 1 - \frac{WSQ(\hat{a}, \hat{b}, \dots)}{Swyy}$$

Bei *ungewichteten* Punkten vereinfachen sich die *gewichteten* Summen  $Swyy$  und  $WSQ$  zu den *ungewichteten* Summen  $Syy$  und  $SQ$ .

Von den Eigenschaften des Bestimmtheitsmaßes für Regressionsgeraden (vgl. 1.5.2) gelten hier auch *einige* (aber nicht *alle*) entsprechend:

#### **Eigenschaften des Bestimmtheitsmaßes $B$ einer Regressionsfunktion**

1.  $B$  ist nur dann *definiert*, wenn nicht alle  $y$ -Werte übereinstimmen und mindestens so viele verschiedene  $x$ -Werte vorliegen wie Parameter  $a, b, \dots$  zu schätzen sind.

2. Das Bestimmtheitsmaß liegt stets zwischen 0 und 1 (inklusive):

$$0 \leq B \leq 1 \quad \text{bzw.} \quad 0\% \leq B \leq 100\%$$

3.  $B = 1$  bzw.  $B = 100\%$  gilt genau dann, wenn die Streuung um die Regressionsfunktion *gleich Null* ist, d.h. wenn *alle* Punkte  $(x_i, y_i)$  *exakt* auf der Regressionsfunktion  $y = \hat{f}(x)$  liegen.
4.  $B = 0$  bzw.  $B = 0\%$  gilt genau dann, wenn die Regressionsfunktion  $y = \hat{f}(x)$  *exakt* mit der parallel zur  $x$ -Achse verlaufenden „Mittelwerts-Geraden“  $y = \bar{y}$  übereinstimmt.
5. Bei der *Beurteilung* von  $B$  ist die *Anzahl*  $n$  der Punkte mit zu berücksichtigen. Stimmt  $n$  z.B. mit der Anzahl der zu schätzenden Parameter  $a, b, \dots$  überein, so liegen alle Punkte auf der Regressionsfunktion, d.h.  $B = 1$ .

Man beachte daß die nicht genannten Eigenschaften (6. bis 8. aus **1.5.2**) des Bestimmtheitsmaßes bei *nicht-linearen* Funktionen *nicht* mehr allgemein gültig sind. Darüber hinaus gibt es auch kein dem *Korrelationskoeffizienten* entsprechendes Pendant bei *nicht-linearen* Funktionen.

### 1.6.2 Die logarithmische Transformation

In diesen Abschnitt wollen wir einige wichtige Eigenschaften des Logarithmus zusammenstellen. Der *Logarithmus zu einer Basis*  $c > 0$  wird mit  $\log_c$  bezeichnet, d.h.

$$y = \log_c(x) \quad \Leftrightarrow \quad x = c^y. \\ (\text{ist äquivalent zu})$$

Prinzipiell kann *jede positive Zahl*  $c \neq 1$  als Basis verwendet werden, aber üblicherweise benutzt man einen der folgende Werte

- Basis  $e = 2,7183\dots$  (*Eulersche Zahl*),  $\ln = \log_e$  heißt *natürlicher Logarithmus*,
- Basis 10,  $\log_{10}$  heißt *dekadischer Logarithmus*,
- Basis 2,  $\log_2$  heißt *dualer Logarithmus*.

Der natürliche Logarithmus wird besonders bei theoretischen Betrachtungen bevorzugt. Dagegen ist der dekadische Logarithmus bei konkreten Rechnungen übersichtlicher, weil die *Vorkommastelle* von  $\log_{10}(x)$  gleichzeitig die *Größenordnung* von  $x$  im üblichen Zehner-System (d.h. die Zehner-Potenz) angibt:

wenn  $\log_{10}(x)$  zwischen 2 und 3 liegt, so liegt  $x$  zwischen  $10^2 = 100$  und  $10^3 = 1000$ .

Logarithmen zur Basis  $c$  lassen sich durch den natürlichen Logarithmus berechnen und umgekehrt

$$\log_c(x) = \frac{\ln(x)}{\ln(c)}, \quad \ln(x) = \frac{\log_c(x)}{\log_c(e)} .$$

Insbesondere gilt für den dekadischen Logarithmus

$$\log_{10}(x) = \frac{\ln(x)}{\ln(10)}, \quad \ln(x) = \frac{\log_{10}(x)}{\log_{10}(e)}.$$

Folglich unterscheiden sich die Logarithmen zu verschiedenen Basen nur um einen festen Faktor, und deshalb kommt es bei *grundlegenden Betrachtungen* meist nicht auf die konkrete Basis an. Wir verwenden in solchen Fällen die Schreibweise “log” (ohne Angabe der Basis  $c$ ), um anzudeuten, daß *jede* Basis verwendet werden kann. Bei konkreten Zahlenwerten wie z.B.  $\log_c(5)$  ist die Angabe der Basis  $c$  aber unverzichtbar, damit klar ist, welches Ergebnis gemeint ist.

Die wichtigste (und zugleich charakteristische) Eigenschaft des Logarithmus (egal zu welcher Basis) ist, daß er die Rechenoperationen vereinfacht, indem er *Produkte* auf *Summen* und *Potenzen* auf *Produkte* reduziert:

$$\log(a \cdot b) = \log(a) + \log(b), \quad \log(b^a) = a \cdot \log(b).$$

Man beachte, daß es *keine* entsprechende Darstellung von  $\log(a + b)$  gibt!

Die *Logarithmus-Funktion (logarithmische Transformation)*  $y = \log_c(x)$  ordnet jeder *positiven* Zahl  $x$  umkehrbar eindeutig eine Zahl  $y \in \mathbb{R}$  zu ( $\mathbb{R}$  bezeichnet die Menge aller *reeller* Zahlen), wobei die Umkehrfunktion durch  $x = c^y$  gegeben ist. Wenn  $x$  gegen 0 abfällt, so konvergiert  $\log_c(x)$  gegen  $-\infty$ . Rein formal kann man setzen:  $\log_c(0) = -\infty$ . Abbildung 1 zeigt die Funktion  $y = \log_{10}(x)$  im Bereich  $0 < x \leq 10$ .

Die *Ableitung* der Logarithmus-Funktion ist

$$\log'_c(x) = \frac{1}{\ln(c)}x^{-1}, \quad \text{speziell} \quad \ln'(x) = x^{-1}.$$

Bei graphischen Darstellungen von Punkten oder Funktionen ist es manchmal vorteilhafter, eine (oder auch beide) Koordinaten-Achsen nicht mit der üblichen *linearen Skala*, sondern mit einer *logarithmischen Skala* (kurz: *log-Skala*) zu versehen. Das Konstruktionsprinzip der log-Skala unter Verwendung der Logarithmus-Funktion ist in Abbildung 1 dargestellt. Das Wesentliche einer log-Skala ist, daß auf ihr die *Abstände* zweier Zahlen erhalten bleiben, wenn man diese Zahlen mit *irgendeinem Faktor multipliziert*. Konkret haben die Zahlen 0,2 und 0,5 denselben Abstand wie 2 und 5. Der Sprung um *eine Größenordnung*, z.B. von 0,1 auf 1 oder von 1 auf 10, ist auf der log-Skala stets *gleich groß* (vgl. Abb. 1). Aus diesem Grund ist eine log-Skala z.B. immer dann (aber nicht nur dann) sinnvoll, wenn man Bereiche darstellen will, die mehrere Größenordnungen überstreichen. Wir werden auch Darstellungen mit *zwei* verschieden skalierten (linear bzw. logarithmisch)  $x$ -Achsen (unten und oben) und/oder *zwei*  $y$ -Achsen (links und rechts) verwenden, wenn dies zweckmäßig ist.

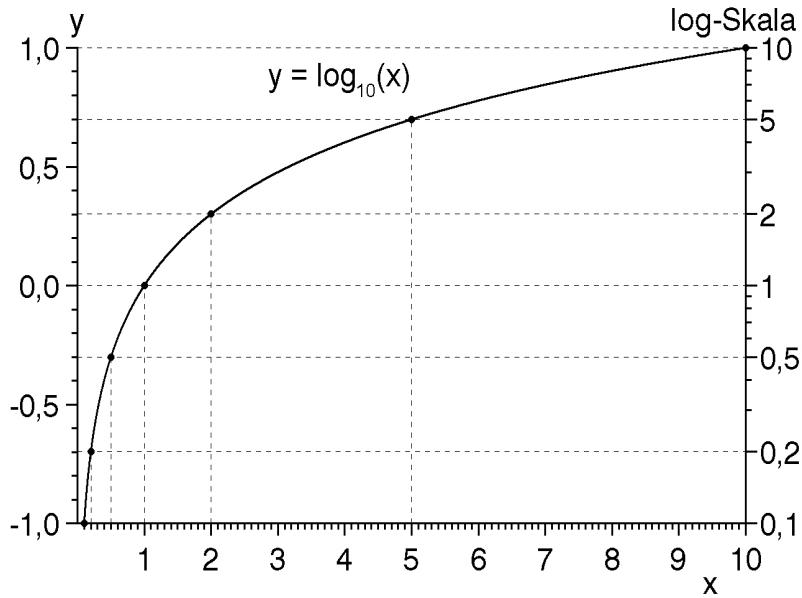


Abb. 1: Die Logarithmus-Funktion  $y = \log_{10}(x)$  und Konstruktion einer *log-Skala* (rechte Vertikale) durch „orthogonale Reflektion“ der  $x$ -Achse an der Funktion.

### 1.6.3 Logarithmischer Zusammenhang

Der *logarithmische* Zusammenhang zwischen einer Zielgröße  $y$  und einer *positiven* Einflußgröße  $x > 0$  ist durch die Funktion gegeben

$$y = f(x) = a + b \log_c(x) \quad (\text{logarithmischer Zusammenhang}),$$

wobei man typischerweise den *dekadischen* oder *natürlichen* Logarithmus verwendet. Ein solcher Zusammenhang ist dadurch charakterisiert, daß eine *multiplikative* Änderung der Einflußgröße  $x$  eine *additive* Änderung der Zielgröße  $y$  bewirkt,

$$f(q \cdot x) = f(x) + b\Delta \quad \text{mit} \quad \Delta = \log_c(q).$$

Speziell für die Basis  $c$  als Faktor  $q$  ergibt sich die Änderung  $\Delta = 1$ .

Betrachtet man nun den Logarithmus der Einflußvariablen  $x$

$$u = \log_c(x),$$

so hängt die Zielvariable  $y$  *linear* von der *transformierten* Einflußgröße  $u$  ab:

$$y = a + bu \quad (\text{linearer Zusammenhang von } y \text{ und } u).$$

Für die beobachteten Wertepaare  $(x_1, y_1), \dots, (x_n, y_n)$  – ggf. mit Gewichtsfaktoren – ergeben sich daher die *optimalen* Werte der Parameter  $a$  und  $b$  nach der Methode der kleinsten Quadrate als die Koeffizienten der *Regressionsgeraden*  $y = \hat{a} + \hat{b}u$  für die *transformierten*

Punkte  $(u_i, y_i)$  mit  $u_i = \log_c(x_i)$ . Hieraus erhält man die *Regressionsfunktion (geschätzte Funktion)* für den logarithmischen Zusammenhang

$$y = \hat{f}(x) = \hat{a} + \hat{b} \log_c(x) \quad (\text{geschätzte Funktion}).$$

Das Bestimmtheitsmaß  $B$  für die Regressionsfunktion (nach 1.6.1) stimmt hier mit dem Bestimmtheitsmaß  $B_{uy}$  der Regressionsgeraden  $y = \hat{a} + \hat{b}u$  (nach 1.5) überein, weil nur die Einflußvariable  $x$ , aber *nicht* die Zielvariable  $y$  transformiert wurde.

### Schadstoffgehalt in Möhren

Bei einer Untersuchung über die Inkorporation des Carcinogens 3,4-Benzpyren in Möhren wurde für sechs vorgegebene Konzentrationen  $x$  [in mg pro kg] im Substrat die von der Möhre inkorporierte Menge  $y$  [in ppm] bezogen auf die Trockensubstanz bestimmt (vgl. Kinder-Osius-Timm 1982, Beispiel 7.33). Als Modell wird eine lineare Abhängigkeit der inkorporierten Menge  $y$  von der *logarithmischen* Konzentration  $u = \log_{10}(x)$  – also ein logarithmischer Zusammenhang von  $x$  und  $y$  – vorgeschlagen. Aus den beobachteten Werten (vgl. auch Abb. 2)

3,4-BP Konzentrat $x$ [mg/kg]	0,1	0,3	1,0	3,0	10,0	30,0
Inkorporierte Menge $y$ [ppm]	0,10	0,75	1,30	1,75	2,25	2,85

ergibt sich die Regressionsgerade für die *transformierte* Einflußgröße  $u = \log_{10} x$ :

$$v = \hat{a} + \hat{b}u \quad \text{mit} \quad \hat{a} = 1,244 \quad \text{und} \quad \hat{b} = 1,072.$$

Die *geschätzte Funktion*  $y = \hat{f}(x) = \hat{a} + \hat{b} \cdot \log_{10}(x)$  erklärt immerhin  $B = 99,6\%$  (Bestimmtheitsmaß der Regressionsfunktion) der Streuung der  $y$ -Werte. Allerdings ist der Stichprobenumfang  $n = 6$  auch relativ gering. Die Regressionsgerade und Regressionsfunktion ist in Abb. 2 dargestellt.

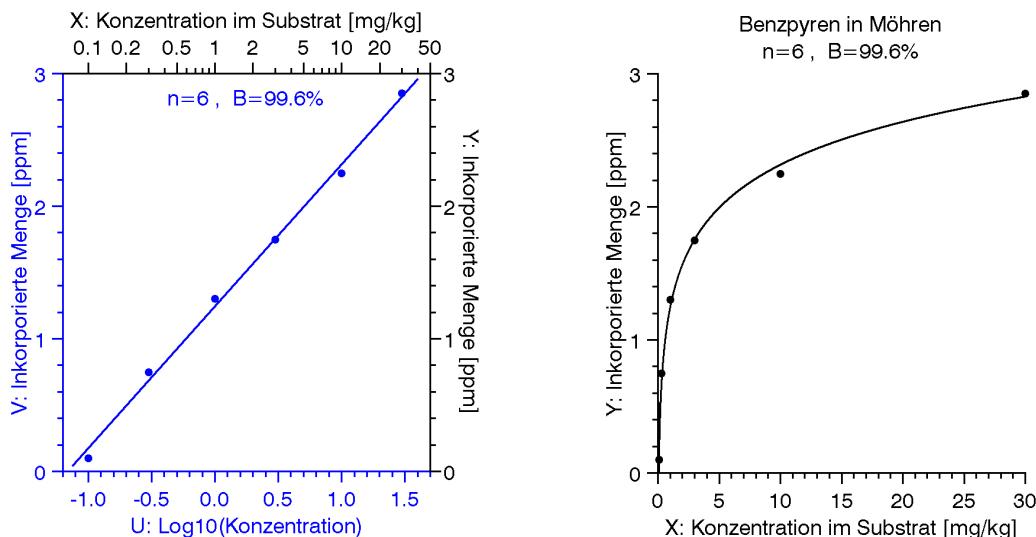


Abb. 2 links: transformierte Punkte und Regressionsgerade (*linearisierte Skala*).  
rechts: beobachtete Punkte mit geschätzter Funktion (*Original-Skala*).

### 1.6.4 Exponentieller Zusammenhang

Die *exponentielle* Abhangigkeit einer *positiven* Zielgroe  $y > 0$  von einer Einflugroe  $x$  ist gegeben durch eine Exponentialfunktion folgender Form

$$y = f(x) = Ae^{bx} \quad (\text{exponentieller Zusammenhang}),$$

wobei der Faktor  $A > 0$  *positiv* ist (vgl. auch das folgende Anwendungs-Beispiel). Ein solcher Zusammenhang ist dadurch gekennzeichnet, da eine *additive* anderung der Einflugroe  $x$  eine *multiplikative* anderung der Zielgroe  $y$  bewirkt

$$f(x + \Delta) = q \cdot f(x) \quad \text{mit} \quad q = e^{b\Delta}$$

Speziell fur die anderung  $\Delta = 1$  um eine Einheit ergibt sich der Faktor  $q = e^b$ .

Eine andere Charakterisierung des exponentiellen Zusammenhangs ist dadurch gegeben, da die *anderung* (Ableitung) der Zielgroe *proportional* zur Zielgroe ist:

$$f'(x) = b \cdot f(x).$$

Betrachtet man nun den naturlichen Logarithmus der Zielvariablen  $y$

$$v = \ln(y),$$

so hangt die *transformierte* Zielvariable  $v$  *linear* von der Einflugroe  $x$  ab,

$$v = a + bx \quad (\text{linearer Zusammenhang von } v \text{ und } x),$$

wobei sich die Parameter  $A$  und  $a$  wie folgt auseinander ergeben:

$$a = \ln(A) \quad \text{bzw.} \quad A = e^a.$$

Fur die beobachteten Wertepaare  $(x_1, y_1), \dots, (x_n, y_n)$  ist es nun naheliegend, die Parameter  $a$  und  $b$  nach der Methode der kleinsten Quadrate als Koeffizienten der *Regressionsgeraden*  $v = \hat{a} + \hat{b}x$  fur die *transformierten* Punkte  $(x_i, v_i)$  mit  $v_i = \ln(y_i)$  zu schatzen. Hierbei wird allerdings die Summe der quadratischen Abweichungen auf der *logarithmisch transformierten* Skala der Zielvariablen minimiert:

$$\sum_i (v_i - (a + bx_i))^2 \longrightarrow \text{Minimum}.$$

Da die *Vorhersage-Fehler* nicht invariant unter der logarithmischen Transformation  $v = \ln(y)$  sind, entspricht dies *nicht exakt* dem Prinzip der minimalen Quadrate auf der *Originalskala* der Zielvariablen

$$\sum_i (y_i - Ae^{bx_i})^2 \longrightarrow \text{Minimum}.$$

Die Minimierung dieser Quadratsumme ist allerdings mathematisch aufwendiger und die zugehörigen optimalen Werte für  $A$  und  $b$  lassen sich nicht explizit durch die beobachteten  $(x, y)$ -Werte darstellen. Man kann aber eine *approximative Lösung* dieser optimalen Werte erhalten, indem man eine *gewichtete Regressionsgeraden*  $v = \hat{a} + \hat{b}x$  bestimmt mit den (von  $y$  abhängigen) Gewichtsfaktoren

$$w_i = y_i^2.$$

Falls die ursprünglichen Werte  $(x_i, y_i)$  bereits mit einer Gewichtung  $w_i^\circ$  vorliegen, so verwendet man das Produkt  $w_i = w_i^\circ y_i^2$  als neues Gewicht der transformierten Werte  $(x_i, v_i)$ .

Aus der so *gewichteten* Regressionsgeraden  $v = \hat{a} + \hat{b}x$  ergibt sich dann als *Regressionsfunktion (geschätzte Funktion)* für den exponentiellen Zusammenhang

$$y = \hat{f}(x) = \hat{A}e^{\hat{b}x} \quad \text{mit} \quad \hat{A} = e^{\hat{a}}.$$

Das Bestimmtheitsmaß  $B$  für die Regressionsfunktion (nach **1.6.1**) stimmt im allgemeinen *nicht* mit dem Bestimmtheitsmaß  $B_{xv}$  der gewichteten Regressionsgeraden  $v = \hat{a} + \hat{b}x$  (nach **1.5.6**) überein, weil die Zielvariable  $y$  transformiert wurde.

### **Abbau von Benzpyren**

Fünf Gruppen von Mäusen wurde die krebserzeugende Substanz Benzpyren in verschiedenen Dosierungen subkutan injiziert. Nach Ablauf von vier Zeiträumen  $x$  [in Tagen] wurde je ein Tier der Gruppe daraufhin untersucht, welche Menge  $y$  [in Mikrogramm] Benzpyren am Applikationsort noch vorhanden (d.h. noch nicht abgebaut) war (vgl. Kinder-Osius-Timm 1982, Beispiel 5.12). Die Abhängigkeit der noch vorhandenen Menge  $y$  von der Zeit  $x$  soll durch einen exponentiellen Zusammenhang modelliert werden (exponentieller Abbau). Aus den beobachteten Werten

Zeit $x$ [Tage]	0	7	28	90	180
Benzpyren $y$ [mg]	65,0	54,0	34,0	6,3	3,5

ergibt sich als gewichtete Regressionsgerade für die *transformierte* Zielgröße  $v = \ln y$  und die Gewichtsfaktoren  $w = y^2$

$$v = \hat{a} + \hat{b}x \quad \text{mit} \quad \hat{a} = 4,157 \quad \text{und} \quad \hat{b} = -0,02181.$$

Die *geschätzte Funktion* (auf der *Original-Skala*)

$$y = \hat{f}(x) = \hat{A}e^{\hat{b}x} \quad \text{mit} \quad \hat{A} = e^{\hat{a}} = 63,91$$

erklärt  $B = 99,5\%$  (Bestimmtheitsmaß der Regressionsfunktion) der Streuung der  $y$ -Werte.

Führt man die Regression auf der transformierten Skala *ohne* Gewichtung durch, so weichen die prognostizierten  $y$ -Werte stärker von den beobachteten Punkten ab als bei Verwendung der Gewichtung (vgl. Abb 3).

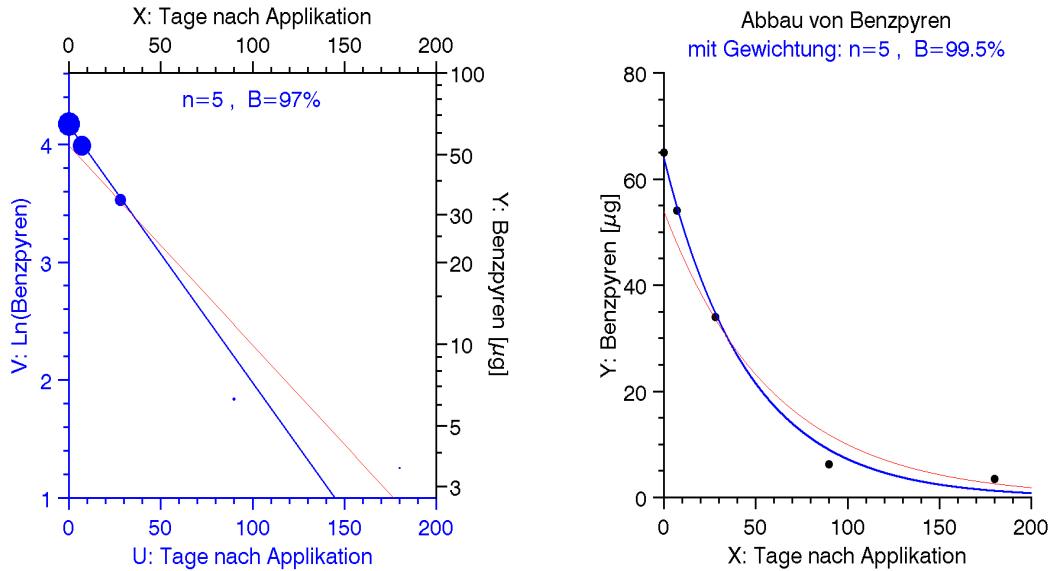


Abb. 3 links: transformierte gewichtete Punkte (Größe entspricht Gewichtung) mit *gewichteter* (fett) und *ungewichteter* (dünn) Regressionsgeraden.  
rechts: beobachtete Punkte auf der *Original-Skala* mit *gewichteter* (fett) und *ungewichtet* (dünn) geschätzter Regressionsfunktion.

### 1.6.5 Potenz-Zusammenhang

Der *Potenz-Zusammenhang* zwischen einer *positiven* Zielgröße  $y > 0$  und einer positiven Einflußgröße  $x > 0$  ist durch eine Potenz-Funktion gegeben,

$$y = f(x) = Ax^b \quad (\text{Potenz-Zusammenhang}),$$

wobei der Faktor  $A > 0$  *positiv* ist (vgl. auch das folgende Anwendungs-Beispiel). Ein solcher Zusammenhang ist dadurch charakterisiert, daß eine *multiplikative* Änderung der Einflußgröße  $x$  auch eine *multiplikative* Änderung der Zielgröße  $y$  bewirkt

$$f(q \cdot x) = q^b \cdot f(x).$$

Wir betrachten nun die Logarithmen *beider* Variablen

$$u = \log_c(x), \quad v = \log_c(y),$$

wobei man typischerweise den natürlichen oder dekadischen Logarithmus wählt. Dann hängt die *transformierte* Zielvariable  $v$  *linear* von der *transformierten* Einflußgröße  $u$  ab,

$$v = a + b u \quad (\text{linearer Zusammenhang von } v \text{ und } u),$$

wobei die Parameter  $A$  und  $a$  sich auseinander ergeben:

$$a = \log_c(A) \quad \text{bzw.} \quad A = c^a.$$

Für die beobachteten Wertepaare  $(x_1, y_1), \dots, (x_n, y_n)$  ist es (wie beim exponentiellen Zusammenhang) nun wieder naheliegend, die Parameter  $a$  und  $b$  nach der Methode der kleinsten Quadrate als Koeffizienten der *Regressionsgeraden*  $v = \hat{a} + \hat{b}u$  für die *transformierten* Punkte  $(u_i, v_i)$  mit  $u_i = \log c(x_i)$  und  $v_i = \log c(y_i)$  zu schätzen. Hierbei wird allerdings die Summe der quadratischen Abweichungen auf der *logarithmisch transformierten* Skala der Zielvariablen minimiert. Die Bestimmung der optimalen Parameterwerte nach der Methode der kleinsten Quadrate auf der *Originalskala* (vgl. 1.6.1) ist wieder aufwendiger. Aber man kann eine *approximative Lösung* dieser optimalen Werte erhalten, indem man eine *gewichtete Regressionsgerade*  $v = \hat{a} + \hat{b}u$  bestimmt mit den Gewichtsfaktoren

$$w_i = (y_i \cdot \ln c)^2.$$

Falls die ursprünglichen Werte  $(x_i, y_i)$  bereits mit einer Gewichtung  $w_i^\circ$  vorliegen, so verwendet man das Produkt  $w_i = w_i^\circ \cdot (y_i \cdot \ln c)^2$  als neues Gewicht der transformierten Werte  $(u_i, v_i)$ .

Aus der so gewichteten Regressionsgeraden  $v = \hat{a} + \hat{b}u$  ergibt sich die *Regressionsfunktion (geschätzte Funktion)* für den Potenz-Zusammenhang

$$y = \hat{f}(x) = \hat{A}x^{\hat{b}} \quad \text{mit} \quad \hat{A} = c^{\hat{a}}.$$

Das Bestimmtheitsmaß  $B$  für die Regressionsfunktion (nach 1.6.1) stimmt hier im allgemeinen *nicht* mit dem Bestimmtheitsmaß  $B_{vv}$  der gewichteten Regressionsgeraden  $v = \hat{a} + \hat{b}u$  (nach 1.5.6) überein, weil die Zielvariable  $y$  transformiert wurde.

### **Länge und Gewicht einer Schlangenart**

Bei 12 Exemplaren einer Schlangenart (*Heterodon nasicus*) wurden die Länge  $x$  [in cm] und das Gewicht  $y$  [in Gramm] bestimmt. Die Abhängigkeit des Gewichts  $y$  von der Länge  $x$  lässt sich durch einen Potenz-Zusammenhang beschreiben (vgl. Batschelet 1980, Example 7.3.1). Aus den beobachteten Werten

Länge $x$ [cm]	25	30	35	40	45	50	55	60	65	70	75	80
Gewicht $y$ [g]	5	15	20	30	40	55	75	95	120	150	185	225

erhält man als Regressionsgerade für die *Transformationen*  $u = \log_{10} x, v = \log_{10} y$ :

$$v = \hat{a} + \hat{b}u \quad \text{mit} \quad \hat{a} = -3,281 \quad \text{und} \quad \hat{b} = 2,959.$$

Die geschätzte Funktion (auf der *Original-Skala*)

$$y = \hat{f}(x) = \hat{A}x^{\hat{b}} \quad \text{mit} \quad \hat{A} = 10^{\hat{a}} = 0,0005233$$

erklärt  $B = 99,97\%$  (Bestimmtheitsmaß der Regressionsfunktion) der Streuung der  $y$ -Werte. Man beachte, daß das Bestimmtheitsmaß  $B_{uv} = 99,92\%$  der Regressionsgeraden für die transformierten Punkte *nicht exakt* mit  $B$  übereinstimmt.

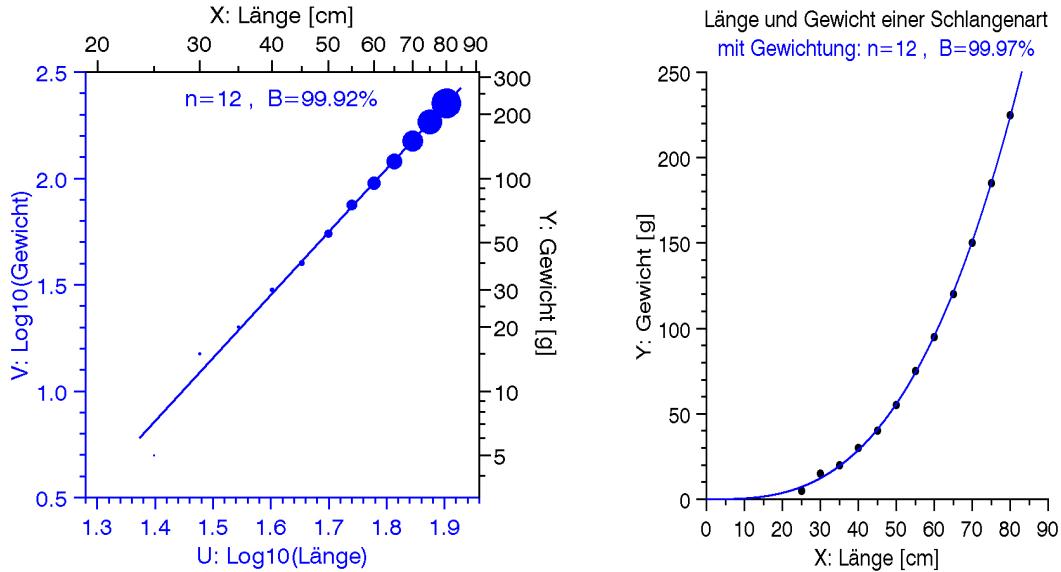


Abb. 4 links: transformierte gewichtete Punkte (Größe entspricht Gewichtung) und gewichtete Regressionsgerade.  
rechts: beobachtete Punkte auf *Original-Skala* mit geschätzter Regressionsfunktion.

### 1.6.6 Linearisierende Transformationen

In den letzten Abschnitten haben wir gesehen, daß sich der logarithmische, exponentielle und Potenz-Zusammenhang zwischen den Variablen  $x$  und  $y$  durch logarithmische Transformation einer oder beider Variablen auf den linearen Zusammenhang zurückführen lassen. Bevor wir weitere konkrete Transformationen (neben dem Logarithmus) behandeln, wollen wir das Prinzip der linearisierenden Transformationen allgemein erläutern.

Gegeben ist ein funktionaler Zusammenhang  $y = f(x)$ . Gesucht sind *Skalen-Transformationen* beider Variablen

$$u = u(x), \quad v = v(y),$$

so daß sich der Zusammenhang  $y = f(x)$  äquivalent als *lineare* Funktion in den *transformierten* Variablen  $u$  und  $v$  darstellen läßt:

$$y = f(x) \Leftrightarrow v = a + b u.$$

Solche Transformationen werden dann linearisierende Transformationen für die Funktion  $y = f(x)$  genannt. Der Fall, daß man nur *eine* der beiden Variablen  $x$  bzw.  $y$  transformiert (wie beim logarithmischen und exponentiellen Zusammenhang), ist in der allgemeinen Formulierung enthalten, indem man für die nicht zu transformierende Variable formal die *identische Transformation*  $u(x) = x$  bzw.  $v(y) = y$  wählt.

Leider gibt es *nicht für jede* Funktion  $y = f(x)$  geeignete linearisierenden Transformationen (z.B. nicht für den quadratischen Zusammenhang, vgl. 1.6.9). Wir wollen jetzt aber nur solche Funktionen mit linearisierenden Transformationen betrachten. Dann ist die Funktion  $y = f(x)$  auch durch die beiden Parameter  $a$  und  $b$  bereits bestimmt sind.

Für beobachtete Wertepaare  $(x_1, y_1), \dots, (x_n, y_n)$  mit *Gewichtsfaktoren*  $w_1^\circ, \dots, w_n^\circ$  (die im *ungewichteten* Fall alle  $= 1$  sind) kann man einerseits die Parameter  $a$  und  $b$  nach der Methode der kleinsten Quadrate als Koeffizienten der *Regressionsgeraden*  $v = \hat{a} + \hat{b} u$  für die *transformierten* Punkte  $(u_i, v_i)$  mit  $u_i = u(x_i)$  und  $v_i = v(y_i)$  schätzen. Hierbei wird die gewichtete Summe der quadratischen Abweichungen auf der *transformierten* Skala der Zielvariablen minimiert,

$$\sum_i w_i \cdot (v_i - (a + bu_i))^2 \quad \longrightarrow \quad \text{Minimum}, \quad (1)$$

wobei es zunächst plausibel erscheint, die gleichen Gewichte wie bei den Originalwerten zu verwenden, d.h.  $w_i = w_i^\circ$ . Dies entspricht aber *nicht exakt* dem Prinzip der gewichteten minimalen Quadrate auf der *Originalskala* der Zielvariablen:

$$\sum_i w_i^\circ \cdot (y_i - f(x_i))^2 \quad \longrightarrow \quad \text{Minimum}. \quad (2)$$

Die Minimierung der Quadratsumme auf der Originalskala wäre zwar vorzuziehen, ist aber auch mathematisch aufwendiger. Deshalb begnügen wir uns hier mit einer *approximativen Lösung* der Minimierung auf der Originalskala, indem wir einen vom ursprünglichen Gewicht  $w_i^\circ$  abweichenden *Gewichtsfaktor*  $w_i$  für die *gewichtete Regressionsgerade* verwenden. Dieser neue *Gewichtsfaktor*  $w_i$  ergibt sich aus folgender Approximation des Vorhersagefehlers für das *transformierte* Paar:  $(u_i, v_i)$

$$v_i - (a + bu_i) = v(y_i) - v(f(x_i)) \approx v'(y_i) \cdot (y_i - f(x_i))$$

mit  $v'(y_i)$  als *Ableitung* der Skalentransformation  $v$  an der Stelle  $y_i$ . Wählt man

$$w_i = w_i^\circ \left( \frac{1}{v'(y_i)} \right)^2 \quad (3)$$

als neuen Gewichtsfaktor der transformierten Werte  $(u_i, v_i)$ , so ist die Minimierung von (1) für die *transformierten Werte* auch eine *approximative* Minimierung von (2) für die Originalwerte.

Die Gewichtsfaktoren beim exponentiellen und Potenz-Zusammenhang wurden auf diese Weise bestimmt (wobei dort die Originalwerte *ungewichtet* waren, d.h.  $w_i^\circ = 1$ ). Wenn *keine*  $y$ -Transformation vorgesehen ist, so verwendet man formal die *identische*  $y$ -Transformation  $v(y) = y$  mit  $v'(y) = 1$  und die neuen Gewichte stimmen mit den bisherigen überein:  $w_i = w_i^\circ$  (wie z.B. beim logarithmischen Zusammenhang).

Durch Einsetzen der optimalen Werte  $\hat{a}$  und  $\hat{b}$  ergibt sich dann die *Regressionsfunktion (geschätzte Funktion)*  $y = \hat{f}(x)$  für den Zusammenhang auf der Originalskala. Das Bestimmtheitsmaß  $B$  für die Regressionsfunktion (nach 1.6.1) beschreibt die Streuungsreduktion auf der *Original-Skala*, d.h. für die  $y$ -Werte. Es stimmt im allgemeinen *nicht* mit dem Bestimmtheitsmaß  $B_{uv}$  für die gewichtete Regressionsgerade  $v = \hat{a} + \hat{b}x$  (nach 1.5.6) überein, welches ja die Streuungsreduktion auf der *transformierten Skala* beschreibt, also für die  $v$ -Werte. Lediglich bei *identischer y-Transformation*  $v(y) = y$  (und *beliebiger x-Transformation*) sind beide Bestimmtheitsmaße gleich:  $B = B_{uv}$  (also z.B. beim logarithmischen Zusammenhang).

Um einen optischen Eindruck zu erhalten wird man sowohl die transformierten Punkte  $(u_i, v_i)$  mit der Regressionsgeraden auf den transformierten Skalen als auch die Punkte  $(x_i, y_i)$  mit der Regressionsfunktion auf den Originalskalen darstellen.

### 1.6.7 Logit-Transformation und logistischer Zusammenhang

Wir wollen jetzt die spezielle Situation betrachten, in der die Zielvariable  $y$  eine *relative Häufigkeit* (oder *Wahrscheinlichkeit*, vgl. 2.1) bzw. *Prozentzahl* ist, deren Werte zwischen 0 und 1 bzw. zwischen 0% und 100% liegen. Typischerweise ist  $y$  eine *Ereignis-Rate* (für das Eintreten eines interessierenden Ereignisses) und wird auch als *Risiko* bezeichnet, wenn es sich um ein unerwünschtes Ereignis (z.B. Krankheit, Tod) handelt. Die Abhängigkeit von  $y$  von einer Einflußvariablen  $x$  (deren Wertebereich keinen Einschränkungen unterworfen ist) soll wieder durch eine Funktion  $y = f(x)$  beschrieben werden. Hierfür sind *lineare* Funktionen  $f(x) = a + bx$  nur bedingt geeignet, weil die Funktionswerte  $a + bx$  nur dann zwischen 0 und 1 liegen, wenn  $x$  zwischen  $-a/b$  und  $(1 - a)/b$  liegt. Um die Probleme zu umgehen, die aus der Einschränkung des Wertebereichs von  $y$  (auf das Intervall von 0 bis 1) entstehen, verwendet man eine Transformation  $v$  der Zielvariablen, deren Werte keinerlei Einschränkungen unterliegen, d.h.  $v$  kann jeden Wert (auf der Zahlengeraden) annehmen. Von den zahlreichen möglichen Transformationen dieser Art wird aus verschiedenen theoretischen Gründen (auf die hier nicht weiter eingegangen werden kann) die sogenannte *Logit-Transformation* bevorzugt verwendet:

$$v = \text{logit}(y) = \ln\left(\frac{y}{1-y}\right) \quad \text{für } 0 < y < 1 \quad (\text{Logit-Transformation}). \quad (1)$$

Wenn  $y$  die Eintritts-Rate (Wahrscheinlichkeit) für ein Ereignis ist, so ist  $(1 - y)$  die Rate (Wahrscheinlichkeit) für das *Nicht-Eintreten* des Ereignisses, und das Verhältnis  $y : (1 - y)$  wird auch als *Chance* für das Ereignis bezeichnet. Folglich ist  $\text{logit}(y)$  gerade der natürliche Logarithmus der Chance, kurz: die *log-Chance*.

Die Logit-Transformation ist im Punkt  $(1/2, 0)$  drehsymmetrisch (vgl. Abb. 5)

$$\text{logit}(1 - y) = - \text{logit}(y), \quad \text{logit}\left(\frac{1}{2}\right) = 0.$$

Für *sehr kleine* Werte von  $y$  stimmt der  $\text{logit}(y)$  näherungsweise mit dem Logarithmus  $\ln(y)$  überein, weil dann der Nenner  $1 - y$  ungefähr gleich 1 ist.

Analog zur log-Skala wird auch eine sogenannte *Logit-Skala* (vgl. Abb. 5, 6) verwendet, wenn man das Intervall von 0 bis 1 in den Randbereichen „progressiv vergrößern“ will, um großenordnungsmäßig variierende  $y$ -Werte übersichtlicher darstellen zu können.

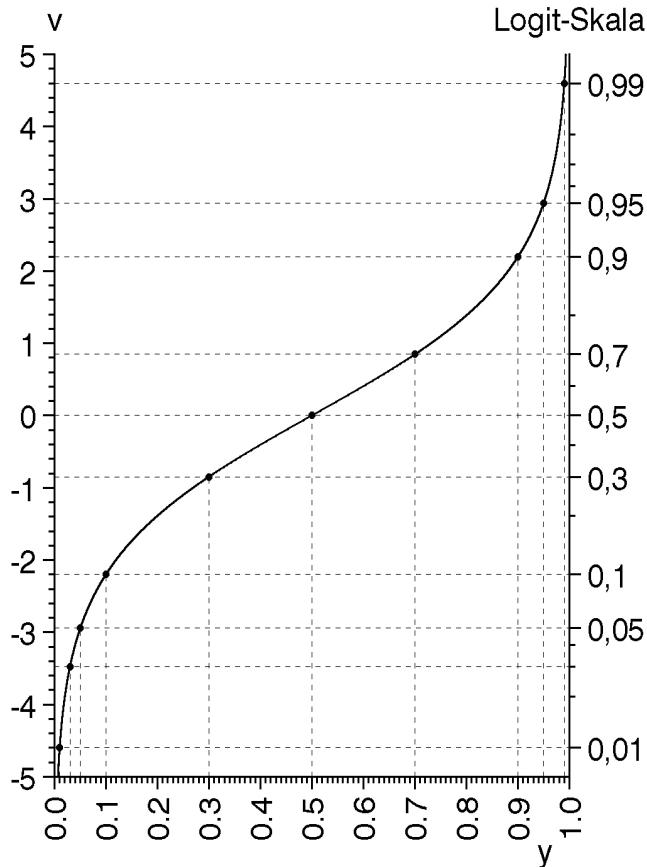


Abb. 5: Die Logit-Funktion und Konstruktion einer *Logit-Skala* (rechte Vertikale durch „orthogonale Reflektion“ der unteren Achse an der Funktion)

Die Umkehrfunktion (inverse Transformation) der Logit-Transformation ist

$$\begin{aligned} y = \Lambda(v) &= \frac{1}{1 + e^{-v}} && (\text{Inverse der Logit-Transformation}), \\ \text{d.h. } y = \Lambda(v) &\Leftrightarrow v = \text{logit}(y). \end{aligned} \quad (2)$$

Nach diesen Vorüberlegungen zur Logit-Transformation wollen wir den logistischen Zusammenhang zwischen  $y$  und  $x$  beschreiben, wobei wir gleichzeitig auch eine Transformation  $u = u(x)$  der Einflußgröße  $x$  zulassen. Typischerweise ist  $u = x$  (d.h. *keine* bzw. die *identische* Transformation) oder  $u = \log(x)$ . Der logistische Zusammenhang zwischen  $y$  und  $u$  ist nun gegeben durch einen linearen Zusammenhang der transformierten Variablen  $v$  und  $u$ :

$$v = \text{logit}(y) = a + bu \quad \text{bzw.} \quad (3)$$

$$y = \Lambda(a + bu) = \frac{1}{1 + e^{-(a+bu)}} \quad (\text{logistischer Zusammenhang}) \quad (4)$$

Für beobachtete Wertepaare  $(x_1, y_1), \dots, (x_n, y_n)$  lassen sich  $a$  und  $b$  wieder schätzen als die Koeffizienten einer *gewichteten Regressionsgeraden*  $v = \hat{a} + \hat{b}u$  für die *transformierten* Punkte  $(u_i, v_i)$  mit  $u_i = u(x_i)$  und  $v_i = \text{logit}(y_i)$ . Wir gehen hierbei davon aus, daß  $y_i$  die *relative Häufigkeit* eines interessierenden Ereignisses bei vorgegebenem Wert  $x_i$  ist, d.h. es ist  $y_i = r_i/m_i$ , wobei  $r_i$  die Anzahl der Eintritte des Ereignisses unter den insgesamt  $m_i$  Untersuchungsobjekten mit gegebenem  $x$ -Wert  $x_i$  ist. Zum Ausgleich der unterschiedlichen Streuung relativer Häufigkeiten verwendet man für  $y_i$  den *Gewichtsfaktor*

$$w_i^\circ = \frac{m_i}{y_i(1 - y_i)}, \quad \text{Gewichtsfaktor für } y_i = \frac{r_i}{m_i} \quad (5)$$

(also den Kehrwert der *geschätzten Varianz* von  $y_i$  nach **3.1.1**). Aus der Ableitung

$$v'(y) = \text{logit}'(y) = \frac{1}{y(1 - y)}$$

ergeben sich mit **1.6.6(3)** die Gewichtsfaktoren für  $v_i = \text{logit}(y_i)$  zu

$$w_i = m_i \cdot y_i(1 - y_i), \quad \text{Gewichtsfaktor für } v_i = \text{logit}(y_i). \quad (6)$$

Man beachte, daß sich die Werte  $y = 0$  und  $y = 1$  nicht transformieren lassen (ihre Logits wären  $-\infty$  und  $+\infty$ ). In der Praxis kann man Beobachtungen mit  $y_i = 0$  oder  $y_i = 1$  *ignorieren* (da ihr Gewichtsfaktor  $w_i = 0$  ist), aber nur wenn dadurch der Gesamtumfang  $\sum m$  nur unwesentlich reduziert wird (z.B. *höchstens* um 10%). Wäre eine solche Reduktion höher oder wenn *keine* Beobachtungen ignoriert werden sollen, so muß man aufwendigere Verfahren einsetzen, auf die hier nicht eingegangen wird (Stichwort: *Maximum-Likelihood Schätzung* bei *logistischer Regression*).

### Toxizität von Ouabain bei Fröschen

Ein Anwendungsgebiet des logistischen Zusammenhangs ist die *Dosis-Wirkungs-Analyse*. Wir betrachten ein historisches Beispiel aus Chapman and Morrell (Quart. J. Pharm. 4, 1931, p. 195) zur Toxizität eines Digitalis-Präparats (Ouabain) im Tierversuch. Hierbei wurden für verschiedene Dosierungen  $x$  [in Mikrogramm pro 100 g Körpergewicht] jeweils  $m = 30$  Frösche mit der Dosis  $x$  geimpft, und dann die Anzahl  $r$  der daran gestorbenen Tiere gezählt. Es ergaben sich folgende Werte

$x$	25	27	29	31	33	35	37
$r$	1	2	7	18	25	28	29

Die Abhängigkeit der Toxizitäts-Rate  $y = r/m$  von der Dosis  $x$  soll durch ein logistisches Modell  $v = a + bu$  beschrieben werden mit  $u = \log_{10}(x)$  als logarithmische Dosis und  $v = \text{logit}(y)$  als Logit der Toxizitäts-Rate  $y$ , d.h.

$$y = f(x) = \Lambda(a + b \log_{10}(x)) = \frac{1}{1 + e^{-(a + b \log_{10}(x))}}.$$

Toxizität des Digitalis-Präparats 'Ouabain' bei Fröschen

Lfd. Nr	Ouabain-Dosis [mg/100g Froschgew.]	Anzahl geimpft	davon gestorben	Toxizitäts- Rate $y = r/m$	$u =$ $\log_{10}(x)$	$v =$ $\text{logit}(y)$	$w =$ $my(1-y)$	Gewicht	$wu$	$wv$	$wuu$	$wvv$	$wuv$
1	25	30	1	3,33 %	1,398	-3,367	0,967	1,351	-3,255	1,889	10,961	-4,550	
2	27	30	2	6,67 %	1,431	-2,639	1,867	2,672	-4,926	3,824	13,001	-7,051	
3	29	30	7	23,33 %	1,462	-1,190	5,367	7,848	-6,384	11,477	7,594	-9,336	
4	31	30	18	60,00 %	1,491	0,405	7,200	10,738	2,919	16,014	1,184	4,354	
5	33	30	25	83,33 %	1,519	1,609	4,167	6,327	6,706	9,608	10,793	10,183	
6	35	30	28	93,33 %	1,544	2,639	1,867	2,882	4,926	4,450	13,001	7,606	
7	37	30	29	96,67 %	1,568	3,367	0,967	1,516	3,255	2,377	10,961	5,105	
Min	25			3,33 %	1,398	-3,367	22,400	33,335	3,241	49,640	67,494	6,310	
Max	37			96,67 %	1,568	3,367						Summen	

Tabelle 1: Berechnung der Summen  $\sum w, \dots, \sum wuv$  für die gewichtete Regressionsgerade der  $(u, v)$ -Werte mit Gewichtsfaktor  $w$  (vgl. Rechen-Box 3).

Die gewichtete Regressionsgerade ergibt sich nach Rechen-Box 3 (mit  $u$  und  $v$  statt der dortigen  $x$  und  $y$ ) aus den Werten in Tab. 1 zu (vgl. Abb 6)

$$v = \hat{a} + \hat{b}u \quad \text{mit} \quad \hat{a} = -66,12 \quad \text{und} \quad \hat{b} = 44,53.$$

Dies liefert den Zusammenhang auf der *transformierten Skala* und die *geschätzte* Funktion (vgl. Abb. 6) auf der *Original-Skala* lautet

$$y = \hat{f}(x) = \Lambda(\hat{a} + \hat{b} \log_{10}(x)) = \frac{1}{1 + e^{-(\hat{a} + \hat{b} \log_{10}(x))}}.$$

Hierdurch wird  $B = 99,88\%$  (Bestimmtheitsmaß der Regressionsfunktion) der Streuung von  $y$  erklärt. Nebenbei bemerkt erklärt die Regressionsgerade  $B_{uv} = 98,77\%$  der Streuung der *transformierten* Werte  $v = \text{logit}(y)$  und insbesondere stimmt  $B_{uv}$  nicht mit  $B$  überein.

Obwohl man die geschätzte Funktion primär für Prognosen von  $y$  bei vorgebenem  $x$  verwendet, ist auch die umgekehrte Fragestellung von Interesse: Welcher  $x$ -Wert  $x_0$  gehört zu einem vorgegebenen  $y$ -Wert  $y_0$  (z.B. 50%)? Da die transformierten Werte  $v_0 = \text{logit}(y_0)$  und  $u_0 = \log_{10}(x_0)$  auf der Regressionsgeraden liegen, ergibt sich

$$u_0 = (v_0 - \hat{a})/\hat{b} \quad \text{und} \quad x_0 = 10^{u_0}.$$

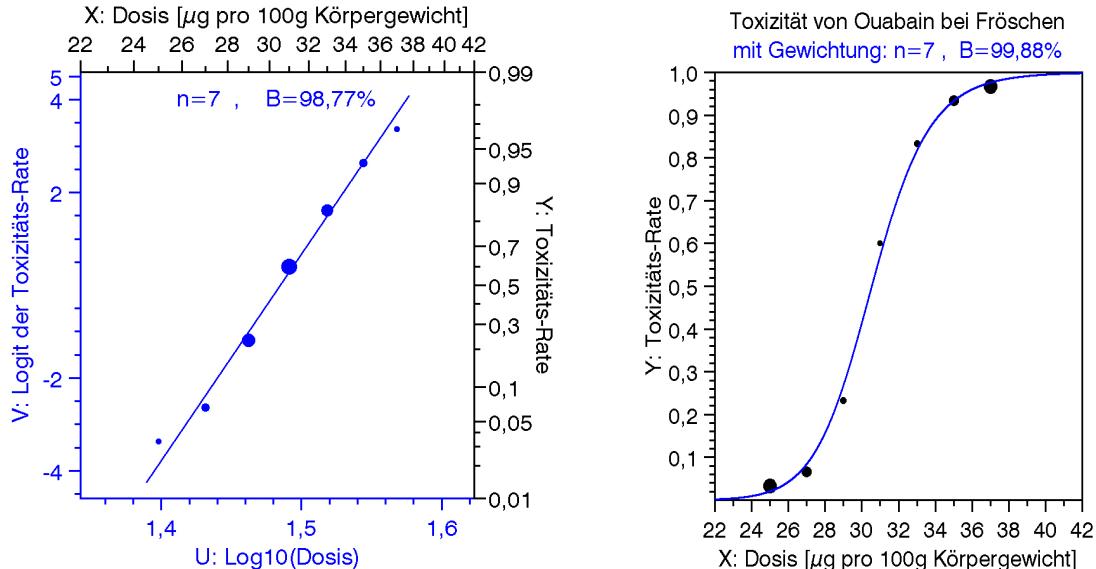


Abb. 6 links: transformierte gewichtete Punkte und *gewichtete* Regressionsgerade  
Die Größe entspricht der Gewichtung aus (6).  
rechts: Punkte mit der Gewichtung (5) (für relative Häufigkeiten)  
auf der *Original-Skala* und geschätzte Regressionsfunktion.

### 1.6.8 Periodischer Zusammenhang

Wir betrachten jetzt eine Zielvariable  $y$ , die periodisch von der Einflußgröße  $x$  abhängt, d.h. während  $x$  eine Periode durchläuft, variiert  $y$  innerhalb eines Intervalls und mit jeder neuen Periode wiederholt sich die Variation von  $y$ . Die Einflußvariable  $x$  ist hierbei typischerweise ein Zeitpunkt und die Periode kann z.B. ein Tag oder ein Jahr sein. Den Ausgangspunkt für die Beschreibung periodischer Zusammenhänge bilden die Sinus-Funktion  $y = \sin(x)$  und die verwandte Cosinus-Funktion  $y = \cos(x)$ . Ursprünglich sind diese Funktionen für *Winkel*  $x$  im Bereich von  $0^\circ$  bis  $360^\circ$  definiert und diese Definition wird dann mit der Periode von  $360^\circ$  fortgesetzt (Abb. 7). Es ist hier jedoch zweckmäßiger, statt des *Winkelmaßes* (GRAD) als Einheit für  $x$  das *Bogenmaß* (RAD) zu verwenden, wobei die Periode dann  $2\pi$  [RAD]  $\hat{=} 360$  [GRAD] ist.

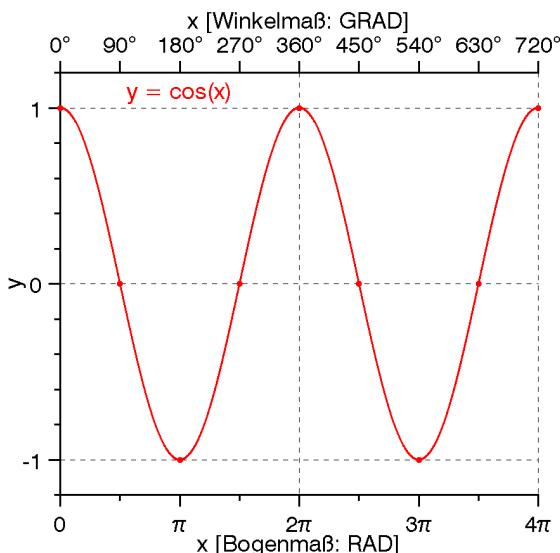
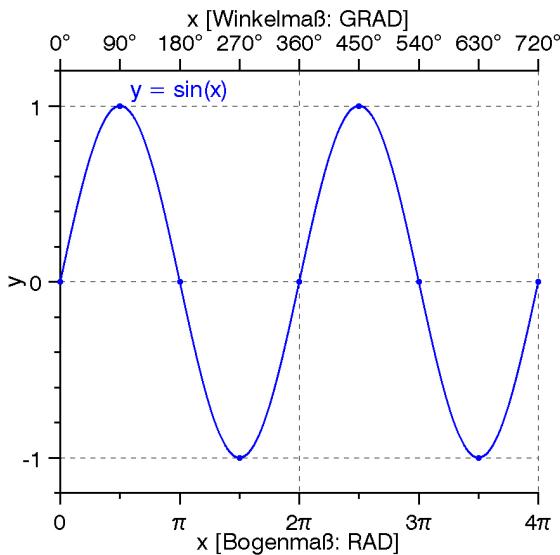
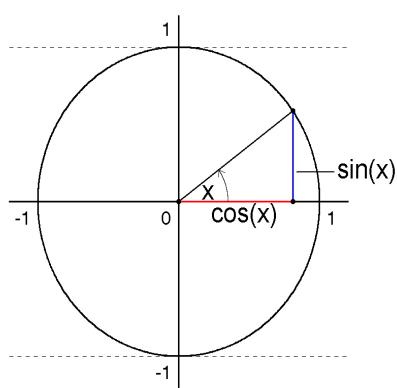


Abb. 7:

Geometrische Darstellung von  $\sin(x)$  und  $\cos(x)$  für den Winkel  $x$  am Einheitskreis (*links oben*). Die Länge des zu  $x$  gehörigen *Kreisbogens* ist das *Bogenmaß (RAD)* von  $x$ .

Durchläuft der Winkel  $x$  (gegen den Uhrzeigersinn) eine Periode von  $0^\circ$  bis  $360^\circ$  (bzw.  $0$  bis  $2\pi$  im Bogenmaß), so erhält man den ersten Teil der Abbildungen links (*oben* für *Sinus* und *unten* für *Cosinus*). Bei der nächsten Periode von  $360^\circ$  bis  $720^\circ$  (bzw.  $2\pi$  bis  $4\pi$  im Bogenmaß) wiederholen sich die Funktionsverläufe.

Wegen des Zusammenhangs von Sinus- und Cosinus-Funktion

$$\sin(x) = \cos(x - \frac{\pi}{2}), \quad \cos(x) = \sin(x + \frac{\pi}{2})$$

genügt es prinzipiell, *eine* der beiden Funktionen zu betrachten, aber wir werden die wichtigsten Aussagen jeweils mit *beiden* Funktionen formulieren.

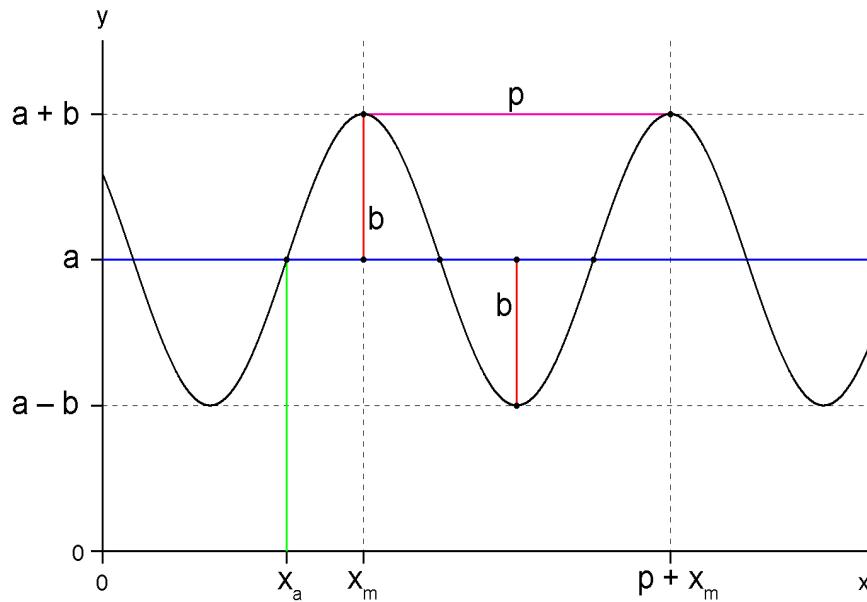
Unter Verwendung der Sinus- bzw. Cosinus-Funktion lassen sich eine Fülle periodische Funktionen konstruieren. Wir betrachten nur die einfache Form (vgl. Abb. 8):

Periodischer Zusammenhang:

$$y = f(x) = a + b \cdot \sin\left(\frac{2\pi}{p}[x - x_a]\right) \quad \text{bzw.}$$

$$y = f(x) = a + b \cdot \cos\left(\frac{2\pi}{p}[x - x_m]\right) \quad \text{mit} \quad x_m = x_a + \frac{1}{4}p.$$

Hierbei ist  $a$  das generelle *y-Niveau*,  $b$  die *Amplitude*,  $p$  die *Periode* und  $x_a$  bzw.  $x_m$  der *Anfangs-* bzw. *Maximalwert* einer sogenannten *Phase* (mit Länge  $p$ ).



**Abb. 8:** Darstellung der periodischen Funktion

$$y = f(x) = a + b \cdot \sin\left(\frac{2\pi}{p}[x - x_a]\right) = a + b \cdot \cos\left(\frac{2\pi}{p}[x - x_m]\right)$$

mit *y-Niveau*  $a$ , Amplitude  $b$ , Periode  $p$ , Anfangswert  $x_a$  und Maximalwert  $x_m$ .

Wenn die Periode  $p$  und der Anfangs- bzw. Maximalwert  $x_a$  bzw.  $x_m$  nicht bekannt sind, so lässt sich dieser Zusammenhang *nicht* durch geeignete Transformationen von  $x$  und  $y$  *linearisieren*. Wenn jedoch (wie im folgenden Anwendungsbeispiel) sowohl die Periode  $p$  als auch der Wert  $x_a$  bzw.  $x_m$  aus allgemeinen Vorüberlegungen heraus bekannt ist, so liefert die Transformation

$$u = \sin\left(\frac{2\pi}{p}[x - x_a]\right) = \cos\left(\frac{2\pi}{p}[x - x_m]\right)$$

einen linearen Zusammenhang zwischen  $y$  und  $u$

$$y = a + bu.$$

Für beobachtete Wertepaare  $(x_1, y_1), \dots, (x_n, y_n)$  – ggf. mit Gewichtsfaktoren – ergeben sich daher die *optimalen* Werte der Parameter  $a$  und  $b$  nach der Methode der kleinsten Quadrate als die Koeffizienten der (ggf. gewichteten) *Regressionsgeraden*  $y = \hat{a} + \hat{b}u$  für die *transformierten* Punkte  $(u_i, y_i)$ . Hieraus erhält man die *Regressionsfunktion (geschätzte Funktion)* für den periodischen Zusammenhang

$$\begin{aligned} y = \hat{f}(x) &= \hat{a} + \hat{b} \sin\left(\frac{2\pi}{p}[x - x_a]\right) \\ &= \hat{a} + \hat{b} \cos\left(\frac{2\pi}{p}[x - x_m]\right) \quad (\text{geschätzte Funktion}). \end{aligned}$$

Das Bestimmtheitsmaß  $B$  für die Regressionsfunktion (nach **1.6.1**) stimmt hier mit dem Bestimmtheitsmaß  $B_{uy}$  der Regressionsgeraden  $y = \hat{a} + \hat{b}u$  (nach **1.5.2** bzw. **1.5.6**) überein, weil nur die Einflußvariable  $x$ , aber *nicht* Zielvariable  $y$  transformiert wurde.

Man kann diese Linearisierung sogar dann noch anwenden, wenn nur die Periode  $p$  *genau* bekannt ist, während man den Wert  $x_a$  bzw.  $x_m$  nur *ungefähr* kennt bzw. vermutet. In diesem Fall wählt man eine Reihe plausibler Werte für  $x_a$  bzw.  $x_m$  und bestimmt für jeden dieser Werte die Regressionsgerade für die transformierten Punkte. Als beste Regressionsgerade wählt man dann die mit der größten Streuungsreduktion, d.h. mit größtem Bestimmtheitsmaß  $B_{uy}$ .

### **Kohlendioxyd-Konzentration in der Luft**

Eine dreijährige Meßreihe der  $CO_2$ -Konzentration  $y$  [in ppm] in der Luft (einer speziellen Region) in Abhängigkeit vom Kalendermonat  $x$  (1=Jan., 2=Feb. etc.) ergab folgende Werte (vgl. Kinder-Osius-Timm 1982, Beispiel 5.13):

Monat	1	2	3	4	5	6	7	8	9	10	11	12
$y(1.\text{Jahr})$	304,1	301,8	298,1	297,2	296,0	293,1	296,5	297,7	297,1	297,6	301,3	299,5
$y(2.\text{Jahr})$	303,5	300,4	298,1	293,5	296,1	295,3	294,4	293,5	296,9	300,2	298,9	303,3
$y(3.\text{Jahr})$	301,3	301,8	301,2	294,4	297,3	295,9	294,8	296,1	298,0	300,1	302,3	301,5

Da die Abgase von Heizungsanlagen erheblich zur  $CO_2$ -Konzentration beitragen, kann man von einer Jahresperiode ( $p = 12$ ) der  $y$ -Werte ausgehen, wobei der Maximalwert  $x_m$  in den Wintermonaten zu erwarten ist. Bei den beobachteten Daten liegen die höchsten  $y$ -Werte eines Jahres in den Monaten Januar (1) und Dezember (12). Daher ist es plausibel, den Maximalwert in der Mitte zwischen Dezember und Januar anzusetzen, was einem Wert von  $x_m = 0,5$  (oder im Hinblick auf die Periode:  $x_m = 12,5$ ) entspricht. Unter Verwendung der Transformation

$$u = \cos\left(\frac{2\pi}{12} [x - x_m]\right)$$

ergibt sich die Regressionsgerade  $v = \hat{a} + \hat{b}u$  auf der *transformierten Skala*. Die *geschätzte periodische Regressionsfunktion* auf der *Original-Skala*

$$y = \hat{f}(x) = \hat{a} + \hat{b} \cos\left(\frac{2\pi}{12} [x - x_m]\right)$$

erklärt  $B = 76,3\%$  der Streuung der  $y$ -Werte.

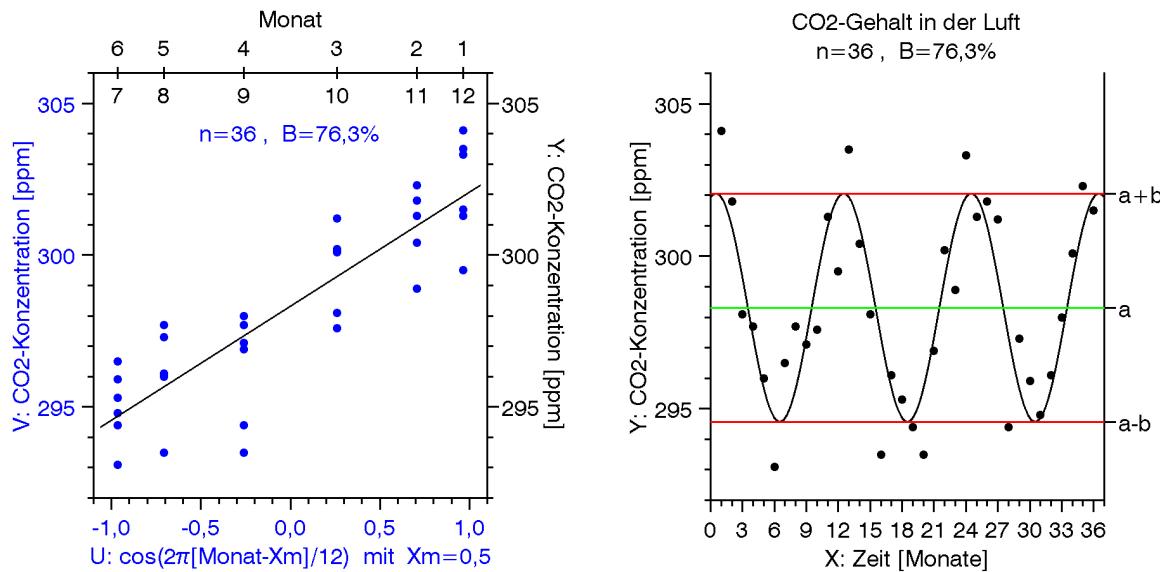


Abb. 9 links: Transformierte Punkte und Regressionsgerade (*linearisierte Skala*).

Für jeden Monat  $x$  (obere  $x$ -Achse) stimmen die transformierten Werte  $u$  (untere  $u$ -Achse) für alle drei Jahre überein.

rechts: Beobachtete Punkte mit geschätzter Funktion (*Original-Skala*).

Zur Überprüfung, ob der Maximalwert  $x_m = 0,5$  optimal gewählt ist, kann man die obige Analyse auch mit alternativen Werten durchführen. Erhöht man den Wert 0,5 jeweils in Schritten von 0,1 so verschlechtert sich die Anpassung (das Bestimmtheitsmaß  $B$  fällt). Verringert man dagegen den Wert, so ergibt sich für  $x_m = 0,4$  zunächst eine geringfügig bessere Anpassung ( $B = 76,5\%$ ), die aber bei einer weiterer Verringerung auf 0,3 etc. wieder schlechter wird.

Die sich für die optimale Wahl von  $x_m$  ergebende Regressionsgerade mit  $\hat{a} = 298,3$  und  $\hat{b} = 3,748$  und die periodische Regressionsfunktion ist in Abb. 9 mit den entsprechenden Punkten dargestellt. Da „nur“  $B = 76,5\%$  der Streuung der  $y$ -Werte erklärt wird, streuen die Punkte noch relativ stark um die geschätzte Funktion.

## ☒ Tipps für *Microsoft Excel*

Bei der Berechnung der gewichteten Regressionsgeraden und ihrer graphischen Darstellung (auf der transformierten Skala) ergeben sich gegenüber den Tipps in **1.5.7** keine neuen Aspekte, lediglich die Bezeichnungen haben sich geändert:  $(u, v)$  statt  $(x, y)$ . Zwei verschiedene  $x$ - bzw.  $y$ -Achsen jeweils mit *transformierter* und *Original-Skala* wie z.B. in Abb. 6 sind in *Excel* nicht vorgesehen. Eine Darstellung der geschätzten *Funktion* auf der *Original-Skala* ist auch etwas aufwendiger (vgl. unten), weil das Zeichnen von *Funktionen* in *Excel* nicht implementiert sind.

**Plotten von Funktionen:** Zur Darstellung einer Funktion  $y = f(x)$  erstellt man zuerst eine *Wertetabelle*, d.h. für eine Reihe von Werten  $x_1, x_2, \dots$  auf der  $x$ -Achse berechnet man (in *Excel*) die zugehörigen Funktionswerte  $f(x_1), f(x_2) \dots$  Dann werden diese Punkte der Funktion als *Datenreihe* in einem neuen (oder bereits vorhandenen) Diagramm dargestellt. Als *Diagrammtyp* wählt man hierbei unter *Punkt (XY)* den Untertyp „*Punkte mit interpolierten Linien ohne Datenpunkte*“, d.h. die Punkte selbst werden nicht sichtbar gezeichnet, sondern nur die sie verbindende *geglättete (interpolierte)* Linie. Wie gut die resultierende Kurve den wahren Funktionsverlauf wiederspiegelt, hängt von den gewählten  $x$ -Werten  $x_1, x_2, \dots$  ab: je *mehr* Werte es sind und je *dichter* sie nebeneinander liegen desto *genauer* und *glatter* wird die Funktion gezeichnet. Typischerweise wird man  $x_1$  als den Anfangswert der  $x$ -Achse wählen und diesen Wert dann schrittweise um eine feste Schrittweite  $\Delta x$  erhöhen, solange bis die gesamte  $x$ -Achse überstrichen ist.

## 1.7 Quadratischer Zusammenhang

Nachdem wir verschiedene Arten des Zusammenhangs unter Verwendung von Transformationen auf *lineare* Funktionen zurückgeführt haben, wollen wir die Klasse der *linearen* Funktionen erweitern, indem wir auch Funktionen mit *Krümmung* lassen. Hierzu betrachten wir *quadratische Funktionen (Polynome 2. Grades)*

$$y = f(x) = a + bx + cx^2 \quad (\text{quadratische Funktion})$$

mit den drei Parametern  $a, b$  und  $c$ . Hierbei gibt der *Krümmungsparameter*  $c$  Auskunft über die *Krümmung* der Funktion. Das Vorzeichen von  $c$  bestimmt die Art der Krümmung: für *positives*  $c > 0$  ist die Krümmung der Funktion *konvex* und für negatives  $c < 0$  ist sie *konkav* (vgl. Abb. 1). Und der Absolutbetrag  $|c|$  von  $c$  ist ein Maß für *Stärke* der Krümmung. Im Grenzfall  $c = 0$  liegt eine *lineare* Funktion (*ohne* Krümmung) vor, und somit umfassen die quadratischen Funktionen die bisher betrachteten linearen Funktionen. Da wir die linearen Funktionen bereits ausführlich behandelt haben, wollen wir im Folgenden davon ausgehen, daß  $c \neq 0$  ist und somit eine quadratische Funktion im *engeren* Sinn (d.h. keine lineare Funktion) vorliegt.

Im Gegensatz zu den linearen Funktionen haben die restlichen beiden Parameter  $a$  und  $b$  hier keine klare geometrische Interpretation. Geometrisch stellt die quadratische Funktion eine *Parabel* dar, die für *positives*  $c$  nach *oben* und für *negatives*  $c$  nach *unten* geöffnet ist. Unter Verwendung der Koordinaten  $(x_m, y_m)$  des *Scheitelpunktes* der Parabel lässt sich die Funktion auch beschreiben durch

$$y = f(x) = y_m + c(x - (x_m))^2 \quad (\text{Scheiteldarstellung}).$$

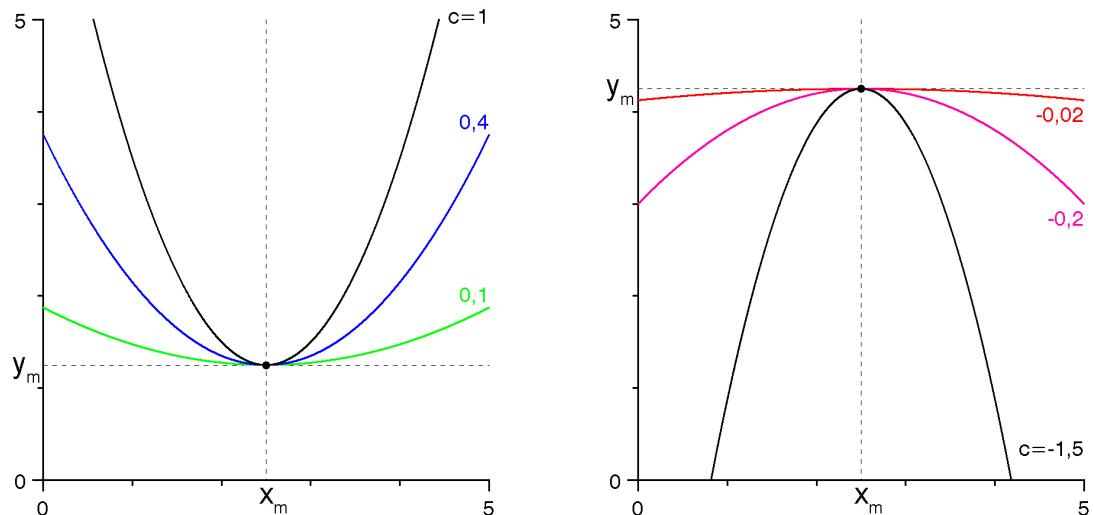


Abb. 1: Die *quadratische* Funktion  $y = y_m + c(x - x_m)^2$  für verschiedene Werte des Krümmungsparameters  $c$  bei jeweils gleichem Scheitelpunkt  $(x_m, y_m)$ .

links: *positive* Krümmung (*konvex*),

rechts: *negative* Krümmung (*konkav*).

Die Scheitelkoordinaten ergeben sich (durch quadratische Ergänzung oder eine Kurvendiskussion) aus  $a$  und  $b$  zu

$$x_m = -\frac{b}{2c}, \quad y_m = a - \frac{b^2}{4c} \quad (\text{Scheitelkoordinaten}). \quad (1)$$

Und umgekehrt erhält man  $a$  und  $b$  aus  $(x_m, y_m)$ ,

$$b = -2cx_m, \quad a = cx_m^2 + y_m.$$

Die Verwendung der Scheitelparameter ist für Interpretationen zweckmäßiger, weil sie direkt die Lage des Minimums bzw. Maximums der Kurve angeben. Für die folgende *Berechnung* der Regressionsfunktion bevorzugen wir jedoch die Parameter  $a, b$  und  $c$ , da sie auch den Grenzfall  $c = 0$  mit einschließen (bei dem die Scheitelkoordinaten nicht definiert sind).

Man beachte, daß sich der quadratische Zusammenhang in seiner allgemeinen Form mit *drei* Parametern nicht durch geeignete Transformationen auf ein linearen Zusammenhang zurückführen läßt, weil letztere durch nur *zwei* Parameter beschrieben werden.

### 1.7.1 Schätzung mit der Methode der kleinsten Quadrate

Für beobachtete Wertepaare  $(x_1, y_1), \dots, (x_n, y_n)$  wollen wir jetzt die optimalen Werte der Parameter  $a, b$  und  $c$  nach der Methode der kleinsten Quadrate bestimmen. Hierbei gehen wir gleich von der allgemeineren Situation mit *Gewichtsfaktoren*  $w_i$  für  $(x_i, y_i)$  aus, die ja den *ungewichteten* Fall (d.h. *konstante* Gewichtsfaktoren  $w_i = 1$ ) mit einschließt. Zur Bestimmung der optimalen Werte  $\hat{a}, \hat{b}, \hat{c}$  der unbekannten Parameter  $a, b, c$  nach der Methode der gewichteten kleinsten Quadrate werden wieder die bereits bekannten gewichteten Summen  $Swxx, Swxy$  (vgl. *Rechen-Box 3*) verwendet. Zusätzlich sind noch weitere Summen  $Swzz, Swxz, Swzy$  erforderlich, die analog unter Verwendung der *quadrierten*  $x$ -Werte  $z_i = x_i^2$  gebildet werden. Mit diesen Bezeichnungen erhält man die Schätzungen (vgl. auch *Rechen-Box 4*):

$$\begin{aligned}\hat{c} &= \frac{1}{D} (Swxx \cdot Swzy - Swxz \cdot Swxy) \quad \text{mit} \quad D = Swxx \cdot Swzz - (Swxz)^2, \\ \hat{b} &= \frac{1}{D} (Swzz \cdot Swxy - Swxz \cdot Swzy), \\ \hat{a} &= \bar{y} - \hat{b}\bar{x} - \hat{c}\bar{z}.\end{aligned}$$

Hierbei muß natürlich  $D \neq 0$  sein, was aber immer der Fall ist, wenn *mindestens drei verschiedene*  $x$ -Werte vorliegen. Und die zugehörigen Scheitelkoordinaten  $(\hat{x}_m, \hat{y}_m)$  ergeben sich dann durch Einsetzen von  $\hat{b}$  und  $\hat{c}$  in die Formeln (1) oben.

**Rechen-Box 4 für (gewichtete) quadratische Regression**(enthält den **ungewichteten** Fall für konstante Gewichtsfaktoren  $w_i = 1$ )

<i>Stichprobenumfang:</i>	$n$	
<i>x-Werte:</i>	$x_1, x_2, \dots, x_n$	
<i>y-Werte:</i>	$y_1, y_2, \dots, y_n$	
<i>z-Werte (quadrierte x-Werte):</i>	$z_1 = x_1^2, z_2 = x_2^2, \dots, z_n = x_n^2$	
<i>Gewichtsfaktoren:</i>	$w_1, w_2, \dots, w_n$	
<i>gewichtete Summen:</i>	$\sum w = \sum_i w_i$	
$\sum wx = \sum_i w_i x_i$	$\sum wx = \sum_i w_i y_i$	$\sum wz = \sum_i w_i z_i$
$\sum wx^2 = \sum_i w_i x_i^2$	$\sum wx^2 = \sum_i w_i y_i^2$	$\sum wz^2 = \sum_i w_i z_i^2$
$\sum wxy = \sum_i w_i x_i y_i$	$\sum wxz = \sum_i w_i x_i z_i$	$\sum wzy = \sum_i w_i z_i y_i$
<i>(gewichtete) Mittelwerte:</i>	$\bar{x} = \frac{1}{\sum w}(\sum wx), \bar{y} = \frac{1}{\sum w}(\sum wy), \bar{z} = \frac{1}{\sum w}(\sum wz)$	

*(gewichtete) Summen von Abweichungs-Quadraten und -Produkten:*

$$\begin{aligned}
 Swxx &= \sum_i w_i \cdot (x_i - \bar{x})^2 &= (\sum wx^2) - \frac{1}{\sum w}(\sum wx)^2 \\
 Swyy &= \sum_i w_i \cdot (y_i - \bar{y})^2 &= (\sum wy^2) - \frac{1}{\sum w}(\sum wy)^2 \\
 Swzz &= \sum_i w_i \cdot (z_i - \bar{z})^2 &= (\sum wz^2) - \frac{1}{\sum w}(\sum wz)^2 \\
 Swxy &= \sum_i w_i \cdot (x_i - \bar{x})(y_i - \bar{y}) &= (\sum wxy) - \frac{1}{\sum w}(\sum wx)(\sum wy) \\
 Swxz &= \sum_i w_i \cdot (x_i - \bar{x})(z_i - \bar{z}) &= (\sum wxz) - \frac{1}{\sum w}(\sum wx)(\sum wz) \\
 Swzy &= \sum_i w_i \cdot (z_i - \bar{z})(y_i - \bar{y}) &= (\sum wzy) - \frac{1}{\sum w}(\sum wz)(\sum wy)
 \end{aligned}$$

*(gewichtete) quadratische Regressionsfunktion  $y = \hat{a} + \hat{b}x + \hat{c}x^2$ :*

$$\begin{aligned}
 \hat{c} &= \frac{1}{D}(Swxx \cdot Swzy - Swxz \cdot Swxy), & D &= Swxx \cdot Swzz - (Swxz)^2 \\
 \hat{b} &= \frac{1}{D}(Swzz \cdot Swxy - Swxz \cdot Swzy), & \hat{a} &= \bar{y} - \hat{b}\bar{x} - \hat{c}\bar{z}
 \end{aligned}$$

*Streuung um die Regressionsfunktion (Parabel):*

$$\begin{aligned}
 WSQ_{quad} &= \sum_i w_i [y_i - (\hat{a} + \hat{b}x_i + \hat{c}x_i^2)]^2 \\
 &= Swyy - 2\hat{b}Swxy - 2\hat{c}Swzy + \hat{b}^2Swxx + \hat{c}^2Swzz + 2\hat{b}\hat{c}Swxz
 \end{aligned}$$

$$\text{Bestimmtheitsmaß der Parabel: } B_{quad} = 1 - \frac{WSQ_{quad}}{Swyy}$$

Obwohl in der *Rechen-Box 4* die beiden Summen  $\sum w z$  und  $\sum w x^2$  wegen  $z = x^2$  stets übereinstimmen, haben wir aus systematischen Gründen *beide* Summen aufgeführt.

Im *ungewichteten* Fall (d.h. alle  $w_i = 1$ ) ist die Summe  $\sum w$  der Gewichte der Stichprobenumfang  $n$ , und alle *gewichteten* Summen reduzieren sich zu *ungewichteten* Summen, die man vereinfachend wieder mit  $\sum x, S_{xx}, \dots$  statt  $\sum wx, Sw_{xx} \dots$  bezeichnet. Ein Beispiel hierfür (Blickwinkel und Reaktionszeit bei Rhesusaffen) folgt anschließend.

Das Bestimmtheitsmaß  $B$  der geschätzten Regressionsfunktion

$$\hat{f}(x) = \hat{a} + \hat{b} x + \hat{c}^2$$

ergibt sich dann wie in **1.6.1** allgemein beschrieben. Die hierfür erforderliche Streuung der  $y$ -Werte um die geschätzte quadratische Funktion

$$WSQ_{quad} = \sum_i w_i [y_i - (\hat{a} + \hat{b} x_i + \hat{c} x_i^2)]^2$$

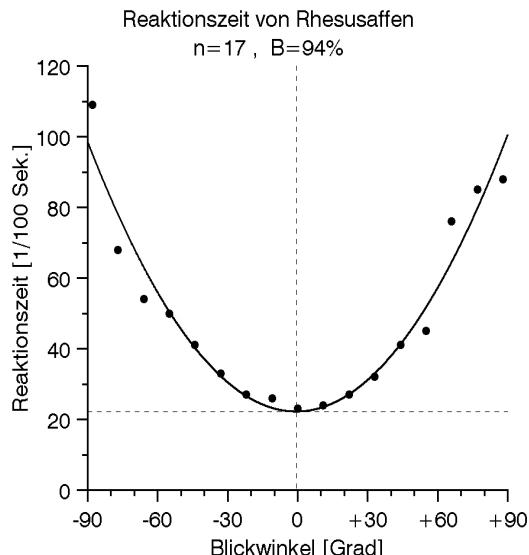
lässt sich auch unter Verwendung der bereits berechneten Summen bestimmen (vgl. *Rechen-Box 4*).

### **Blickwinkel und Reaktionszeit bei Rhesusaffen**

Als Wirkung eines neurotoxischen Präparats sollen die Veränderungen der (blickwinkelspezifischen) Reaktionszeiten eines Rhesusaffen auf optische Reize untersucht werden. Die Aufnahme der Reaktionszeiten des unbehandelten Tieres ergab die Werte der folgenden Tabelle (vgl. Kinder-Osius-Timm 1982, Beispiel 5.11).

Blickwinkel x [Grad]	Reaktionszeit y [1/100 Sek.]
+88	88
+77	85
+66	76
+55	45
+44	41
+33	32
+22	27
+11	24
0	23
-11	26
-22	27
-33	33
-44	41
-55	50
-66	54
-77	68
-88	109

Abb. 2: Beobachtete Punkte und geschätzte Funktion  
 $\hat{f}(x) = \hat{a} + \hat{b} x + \hat{c}^2$



Die Daten sollen durch eine *quadratische* Funktion angenähert werden. Da hier *keine* Gewichtung vorzusehen ist, werden alle Gewichtsfaktoren *konstant*  $w_i = 1$  gesetzt und man berechnet die *ungewichteten* Summen laut *Rechen-Box 4*:

$$\begin{aligned}\sum x &= 0; & \sum y &= 849; & \sum z &= 49\,368; \\ \sum x^2 &= 49\,368; & \sum y^2 &= 53\,405; & \sum z^2 &= 256\,861\,704; \\ \sum xy &= 583; & \sum xz &= 0; & \sum zy &= 3\,548\,083.\end{aligned}$$

Mit  $\sum w = n = 17$  ergeben sich die (ungewichteten) Summen

$$\begin{aligned}S_{xx} &= 49\,368; & S_{yy} &= 11\,005; & S_{zz} &= 113\,497\,032; \\ S_{xy} &= 583; & S_{xz} &= 0; & S_{zy} &= 1\,082\,587;\end{aligned}$$

aus denen man dann die Schätzungen der Parameter erhält:

$$\hat{a} = 22,24; \quad \hat{b} = 0,01181; \quad \hat{c} = 0,009538.$$

Die quadratische Funktion beschreibt den Zusammenhang zufriedenstellend:  $B_{quad} = 94\%$  der Streuung der  $y$ -Werte wird durch die Parabel erklärt (vgl. Abb. 2). Die Scheitelwerte haben hier eine direkte Interpretation:  $\hat{y}_m = 22,2$  Sek. ist die minimale Reaktionszeit und  $\hat{x}_m = -0,61^\circ$  der zugehörige Blickwinkel, der (erwartungsgemäß) nur gering von  $0^\circ$  abweicht.

Man könnte hier auch eine einfachere Form des quadratischen Zusammenhangs verwenden, bei dem man von der naheliegenden Annahme  $\hat{x}_m = 0$  ausgeht. Dann reduziert sich die Scheiteldarstellung zu

$$y = f(x) = c u + y_m$$

mit  $u = x^2$ . Dies entspricht einem linearen Zusammenhang von  $y$  und der transformierten Variablen  $u$ . Die sich aus der zugehörigen Regressionsgerade ergebenden Schätzungen  $\hat{c}$  und  $\hat{y}_m$  stimmen im Rahmen der angegebenen Genauigkeit mit obigen Schätzungen überein.

Man beachte, daß hier die Reduktion auf einen *linearen* Zusammenhang von  $y$  und  $u = x^2$  nur möglich ist, weil wir die Kenntnis der Scheitelposition  $x_m$  *vorausgesetzt* haben, bzw.  $x_m = 0$  *postuliert* haben. Generell ergibt sich bei *Kenntnis* von  $x_m$  mit der Transformation  $u = (x - x_m)^2$  stets ein linearer Zusammenhang zwischen  $y$  und  $u$ . Im allgemeinen Fall, d.h. bei Unkenntnis von  $x_m$ , gibt es allerdings keine linearisierende Transformation von  $x$ .

Das Beispiel wird in **3.6.1e** noch einmal aus anderer Sicht behandelt.

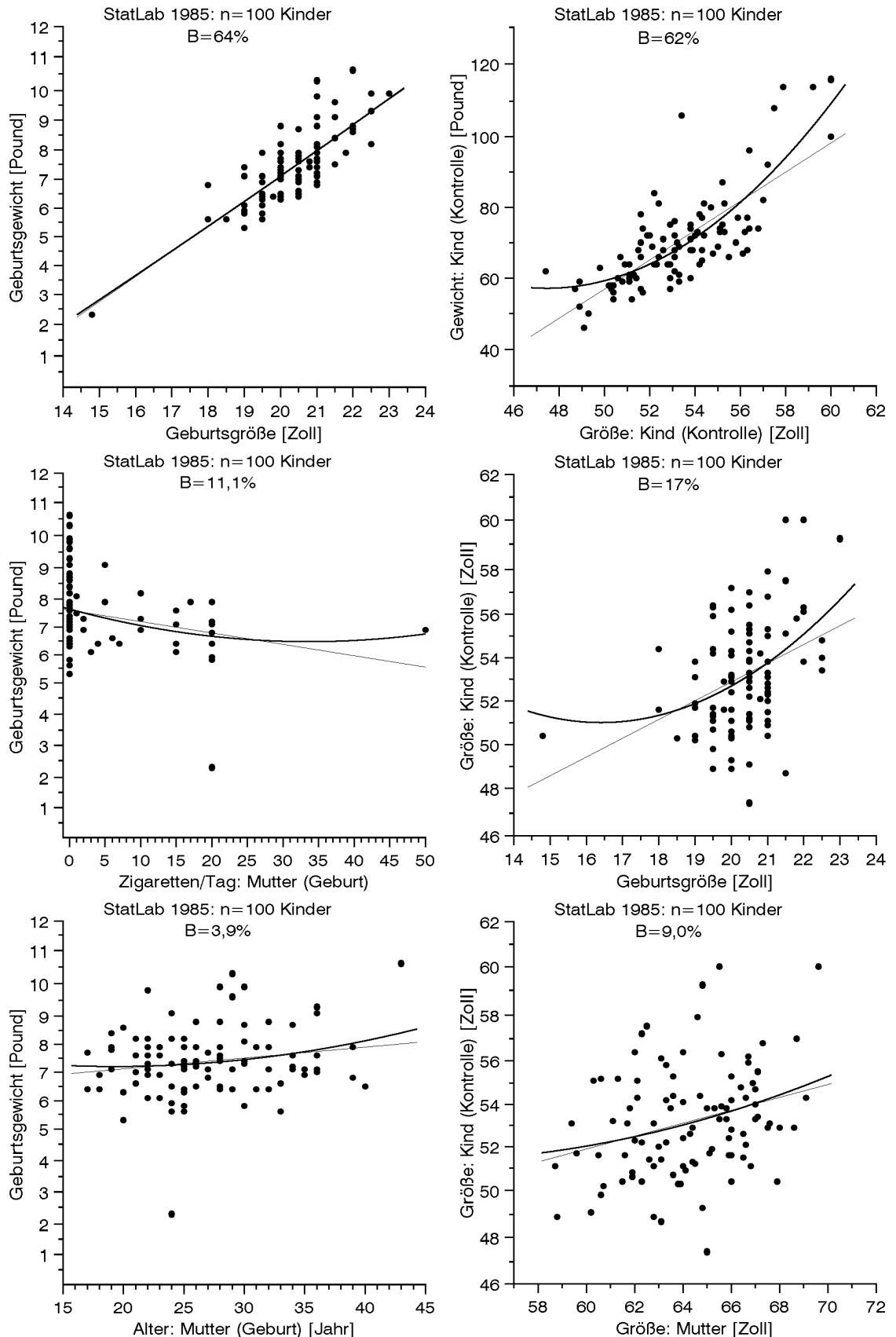
### 1.7.2 Quadratischer vs. linearer Zusammenhang

Die *quadratischen* Funktionen  $y = a + bx + cx^2$  sind gegenüber den *linearen* Funktionen  $y = a + bx$  wesentlich flexibler, weil sie sowohl *gekrümmte Kurven* (für  $c \neq 0$ ) als auch *Geraden* (für  $c = 0$ ) zulassen. Insbesondere kann der quadratische Zusammenhang eine *Trendwende* wiederspiegeln: für *wachsendes*  $x$  *fällt* (oder *steigt*)  $y$  zunächst bis  $x$  den Scheitelpunkt  $x_m$  erreicht und *steigt* (oder *fällt*) dann wieder (vgl. Abb. 1).

Wenn nicht bereits aus Vorkenntnissen klar ist, welcher Zusammenhang vorliegen muß, so sollte man die beobachteten Punkte zunächst graphisch darstellen. Falls diese Punktwolke einen *linearen Trend* aufweist, so wird man zunächst die Regressionsgrade ermitteln. Wenn die Punktwolke jedoch eine *Krümmung* erkennen läßt, so hat man verschiedene Möglichkeiten. Einerseits kann man die quadratische Regressionsfunktion bestimmen. Andererseits kann man die  $x$ -Werte und/oder die  $y$ -Werte *transformieren* und das obige Spiel mit den *transformierten* Punkten erneut durchführen.

Selbst wenn man der Meinung ist, daß eine Regressionsgerade die beobachteten (oder transformierten) Punkte bereits gut erklärt, so ist es empfehlenswert, zusätzlich zu Prüfen, ob die Punktwolke nicht doch eine gewisse Krümmung aufweist. Hierzu berechnet man die quadratische Regressionsfunktion und ihr Bestimmtheitsmaß. Wenn dieses Bestimmtheitsmaß das der Regressionsgeraden deutlich übersteigt, so wird man dem quadratischen Zusammenhang gegenüber dem linearen den Vorzug geben. Sind dagegen beide Bestimmtheitsmaße ungefähr gleich groß, so liegt keine wesentliche Krümmung vor und man wird sich für den einfacheren linearen Trend entscheiden.

Zum Vergleich sind in Abb. 3 für einige Merkmale der StatLab-Auswahl 1985 die beobachteten Punkte zusammen mit der quadratischen Regressionsfunktion und Regressionsgeraden dargestellt. Für Geburtsgröße und Geburtsgewicht (links oben) stimmen beide Funktionen im Rahmen der Zeichengenauigkeit (fast) überein, während sie in den anderen fünf Fällen besonders in den Randbereichen der Daten unterschiedlich stark voneinander abweichen. Für die stärkere Abweichung in den Randbereichen sind hierbei im wesentlichen *einzelne* beobachtete Punkte in diesem Bereich verantwortlich. Ein Vergleich mit **1.5** Abb.4 zeigt, daß die Bestimmtheitsmaße der Regressionsgeraden in allen sechs Fällen nur unwesentlich geringer sind als die der quadratischen Funktion, und somit der quadratische Zusammenhang die Daten nicht wesentlich besser erklärt als der lineare Trend.



**Abb. 3:** Quadratische Regressionsfunktion mit Bestimmtheitsmaß  $B_{quad}$  und Regressionsgerade (dünn) für Merkmale der StatLab-Auswahl 1985 (vgl. auch 1.5 Abb.4).

### 1.7.3 Quadratischer Zusammenhang bei transformierten Variablen

Analog zu den linearisierenden Transformationen, lässt sich die Anwendbarkeit des quadratischen Zusammenhangs wieder entscheidend erweitern, indem man zusätzlich Transformationen  $u = u(x)$  und  $v = v(y)$  verwendet. Durch Zulassen der *identischen* Transformation werden hierbei auch die Fälle mit einbezogen, in denen eine oder beide Variablen  $x$  und  $y$  *nicht* transformiert wird.

Durch einen quadratischen Zusammenhang der *transformierten* Variablen

$$v = a + b u + c u^2$$

ergeben sich dann – je nach Art der verwendeten Transformationen – neue Formen des Zusammenhangs zwischen den ursprünglichen Variablen  $x$  und  $y$ . Die Bestimmung der optimalen Parameterwerte kann dann (analog 1.6.6) nach der Methode der kleinsten Quadrate auf der *transformierten* Skala erfolgen. Eine approximative Lösung zur Minimierung auf der *Originalskala* ergibt sich wieder durch Verwendung der Gewichtsfaktoren aus 1.6.6 (1) bzw. (2).

Weitere allgemeinere Formen des Zusammenhangs ergeben sich, wenn man die quadratische Funktion um einen *kubischen* Terms erweitert zu

$$v = a + b u + c u^2 + d u^3 \quad (\text{kubische Funktion}) .$$

Durch Hinzunahme höherer Potenzen von  $u$  erhält man *Polynome*, auf die wir hier jedoch nicht weiter eingehen wollen.

### Säuglingssterblichkeit und Geburtsgewicht

Karn und Penrose haben in ihrem Artikel *Birth weight and gestation time in relation to maternal age, parity and infant survival* (Annals of Eugenetics 16, 1951-52, 147-164) erstmals den Einfluß verschiedener Variablen auf die Säuglingssterblichkeit unter Verwendung statistischer Modelle analysiert. Die dort untersuchten Daten stammen aus Krankenhausunterlagen der Jahre 1935-1946 in England. Uns interessiert hier nur die Abhängigkeit der Säuglingssterblichkeit vom Geburtsgewicht, und wir betrachten hierzu die insgesamt 7 036 *männlichen* Neugeborenen dieser Studie. In der Tabelle 1 sind für jedes Geburtsgewicht  $x_i$  [in engl. Pounds] die zugehörige Anzahl  $m_i$  der Geborenen, die Zahl  $r_i$  der davon innerhalb von 28 Tagen Gestorbenen (inkl. Totgeburten) und die Sterberate  $y_i = r_i/m_i$  angegeben.

Säuglingssterblichkeit in England 1935-46: Jungen						
	Geburts-	Anzahl	Anzahl	Sterberate	Logit der	
	Gewicht	Geborene	Gestorbene		Sterberate	
Nr	[Pound]	x	m	r	y = r / m	v=logit(y)
1	1,0	13	13	1,000	∞	∞
2	1,5	8	8	1,000	∞	∞
3	2,0	20	19	0,950	2,944	
4	2,5	15	13	0,867	1,872	
5	3,0	39	29	0,744	1,065	
6	3,5	46	20	0,435	-0,262	
7	4,0	60	14	0,233	-1,190	
8	4,5	99	18	0,182	-1,504	
9	5,0	170	16	0,094	-2,264	
10	5,5	360	28	0,078	-2,473	
11	6,0	721	16	0,022	-3,786	
12	6,5	1 002	25	0,025	-3,666	
13	7,0	1 403	41	0,029	-3,503	
14	7,5	1 141	26	0,023	-3,759	
15	8,0	942	18	0,019	-3,938	
16	8,5	459	16	0,035	-3,321	
17	9,0	301	4	0,013	-4,307	
18	9,5	132	5	0,038	-3,235	
19	10,0	62	6	0,097	-2,234	
20	10,5	25	2	0,080	-2,442	
21	11,0	15	2	0,133	-1,872	
22	11,5	2	0	0,000	-∞	
23	13,0	1	1	1,000	∞	∞
Summe		7 036	340			

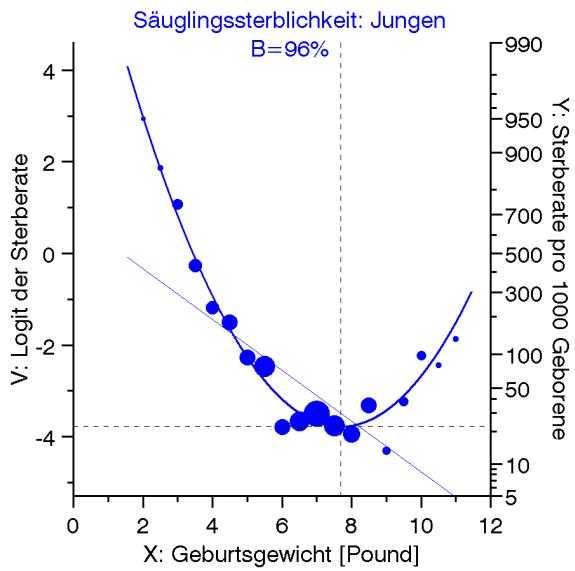


Tabelle 1: Daten zur Säuglingssterblichkeit nach Karn & Penrose

Abb. 4: Beobachtete Punkte (Größe entspricht Gewichtungsfaktor) mit geschätzter quadratischer Funktion und ihren Scheitelwerten. Zum Vergleich ist auch die Regressionsgerade (dünn) eingezeichnet.

Zur Beschreibung des Zusammenhangs verwenden wir wie in 1.6.7 die logistische Transformation  $v = \text{logit}(y)$  der Sterberate, die (im Gegensatz zu  $y$ ) jeden Wert zwischen  $-\infty$  und  $+\infty$  annehmen kann. Die Darstellung der  $(x, v)$ -Punkte in Abb. 4 zeigt eine klare Trendwende: die (transformierte) Sterberate fällt zunächst bei steigendem Geburtsgewicht und steigt dann wieder an. Deshalb ist ein linearer logistischer Zusammenhang  $v = a + bx$  hier nicht angemessen und wir verwenden den quadratischen logistischen Zusammenhang

$$\text{logit}(y) = v = a + b x + c x^2.$$

Zur Schätzung der Parameter  $a, b$  und  $c$  nach der Methode der (gewichteten) kleinsten Quadrate verwenden wir wie schon in 1.6.7 die Gewichtsfaktoren  $w_i = m_i \cdot y_i(1-y_i)$  (nicht zu verwechseln mit dem Geburtsgewicht), um die stark variierende Streuung der  $v$ -Werte zu berücksichtigen (vgl. Abb. 4). Für die beiden kleinsten und größten Geburtsgewichte ist  $w_i = 0$ , weil dort  $y_i = 1$  oder  $y_i = 0$  gilt. Die zugehörigen insgesamt 24 Neugeborenen (die nur einen sehr geringen Anteil der 7 036 Geborenen ausmachen) werden daher bei der Schätzung nicht berücksichtigt. Die erforderlichen Daten zur Auswertung nach Rechenbox 4 (hier mit  $v$  statt  $y$ ) sind in Tabelle 2 zusammengestellt.

Säuglingssterblichkeit in England 1935-46: Jungen

Nr	w	wx	wv	wxx=wz	wzz	wvv	wxz	wxv	wzv
3	0,95	1,90	2,80	3,80	15,20	8,24	7,60	5,59	11,19
4	1,73	4,33	3,24	10,83	67,71	6,07	27,08	8,11	20,28
5	7,44	22,31	7,92	66,92	602,31	8,43	200,77	23,75	71,25
6	11,30	39,57	- 2,97	138,48	1 696,36	0,78	484,67	- 10,38	- 36,33
7	10,73	42,93	- 12,77	171,73	2 747,73	15,19	686,93	- 51,07	- 204,29
8	14,73	66,27	- 22,15	298,23	6 039,10	33,32	1 342,02	- 99,68	- 448,56
9	14,49	72,47	- 32,82	362,35	9 058,82	74,32	1 811,76	- 164,10	- 820,50
10	25,82	142,02	- 63,86	781,12	23 628,95	157,91	4 296,17	- 351,21	- 1 931,66
11	15,64	93,87	- 59,23	563,22	20 275,84	224,21	3 379,31	- 355,35	- 2 132,12
12	24,38	158,45	- 89,35	1 029,90	43 513,13	327,54	6 694,33	- 580,80	- 3 775,20
13	39,80	278,61	- 139,43	1 950,29	95 564,25	488,45	13 652,04	- 976,02	- 6 832,14
14	25,41	190,56	- 95,49	1 429,17	80 391,04	358,92	10 718,80	- 716,21	- 5 371,57
15	17,66	141,25	- 69,54	1 129,99	72 319,18	273,85	9 039,90	- 556,28	- 4 450,27
16	15,44	131,26	- 51,28	1 115,70	80 609,59	170,31	9 483,48	- 435,91	- 3 705,23
17	3,95	35,52	- 17,00	319,69	25 895,24	73,23	2 877,25	- 153,01	- 1 377,06
18	4,81	45,70	- 15,56	434,16	39 182,69	50,34	4 124,49	- 147,83	- 1 404,39
19	5,42	54,19	- 12,10	541,94	54 193,55	27,04	5 419,35	- 121,05	- 1 210,46
20	1,84	19,32	- 4,49	202,86	22 365,32	10,98	2 130,03	- 47,19	- 495,45
21	1,73	19,07	- 3,24	209,73	25 377,73	6,07	2 307,07	- 35,69	- 392,58
$\Sigma$	243,28	1 559,60	- 677,33	10 760,12	603 543,73	2 315,18	78 683,07	- 4 764,32	- 34 485,10
	<i>x-Mittelwert</i>	<i>z-Mittelwert</i>	<i>y-Mittelwert</i>	<i>Swxx</i>	<i>Swzz</i>	<i>Swvv</i>	<i>Swxz</i>	<i>Swxv</i>	<i>Swzv</i>
	6,4107	44,2294	- 2,7842	761,942	127629,518	429,367	9702,814	- 422,127	- 4527,078
Nenner <i>D</i>	Koeffizienten der Regressionsparabel <i>a</i> <sup>^</sup> <i>b</i> <sup>^</sup> <i>c</i> <sup>^</sup>			Scheitelkoordinaten <i>xs</i> <i>vs</i>		<i>ys</i>	Streuung <i>WSQ</i>	Bestimmtheit <i>B</i>	
3101719,41	8,5637 - 3,2081 0,2084			7,70 - 3,78		2,2% 18,68		95,7%	

**Tabelle 2:** Analyse der Säuglingssterblichkeits-Daten aus Tabelle 1 nach *Rechen-Box 4* (hier mit *v* statt *y*).

Die quadratische Regressionsfunktion  $v = \hat{a} + \hat{b}x + \hat{c}x^2$  erklärt  $B_{quad} = 96\%$  der Streuung der *v*-Werte und ihr Verlauf in Abb. 4 stimmt gut mit den beobachteten Werten überein. Der Scheitelpunkt  $x_s = 7,7$  und  $v_s = -3,78$  repräsentiert hier das „günstigste“ Geburtsgewicht mit der niedrigsten Sterblichkeit von nur  $y_s = 2,2\%$ .

Nebenbei sei noch angemerkt, daß die logistische *Regressionsgerade*  $v = \hat{a} + \hat{b}x$  auch schon  $B_{xv} = 54\%$  der Streuung der *v*-Werte erklärt. Allerdings zeigt Abb. 4 klar, daß die Regressionsgerade die beobachteten Punkte völlig unzureichend beschreibt. Die Lehre hieraus ist, daß es nicht ausreicht, eine Regressionsgerade und ihr Bestimmtheitsmaß zu berechnen, sondern man muß sie auch zusammen mit den beobachteten Punkten *graphisch darstellen*, um die Güte der Anpassung beurteilen zu können.

### Tipps für Microsoft Excel

**Berechnung:** In einer Tabelle (wie z.B. Tab. 2) bestimmt man zunächst alle *gewichteten* Summen  $\sum w, \dots, \sum wzy$ , aus denen man dann (mit *Rechen-Box 4*) alle weiteren Größen berechnen kann. Man kann hierbei sogar auf die Spalten mit den Produkten  $wx, \dots, wzv$  verzichten, weil sich die zugehörigen Spaltensummen direkt mit der *Excel-Funktion SUMMENPRODUKT* berechnen lassen (vgl. Tipps in 1.5.6).

**Graphische Darstellung:** Die *beobachten* Datenpaare lassen sich auf der *transformierten* bzw. *Original-Skala* (wie in 1.5.5 erläutert) darstellen. Und das Zeichnen von *Funktionen* (wie z.B. in Abb. 4) wurde bereits in 1.5.6 erläutert.

## 2 Grundbegriffe der Stochastik

In der *beschreibenden Statistik* (Kapitel 1) haben wir die zu analysierenden *Daten* als *gegeben* hingenommen und nicht genauer hinterfragt, auf welche Art und Weise sie *entstanden sind*. Will man aus den beobachteten Daten jedoch *Schlußfolgerungen* ziehen, die wesentlich über die *Beschreibung* der Daten hinausgehen, so muß man auch die Entstehung der Daten mit berücksichtigen. Zum Beispiel sind Verallgemeinerungen von den *StatLab-Daten* auf *alle* amerikanischen Kinder nur dann zulässig, wenn die Datenerhebung auch *repräsentativ* für diese Gesamtheit war. Analysiert man die Erhebung der Daten, so stößt man auf folgendes *Problem*:

Experimentelle oder beobachtete Daten sind nicht *exakt reproduzierbar*. Ursachen hierfür sind z.B. Variabilitäten (der Untersuchungsobjekte oder Versuchsbedingungen), nicht kontrollierbare Meßfehler etc.

Zur Lösung dieses Problems geht man von einer *zufälligen Variabilität* der Daten aus und berücksichtigt den *Zufall* bei der Auswertung. Die mathematische Disziplin zur Modellierung und Untersuchung des *Zufalls* ist die *Stochastik*. Sie enthält die *Wahrscheinlichkeitstheorie* und die darauf aufbauende *mathematische (deduktive) Statistik*. Die Wahrscheinlichkeitstheorie beschäftigt sich mit:

- der Konstruktion von Modellen für stochastische *Vorgänge*:  
Zufallsexperimente oder Zufalls-Beobachtungen,
- der Analyse solcher (stochastischer) Modelle,
- der Berechnung von Wahrscheinlichkeiten für interessierende Ereignisse.

Zu den Aufgaben der *deduktiven Statistik* zählt die *Planung* und *Auswertung* konkreter Studien (Experimente, Beobachtungen). Die Auswertung beschäftigt sich primär mit

- dem Schätzen unbekannter Modell-Parameter,
- dem Testen von Hypothesen über Modell-Parameter.

Und bei der Planung stehen im Vordergrund

- Auswahl eines geeigneten Auswertungs-Modells nebst zugehörigem Design der Datenerhebung,
- Kalkulation des erforderlichen Stichprobenumfangs, der eine zuverlässige statistische Auswertung erlaubt.

Bevor wir zur *deduktiven Statistik* kommen, wollen wir die hierfür notwendigen Grundlagen der *Wahrscheinlichkeitsrechnung* kurz zusammenstellen und erläutern.

## 2.1 Wahrscheinlichkeiten und Zufallsvariablen

### 2.1.1 Definitionen und Beispiele

**Stochastischer Vorgang:** Ausgangspunkt unserer Betrachtungen ist ein *stochastischer Vorgang*, d.h. ein *Zufallsexperiment* oder eine *Zufalls-Beobachtung*, der charakterisiert ist durch folgende Eigenschaften:

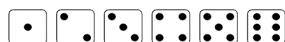
- Es gibt *mehrere mögliche Ergebnisse*.  
Die Menge  $\Omega$  aller *Ergebnisse* heißt der *Ergebnis- oder Stichprobenraum*.
- Das Ergebnis ist *nicht exakt vorhersagbar* und *nicht reproduzierbar*.
- Der Vorgang ist (prinzipiell bzw. in Gedanken) *wiederholbar*.

Wir geben zunächst einige Beispiele für solche stochastischen Vorgänge an.

stochastischer Vorgang	Ergebnisse	Ereignisse (Beispiele)
Werfen eines Würfels	Zahlen: $1, 2, \dots, 6$	<i>gerade Zahl:</i> $\{2, 4, 6\}$
Roulette-Spiel	Zahlen: $0, 1, \dots, 36$	<i>Rot:</i> $\{1, 3, \dots, 34, 36\}$ <i>1. Dutzend:</i> $\{1, 2, \dots, 12\}$
Wahlumfrage: „Sonntagsfrage stellen“	alle zugelassenen Parteien	<i>Ampel-Koalition:</i> $\{ \text{SPD}, \text{FDP}, \text{Grüne} \}$
Ziehung aus Urne mit markierten Kugeln	alle verschiedenen Markierungen	spezielle Markierung, z.B. <i>schwarz, weiß</i>
Messung eines Gewichts [kg]	positive Zahlen	<i>Übergewicht, Untergewicht</i>

Tabelle 1: Einige Beispiele für stochastische Vorgänge, die im folgenden Text näher erläutert werden.

**Werfen eines Würfels:**



Die möglichen Ergebnisse sind die Zahlen von 1 bis 6, d.h. der Ergebnisraum ist  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .  $\square$

**Roulette-Spiel:** Hier sind die Zahlen von 0 (Zero) bis 36 als Ergebnisse möglich (vgl. Abb. 1) und somit ist  $\Omega = \{0, 1, \dots, 36\}$ .  $\square$

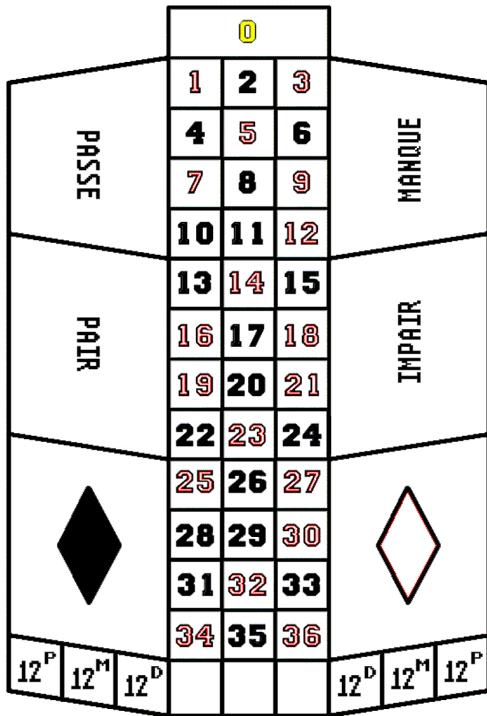
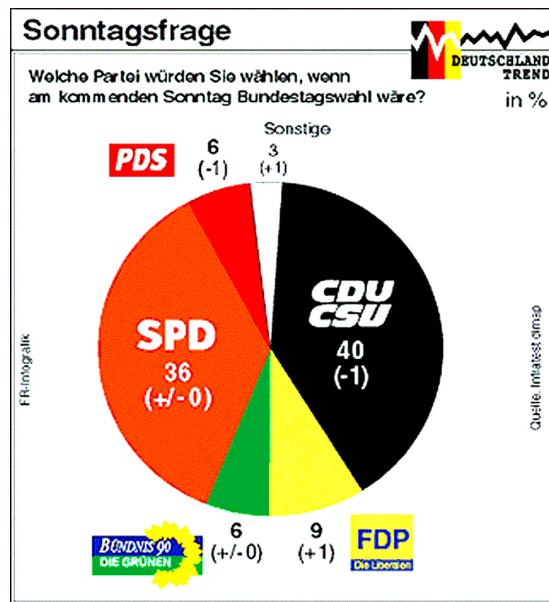


Abb. 1 (links): Roulette-Spielfeld

Abb. 2 (rechts): Wahlumfrage



Untersuchungsanlage

Grundgesamtheit: Wahlberechtigte Bevölkerung in Deutschland

ab 18 Jahren

Stichprobe: Repräsentative Zufallsauswahl/Randomstichprobe

Erhebungsverfahren: Computergestützte Telefoninterviews (CATI)

Fallzahl: 1.300 Befragte (900 West, 400 Ost)

Erhebungszeitraum: 21. - 26. Februar 2002

Fehlertoleranz: 1,2\* bis 2,7\*\* Prozentpunkte

\* bei einem Anteilswert von 5%

\*\* bei einem Anteilswert von 50%

Durchführendes Institut: Infratest dimap

**Wahlumfrage:** Aus allen Wahlberechtigten Deutschlands wird *zufällig* eine Person ausgewählt und ihr die „*Sonntagsfrage*“ gestellt: „Welche Partei würden Sie wählen, wenn am kommenden Sonntag Bundestagswahl wäre?“ Die möglichen Ergebnisse einer solchen Einzelbefragung sind alle zur Wahl zugelassenen Parteien: SPD, CDU, Bündnis 90/Die Grünen, PDS, FDP etc. Um zu sinnvollen Schlußfolgerungen zu kommen, wird dieser Vorgang einer *Einzelbefragung* natürlich wiederholt, d.h. es werden bei größeren Umfragen z.B. 1300 (vgl. Abb. 2) und bei Blitzumfragen z.B. 100 Befragungen durchgeführt. □

**Ziehen aus einer Urne:** Wir betrachten eine *Urne* mit  $n$  Kugeln, die *unterschiedlich markiert* sind, etwa durch *Zahlen* (z.B. von 1 bis 49 bei den Lotto-Zahlen) oder durch *Farben* (schwarz, weiß, rot, blau ...). Das zufällige Ziehen einer Kugel aus der Urne ist dann ein *stochastischer Vorgang*, von dem uns als Ergebnis nur die *Markierung* (Zahl bzw. Farbe) der gezogenen Kugel interessiert. Der Ergebnisraum  $\Omega$  ist daher die Menge der unterschiedlichen Markierungen (Zahlen bzw. Farben).

Das Ziehen aus einer Urne ist ein universelles Modell für viele stochastische Vorgänge. Enthält die Urne z.B. nur 6 Kugeln mit den Zahlen von 1 bis 6, so entspricht das Ziehen einer Kugel prinzipiell dem Werfen eines (idealen bzw. homogenen) Würfels. Und das Roulette-Spiel entspricht dem Ziehen aus einer Urne mit 37 Kugeln und den Markierungen von 0 bis 36. Sogar die Einzelbefragung einer Wahlumfrage lässt sich als Ziehen einer Kugel (Person) aus der „Urne aller Wahlberechtigten“ auffassen, wobei die Markierung der Kugel diejenige Partei angibt, der die jeweilige Person ihre Stimme geben würde. □

**Messung des Körpergewichts:** Das Ergebnis einer Messung des Körpergewichts [in kg] einer Person ist eine *Dezimalzahl*. Obwohl man das Gewicht typischerweise nur auf höchstens eine Nachkommastelle angibt, kommen bei beliebig genauer Messung alle Zahlen eines *Zahlen-Intervalls* als Meßergebnis in Frage. Die genaue Bestimmung *aller möglichen* Ergebnisse ist hier nicht ganz einfach, weil das kleinstmögliche und größtmögliche Gewicht nicht ohne weiteres angebar sind. Man wählt daher den Ergebnisraum  $\Omega$  vorsichtshalber groß genug, z.B. von 0 bis 300 [kg] auch wenn dann nicht alle diese Ergebnisse biologisch möglich sind.

Eine solche Gewichtsmessung kann auf drei verschiedene Weisen als stochastischer Vorgang interpretiert werden. *Erstens* kann das Meßverfahren (also die Waage) aufgrund zufälliger Störungen ungenau sein, sodaß eine Wiederholungsmessung nicht zum exakt gleichen Ergebnis kommt. *Zweitens* wird man selbst bei genauerer Messung zu einem späteren Zeitpunkt ein anderes Ergebnis erhalten, weil das Körpergewicht wegen Nahrungsaufnahme und Ausscheidung im Laufe eines Tages variiert. Wählt man schließlich die Person zunächst *zufällig* aus einer interessierenden Gesamtheit aus und bestimmt ihr Gewicht, so ergibt sich die dritte Interpretation einer Gewichtsmessung als stochastischer Vorgang.  $\square$

## Ereignisse und Wahrscheinlichkeiten

Ein *Ereignis*  $A$  ist eine *Menge* von Ergebnissen, also eine *Teilmenge* des Stichprobenraums  $\Omega$ , d.h. in Mengen-Schreibweise:  $A \subset \Omega$  (*lies*:  $A$  ist enthalten in  $\Omega$ ). Man sagt, dass das Ereignis  $A$  bei einem Ergebnis  $\omega$  *eingetreten* ist, wenn das Ergebnis  $\omega$  in  $A$  liegt, d.h.  $\omega \in A$  gilt. Wenn  $A$  nur *ein einziges* Ergebnis  $\omega$  enthält, d.h.  $A = \{\omega\}$ , so heißt  $A$  auch ein *Elementar-Ereignis*.

Zur Quantifizierung des „Zufalls“ wird jedem *Ereignis*  $A$  eine Zahl  $P(A)$  aus dem Intervall von  $0 = 0\%$  bis  $1 = 100\%$  zugeordnet, die man als *Wahrscheinlichkeit* für das *Eintreten* von  $A$  bezeichnet. Diese Wahrscheinlichkeit  $P(A)$  ist interpretierbar als der *Grenzwert der relativen Häufigkeiten* für das Eintreten von  $A$  in einer *unendlichen Folge von unabhängigen Wiederholungen* des Experiments (*Häufigkeitsinterpretation der Wahrscheinlichkeit*, vgl. Abb. 3).

**Werfen eines Würfels** (Fortsetzung): Bei einem *homogenen* (oder *idealen*) Würfel hat jedes der 6 möglichen Ergebnisse  $\omega \in \{1, \dots, 6\}$  dieselbe Wahrscheinlichkeit von  $\frac{1}{6}$ . Das Ereignis „*gerade Zahl*“ entspricht der Menge  $\{2, 4, 6\}$  und das Ereignis „*mindestens eine 5*“ entspricht der Menge  $\{5, 6\}$ .  $\square$

**Roulette-Spiel** (Fortsetzung): Beim (fairen) Roulette hat jedes der 37 möglichen Ergebnisse  $\omega \in \{0, 1, \dots, 36\}$  dieselbe Wahrscheinlichkeit von  $\frac{1}{37}$ . Interessierende Ereignisse, auf deren Eintreten die Mitspieler setzen können, sind z.B. (vgl. Abb. 1):

*Rot:*  $\{1, 3, \dots, 34, 36\}$  (alle *roten* Zahlen)

*1. Dutzend* ( $12^P$ ):  $\{1, 2, \dots, 12\}$  (die ersten 12 Zahlen von 1 bis 36).  $\square$

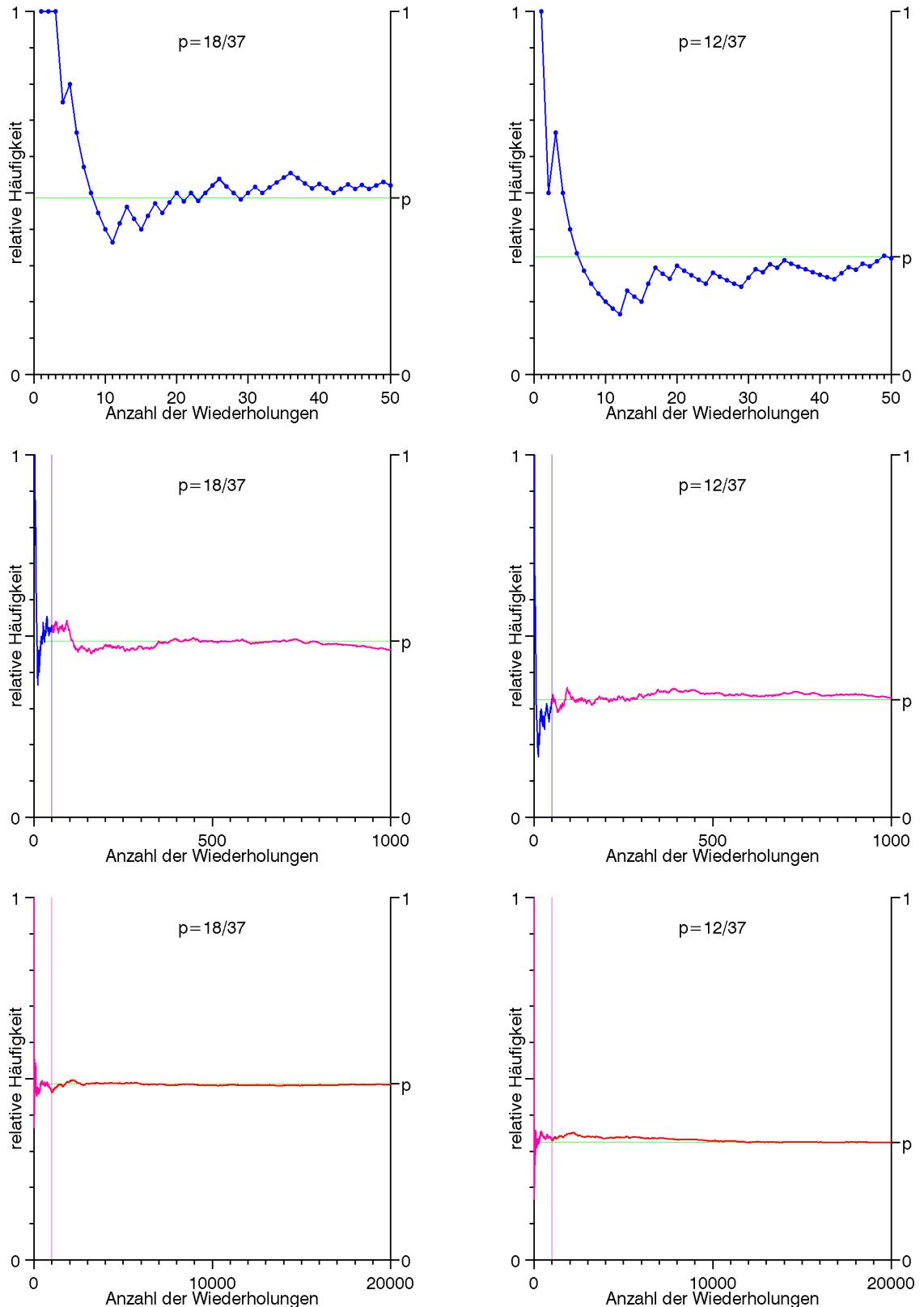
**Wahlumfrage (Fortsetzung):** Die Wahrscheinlichkeit bei einer Einzelbefragung, eine bestimmte Partei (z.B. die *SPD*) als Ergebnis zu erhalten, entspricht gerade dem aktuellen Stimmanteil dieser Partei (*SPD*) unter allen Wahlberechtigten (und ist *unbekannt*). Interessierende Ereignisse sind hier Koalitionen, z.B. die *Große Koalition*  $\{SPD, CDU\}$  oder die *Ampel-Koalition*  $\{SPD, FDP, Grüne\}$ .  $\square$

**Ziehen aus einer Urne (Fortsetzung):** Die Wahrscheinlichkeit, eine Kugel mit einer speziellen Markierung (z.B. *schwarz*) zu ziehen, entspricht genau dem Anteil der so *markierten* Kugeln unter *allen N* Kugeln, bei  $M$  *schwarzen* Kugeln also  $\frac{M}{N}$ .  $\square$

**Messung des Körpergewichts (Fortsetzung):** Beim *Geburtsgewicht* eines zufällig ausgewählten Neugeborenen interessiert man sich z.B. für die Ereignisse *Untergewicht* (d.h. Gewicht  $\leq 2,5 \text{ kg}$ ), *Normalgewicht*, *Übergewicht*. Diese Ereignisse sind dadurch definiert, daß das Gewicht in einem bestimmten *Intervall* liegt. Selbst ein auf 100 Gramm *gerundetes Gewicht* (genauer werden Geburtsgewichte meist nicht bestimmt) von z.B. 3,5 kg bedeutet genau genommen, daß das exakte Gewicht im Intervall von 3,45 kg (inkl.) bis 3,55 kg (exkl.) liegt, weil alle Werte dieses Intervalls auf 3,5 kg gerundet werden. Wie wir noch sehen werden (in 2.4) ist es nicht sinnvoll, jedem exakten Ergebnis von z.B. 3,500000... kg (mit unendlich vielen Nullen!) eine positive Eintrittswahrscheinlichkeit zuzuordnen, sondern nur den *Intervallen* (und somit auch den in der Praxis auftretenden *gerundeten* Gewichten).  $\square$

*Anmerkung für Fortgeschrittene:* Wenn der Stichprobenraum  $\Omega$  (wie beim Geburtsgewicht oben) ein Zahlenintervall (d.h. ein Abschnitt auf der Zahlengeraden) ist, so läßt sich nicht *jeder* Teilmenge von  $\Omega$  in sinnvoller Weise eine Wahrscheinlichkeit zuordnen. In diesem Fall bezeichnet man nur die *praktisch relevanten* Mengen als *Ereignisse* und ordnet auch nur diesen Wahrscheinlichkeiten zu. Zu diesen Ereignissen gehören alle Teil-Intervalle von  $\Omega$  sowie Vereinigungen von endlich (und auch abzählbar unendlich) vielen Intervallen (andere Ereignisse werden wir hier auch nicht betrachten).

**Abb. 3:** Relative Häufigkeiten für die Roulette-Ereignisse *Rot* (links) bzw. *1. Dutzend* (rechts) mit Wahrscheinlichkeit  $p = \frac{18}{37}$  bzw.  $p = \frac{12}{37}$  bei wachsender Zahl von Wiederholungen. Der Anfangsabschnitt (bis zum vertikalen Strich) ist eine komprimierte Darstellung der darüberstehenden Grafik.



### Zufallsvariable

Eine *Zufallsvariable*  $X$  ist ein durch *Zahlen meßbarer Aspekt* des Ergebnisses, d.h jedem Ergebnis  $\omega \in \Omega$  ist eine Zahl  $x = X(\omega)$  als Ergebnis (*Realisierung*) von  $X$  zugeordnet. Eine Zufallsvariable  $X$  reduziert jedes Ergebnis  $\omega$  auf eine Zahl  $x = X(\omega)$  und stellt damit typischerweise eine Vereinfachung dar (Reduktion auf das Wesentliche). Bei einem zufällig ausgewählten Neugeborenen einer Population sind aus medizinischer Sicht z.B. das Geburtsgewicht, die Geburtsgröße und die Schwangerschaftsdauer (als Zufallsvariablen) von Interesse, während andere Aspekte (z.B. der Tag der Geburt, der Name) unwesentlich sind.

Ein *Ereignis* für eine Zufallsvariable  $X$  ist durch eine *Zahlenmenge*  $A$  gegeben und wird geschrieben  $\{X \in A\}$  (*lies:  $X$  liegt in  $A$* ). Für die Wahrscheinlichkeit, daß  $X$  in  $A$  liegt, schreibt man  $P\{X \in A\}$ . Wenn  $A$  ein *Intervall* ist, z.B.

$$A = [a, b] = \{x \mid a \leq x \leq b\}, \quad \text{---} \overset{a}{/} \overset{b}{\backslash} \text{---} \quad A$$

so schreibt man die Wahrscheinlichkeit auch suggestiv

$$P\{X \in A\} = P\{X \in [a, b]\} = P\{a \leq X \leq b\}.$$

**Indikator-Variablen:** Interessiert man sich nur dafür, ob ein spezielles *Ziel-Ereignis*  $C$  *eingetreten* ist oder nicht, so kann man dies durch eine *Indikator-Variable*  $X$  kodieren, die nur die Werte 1 (bei Eintritt von  $C$ ) und 0 (bei Nicht-Eintritt von  $C$ ) annimmt, d.h.

$$X(\omega) = \begin{cases} 1 & \text{falls } \omega \in C \quad (C \text{ ist eingetreten}), \\ 0 & \text{falls } \omega \notin C \quad (C \text{ ist nicht eingetreten}). \end{cases}$$

Die Zufallsvariable  $X$  nimmt den Wert 1 mit Wahrscheinlichkeit  $P(C)$  an, und den Wert 0 mit der komplementären Wahrscheinlichkeit

$$P\{X = 1\} = P(C), \quad P\{X = 0\} = 1 - P(C).$$

**Kodierungen:** Ein interessierendes qualitatives Merkmal (z.B. die Blutgruppe im *AB0*-System) wird meistens durch Zahlen kodiert (z.B. durch eine laufende Nummer, vgl. **1.2**) damit es im Computer einfacher zu verarbeiten ist. Eine solche Kodierung ist eine Zufallsvariable  $X$ , die jedem Ergebnis  $\omega$  den zugehörigen Kode  $x = X(\omega)$  zuordnet, z.B. bei den Blutgruppen  $\Omega = \{0, A, B, AB\}$

$$X(0) = 1, \quad X(A) = 2, \quad X(B) = 3, \quad X(AB) = 4.$$

**Diskrete und stetige Zufallsvariablen:** Eine Zufallsvariable heißt *diskret*, wenn ihre möglichen Werte *isoliert* sind und sich *abzählen* lassen:  $x_1, x_2, x_3, \dots$  (es können auch *unendlich* viele Werte sein, z.B. *alle* natürlichen Zahlen  $x_1 = 1, x_2 = 2, x_3 = 3, \dots$ ).

Typische Beispiele hierfür sind Zufallsvariablen, deren Werte im Bereich der ganzen Zahlen liegen, also *Anzahlen* (z.B. Mutanten einer Zellkultur) darstellen. Auch Kodierungen und Indikator-Variablen sind diskret.

Dagegen kann eine *stetige* Zufallsvariable kontinuierlich *alle* Werte eines (endlichen oder unendlichen) Zahlen-Intervalls annehmen, z.B. alle *positiven* Zahlen. Typische Beispiele hierfür sind Messungen eines Gewichts, einer Länge oder Zeit, wobei die Ergebnisse prinzipiell beliebig genau (also *nicht* gerundet) bestimmt werden.

Zufallsvariable $X$	Werte von $X$	Ereignisse für $X$
Augensumme zweier Würfel	$2, 3, \dots, 12$ (diskret)	$\{X \geq 10\}$ bzw. $\{10, 11, 12\}$
Gewinn eines Tips beim Lotto „6 aus 49“ in Euro	$0 - 50$ Mio. (diskret)	$\{X \geq 1$ Mio. $\}$ „ <i>Millionär werden</i> “
Geburtsgewicht [g]	positive Zahlen (stetig)	<i>Untergewicht:</i> $\{X \leq 2500\}$
Wirkstoffgehalt [mg] in Tablette	positive Zahlen (stetig)	<i>Überdosis, Unterdosierung</i>

Tabelle 2: Beispiele für Zufallsvariablen von stochastischen Vorgängen

### 2.1.2 Eigenschaften von Wahrscheinlichkeiten

Wir stellen jetzt wichtige Eigenschaften und Rechenregeln von Wahrscheinlichkeiten zusammen. Die Eigenschaften sind zwar *plausibel* (weil sie analog für *relative Häufigkeiten* gelten), aber sie lassen sich nicht *beweisen*. Streng genommen *charakterisieren* diese Eigenschaften den Wahrscheinlichkeitsbegriff. In der *Wahrscheinlichkeitstheorie* geht man deshalb von einigen dieser Eigenschaften als grundlegenden *Axiomen* aus (die zuerst von *Kolmogorov* formuliert wurden) und leitet alle weiteren Eigenschaften daraus formal ab.

Für jede der folgenden Eigenschaft gebt wir zwei Formulierungen an. Die erste bezieht sich auf Ereignisse im *Ergebnisraum*  $\Omega$  – also auf Teilmengen  $A, B, \dots$  von  $\Omega$  – und die zweite auf Ereignisse für eine *Zufallsvariable*  $X$  – also auf Teilmengen  $A, B, \dots$  der Menge  $\mathbb{R}$  aller *reellen* Zahlen.

(1) Normierung:

$$0 \leq P(A) \leq 1 \quad \text{bzw.} \quad 0 \leq P\{X \in A\} \leq 1.$$

(2) Das sichere Ereignis  $\Omega$  tritt stets ein:

$$P(\Omega) = 1 \quad \text{bzw.} \quad P\{X \in \mathbb{R}\} = 1$$

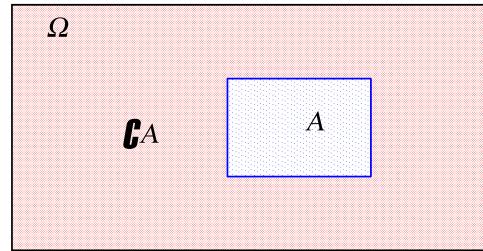
$\mathbb{R} = \text{Menge aller (ellen) Zahlen.}$

(3) Das unmögliche Ereignis  $\emptyset$  (leere Menge) tritt nie ein:

$$P(\emptyset) = 0 \quad \text{bzw.} \quad P\{X \in \emptyset\} = 0.$$

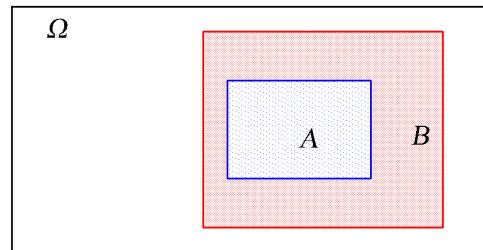
(4) Das zu  $A$  komplementäre Ereignis  $\complement A = \{e \in \Omega \mid e \notin A\}$  (lies: nicht  $A$  oder Komplement von  $A$ ) hat die zu  $P(A)$  komplementäre Wahrscheinlichkeit

$$P(\complement A) = 1 - P(A) \quad \text{bzw.} \quad P\{X \notin A\} = 1 - P\{X \in A\}.$$



(5) Wenn ein Ereignis  $B$  ein anderes Ereignis  $A$  umfaßt, d.h. es gilt  $A \subset B$ , dann ist die Wahrscheinlichkeit für  $B$  mindestens so groß wie die für  $A$

$$P(A) \leq P(B) \quad \text{bzw.} \quad P\{X \in A\} = 1 - P\{X \in B\}.$$



- (6) Die Wahrscheinlichkeit der *Vereinigung zweier Ereignisse* ist die *Summe* der Wahrscheinlichkeiten beider Ereignisse *abzüglich* der Wahrscheinlichkeit des *Durchschnitts* beider Ereignisse:

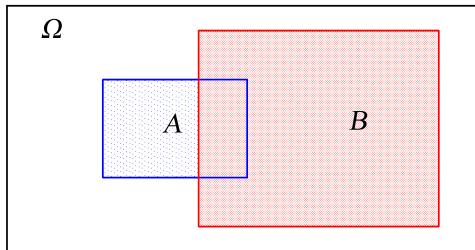
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{bzw.}$$

$$P\{X \in A \text{ oder } X \in B\} = P\{X \in A\} + P\{X \in B\} - P\{X \in A \text{ und } X \in B\}$$

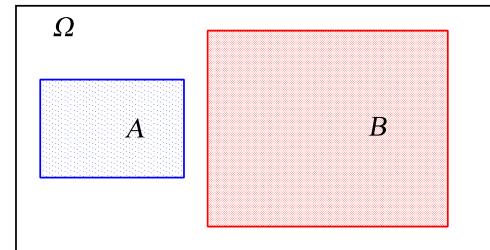
Wenn die Ereignisse *sich gegenseitig ausschließen* (*disjunkt sind*), d.h. ihr Durchschnitt  $A \cup B = \emptyset$  ist leer, so gilt speziell

$$P(A \cup B) = P(A) + P(B)$$

$$P\{X \in A \text{ oder } X \in B\} = P\{X \in A\} + P\{X \in B\}$$



Allgemeiner Fall



Spezialfall: A und B disjunkt

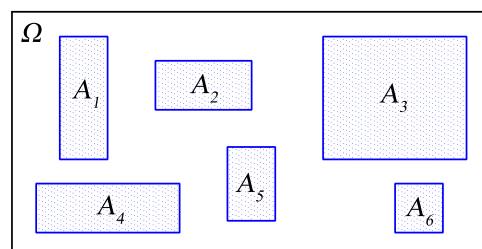
- (7) Wenn die Ereignisse  $A_1, A_2, A_3, \dots$  *paarweise disjunkt sind* (d.h. *sich gegenseitig ausschließen*), so ist die Wahrscheinlichkeit ihrer Vereinigung gleich der Summe der Wahrscheinlichkeiten dieser Ereignisse:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup \dots) &= \sum_i P(A_i) \\ &= P(A_1) + P(A_2) + P(A_3) + \dots \end{aligned}$$

bzw.

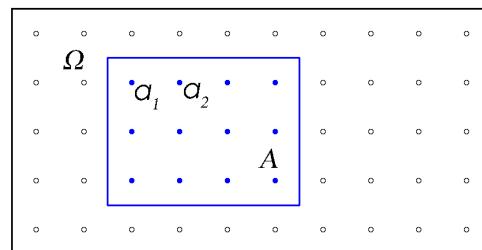
$$\begin{aligned} P\{X \in A_i \text{ für irgendein } i\} &= \sum_i P\{X \in A\} \\ &= P\{X \in A_1\} + P\{X \in A_2\} + P\{X \in A_3\} + \dots \end{aligned}$$

Hierbei sind auch *unendlich viele* Ereignisse  $A_1, A_2, A_3, \dots$  zugelassen.



- (8) Besteht das Ereignis  $A = \{a_1, a_2, \dots, a_n\}$  aus  $n$  verschiedenen Ergebnissen  $a_1, a_2, \dots, a_n$ , so ist die Wahrscheinlichkeit für  $A$  die *Summe* der Wahrscheinlichkeiten aller Ergebnisse aus  $A$ :

$$\begin{aligned} P(A) &= \sum_i P\{a_i\} \\ &= P\{a_1\} + P\{a_2\} + \dots + P\{a_n\} \quad \text{bzw.} \\ P\{X \in A\} &= \sum_i P\{X = a_i\} \\ &= P\{X = a_1\} + P\{X = a_2\} + \dots + P\{X = a_n\} \end{aligned}$$



	0	
1	2	3
4	5	6
7	8	9
10	11	12
13	14	15
16	17	18
19	20	21
22	23	24
25	26	27
28	29	30
31	32	33
34	35	36

zu (4)  
 $A = 2.$  Dutzend

	0	
1	2	3
4	5	6
7	8	9
10	11	12
13	14	15
16	17	18
19	20	21
22	23	24
25	26	27
28	29	30
31	32	33
34	35	36

zu (6)  
 $B = 2.$  Kolonne

	0	
1	2	3
4	5	6
7	8	9
10	11	12
13	14	15
16	17	18
19	20	21
22	23	24
25	26	27
28	29	30
31	32	33
34	35	36

Ereignisse  
zu (7) und (8)

Abb. 4: Ereignisse beim Roulette-Spiel

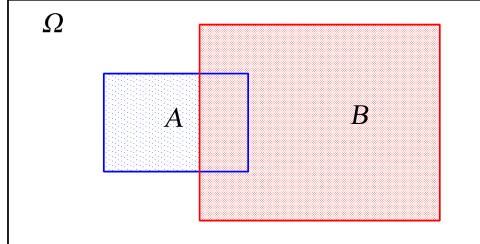
### 2.1.3 Bedingte Wahrscheinlichkeit

#### Zahlen-Lotto „6 aus 49“

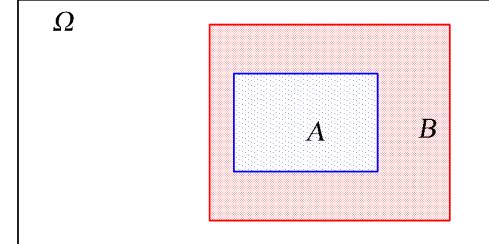
Beim Lotto „6 aus 49“ gibt es  $\binom{49}{6} \approx 14 \cdot 10^6$  verschiedene Ergebnisse einer Ziehung (vgl. 2.3.4). Die Wahrscheinlichkeit, daß ein fester Tip  $\omega$  eines Spielers, also z.B.  $\omega = \{4, 9, 16, 25, 36, 49\}$  gezogen wird, ist daher sehr gering:  $P\{\omega\} = 1/\binom{49}{6} \approx 7 \cdot 10^{-8}$ . Angenommen, der Spieler verfolgt die Ziehung „live“ und hat nach Ziehung der ersten fünf Zahlen  $\{4, 9, 16, 36, 49\}$  bereits „5 Richtige“. Die Wahrscheinlichkeit, daß er auch noch „6 Richtige“ erzielt, beträgt jetzt  $1/44 \approx 2,3\%$ , weil hierzu nur noch die fehlende Zahl 25 seines Tips aus den verbleibenden 44 Kugeln gezogen werden muss.  $\square$

Wie das Lotto-Beispiel zeigt, kann sich die Wahrscheinlichkeit für ein Ereignis dramatisch ändern, wenn man zusätzliche Informationen mitberücksichtigt. Wir wollen dies jetzt allgemeiner betrachten. Sind  $A$  und  $B$  zwei Ereignisse, so fragen wir nach der Wahrscheinlichkeit von  $A$  unter der Bedingung, daß das Ereignis  $B$  eintritt. Diese *bedingte Wahrscheinlichkeit*  $P(A | B)$  von  $A$  unter der Bedingung  $B$  ist definiert durch (die zweite Formulierung bezieht sich auf Zufallsvariablen  $X$  und  $Y$ ):

$$(1) \quad P(A | B) = \frac{P(A \cap B)}{P(B)} \quad \text{bzw.} \quad P\{X \in A | Y \in B\} = \frac{P\{X \in A \text{ und } X \in B\}}{P\{Y \in B\}}$$



Allgemeiner Fall



Spezialfall: A ist in B enthalten

Bei der Berechnung der auf  $B$  bedingten Wahrscheinlichkeit  $P(A | B)$  von  $A$  wird nur der in  $B$  liegende Teil von  $A$  berücksichtigt, und die Wahrscheinlichkeit dieses *Durchschnitts*  $A \cap B$  wird bezogen (d.h. dividiert durch) die Wahrscheinlichkeit  $P(B)$  der Bedingung  $B$ . Hierbei wird für den Nenner  $P(B) > 0$  vorausgesetzt. – Wenn die Bedingung  $B$  das Ereignis  $A$  umfaßt (d.h.  $A \subset B$ ), so reduziert sich der Durchschnitt  $A \cap B$  zu  $A$ . Dies war im obigen Lotto-Beispiel der Fall, weil „mindestens 5 Richtige“ ( $B$ ) natürlich „6 Richtige“ ( $A$ ) umfassen.

Das Betrachten von auf  $B$  bedingten Wahrscheinlichkeiten läuft darauf hinaus, daß man den bisherigen Ergebnisraum  $\Omega$  auf die Teilmenge  $B$  einschränkt, und die ursprünglichen Wahrscheinlichkeiten durch Division von  $P(B)$  neu normiert.

Multipliziert man eine *bedingte* Wahrscheinlichkeit mit der Wahrscheinlichkeit ihrer *Bedingung*, so ergibt sich wieder eine *unbedingte* Wahrscheinlichkeit:

(2)

$$\begin{aligned} P(A | B) \cdot P(B) &= P(A \cap B) \text{ bzw.} \\ P\{X \in A | Y \in B\} \cdot P\{Y \in B\} &= P\{X \in A \text{ und } Y \in B\}. \end{aligned}$$

**Roulette-Spiel:** Für  $A = 2.$  Dutzend und  $B = 2.$  Kolonne (vgl. Abb. 4) ergibt sich (beim idealen Roulette)

$$P(A) = \frac{12}{37} = P(B) \quad P(A \cap B) = \frac{4}{37} \quad P(A | B) = \frac{1}{3} \neq P(A). \quad \square$$

### 2.1.4 Unabhängigkeit zweier Zufallsvariablen

#### Werfen mit zwei Würfeln:

Beim gleichzeitigen Werfen eines roten und eines grünen Würfels ergeben sich als mögliche Ergebnisse alle Kombinationen der sechs Zahlen des roten Würfels mit den sechs Zahlen des grünen Würfels. Die insgesamt 36 möglichen Ergebnisse für beide Würfel lassen sich durch Punkte in der Ebene darstellen (vgl. Abb. rechts). Man beachte, daß die beiden Ergebnisse

(1,6): roter Würfel = 1, grüner Würfel = 6

(6,1): roter Würfel = 6, grüner Würfel = 1  
nicht übereinstimmen.  $\square$

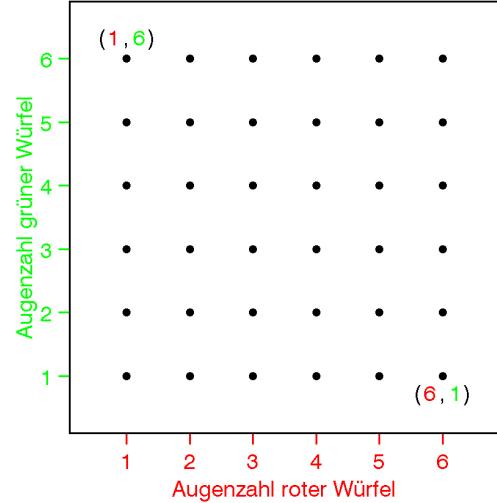


Abb. 5: Ergebnisse bei zwei Würfeln

Wir betrachten jetzt allgemein zwei voneinander unabhängige stochastische Vorgänge mit den Ergebnisräumen  $\Omega_1$  bzw.  $\Omega_2$  und fassen das simultane (gemeinsame, gleichzeitige) Ausführen beider Einzel-Vorgänge als einen neuen *simultanen Vorgang* auf. Die Ergebnisse des simultanen Vorgangs lassen sich als geordnete Paare  $(\omega_1, \omega_2)$  auffassen, wobei  $\omega_1 \in \Omega_1$  und  $\omega_2 \in \Omega_2$  Ergebnisse der Einzelvorgänge sind, d.h. der Ergebnisraum des *simultanen Vorgangs* ist der *Produktraum* (vgl. Abb. 6)

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1 \text{ und } \omega_2 \in \Omega_2\}. \quad (1)$$

Sind  $A_1 \subset \Omega_1$  bzw.  $A_2 \subset \Omega_2$  irgend zwei Einzel-Ereignisse des ersten bzw. zweiten Vorgangs, so beschreibt das *Produkt-Ereignis* (vgl. Abb. 6)

$$A = A_1 \times A_2 = \{(a_1, a_2) \mid a_1 \in A_1 \text{ und } a_2 \in A_2\} \quad (2)$$

das simultane Eintreten beider Einzel-Ereignisse  $A_1$  und  $A_2$  bei beiden Einzel-Vorgängen. Die Wahrscheinlichkeit für das simultane Eintreten beider Einzel-Ereignisse ist dann gleich

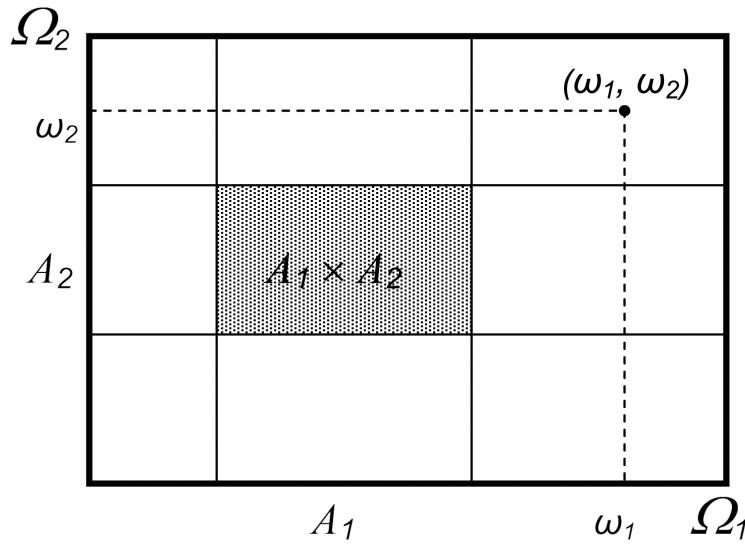


Abb. 6: Ergebnisraum aus (1) zweier simultaner Vorgänge und das simultane Ereignis aus (2).

dem *Produkt* beider Wahrscheinlichkeiten für das *einzelne* Eintreten der Ereignisse, d.h. es gilt die

*Produktregel für unabhängige Ereignisse  $A_1$  und  $A_2$ : Die Wahrscheinlichkeit für das Eintreten beider Ereignisse ist gleich dem Produkt beider Wahrscheinlichkeiten*

$$P(A_1 \times A_2) = P(A_1) \times P(A_2). \quad (3)$$

Ist  $X_1$  bzw.  $X_2$  eine interessierende Zufallsvariable des ersten bzw. zweiten stochastischen Vorgangs, so lautet die

*Produktregel für Ereignisse  $A_1$  und  $A_2$  unabhängiger Zufallsvariablen  $X_1$  und  $X_2$*

$$P\{X_1 \in A_1 \text{ und } X_2 \in A_2\} = P\{X_1 \in A_1\} \times P\{X_2 \in A_2\}. \quad (4)$$

Speziell ergibt sich für Ereignisse  $A_1 = \{a_1\}$  und  $A_2 = \{a_2\}$

$$P\{X_1 = a_1 \text{ und } X_2 = a_2\} = P\{X_1 = a_1\} \times P\{X_2 = a_2\}. \quad (5)$$

**Warnung:** Die Produktregeln (3), (4) und (5) gelten nur bei *unabhängigen stochastischen Vorgängen*. Genauer gesagt ist die (*stochastische*) *Unabhängigkeit* zweier Vorgänge bzw. Zufallsvariablen  $X_1$  und  $X_2$  dadurch *charakterisiert* (d.h. definiert), daß die Bedingung (3) bzw. (4) für *alle* Ereignisse  $A_1$  und  $A_2$  gilt. Ein Beispiel für *abhängige Zufallsvariablen* sind die Augenzahl  $X_1$  auf der *oberen* Seite eines Würfels und die gegenüberliegende Zahl  $X_2$  auf der *unteren* Seite, wobei  $X_1 + X_2 = 7$ . Hier gilt

$$P\{X_1 = 1 \text{ und } X_2 = 6\} = \frac{1}{6} \neq P\{X_1 = 1\} \times P\{X_2 = 6\} = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

Wenn  $X_1$  und  $X_2$  (stochastisch) unabhängig sind, so hängen die auf  $X_1$  bedingten Wahrscheinlichkeiten für  $X_2$  nicht mehr von der Bedingung ab, d.h. es gilt

$$P\{X_2 \in A_2 \mid X_1 \in A_1\} = P\{X_2 \in A_2\} \quad \text{falls } X_1 \text{ und } X_2 \text{ unabhängig sind.} \quad (6)$$

**Unabhängiges Ziehen aus zwei Urnen:** Bei einer Urne mit *weißen* und *schwarzen* Kugeln bezeichne  $p_w$  bzw.  $p_s$  die Wahrscheinlichkeit, daß eine zufällig gezogene Kugel weiß bzw. schwarz ist, z.B.  $p_w = 40\%$  bzw.  $p_s = 60\%$ . Und bei einer zweiten Urne mit *roten* und *blauen* Kugeln sei  $p_r$  bzw.  $p_b$  die Wahrscheinlichkeit für das Ziehen einer roten bzw. blauen Kugel, z.B.  $p_r = 30\%$  bzw.  $p_b = 70\%$ . Werden *unabhängig voneinander* je eine Kugel aus jeder Urne gezogen, so lassen sich die Wahrscheinlichkeiten für die vier möglichen Farbkombinationen der gezogenen Kugeln in einer Tabelle darstellen (in der auch die konkreten Wahrscheinlichkeiten für obige Zahlenbeispiele eingetragen sind):

Urne 1	Urne 2		insgesamt
	rot	blau	
weiß	$p_w \cdot p_r = 0,12$	$p_w \cdot p_b = 0,28$	$p_w = 0,40$
schwarz	$p_s \cdot p_r = 0,18$	$p_s \cdot p_b = 0,43$	$p_s = 0,60$
insgesamt	$p_s = 0,30$	$p_b = 0,70$	1

### 2.1.5 Unabhängigkeit mehrerer Zufallsvariablen

Wir wollen jetzt die Ausführungen des letzten Abschnitts von *zwei* auf eine *beliebige Anzahl  $K$  voneinander unabhängiger* stochastische Vorgänge verallgemeinern und das *simultane (gemeinsame, gleichzeitige)* Ausführen aller  $K$  Einzel-Vorgänge als einen neuen *simultanen Vorgang* auffassen. Aus den Ergebnisräumen  $\Omega_1, \Omega_2, \dots, \Omega_K$  der einzelnen Vorgänge ergibt sich der Ergebnisraum des simultanen Vorgangs als ein *Produktraum*

$$\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_K = \{(\omega_1, \omega_2, \dots, \omega_K) \mid \omega_1 \in \Omega_1, \omega_2 \in \Omega_2, \dots, \omega_K \in \Omega_K\}. \quad (1)$$

Für jedes  $k$  sei  $A_k \subset \Omega_k$  ein Ereignis des  $k$ -ten Vorgangs. Dann beschreibt das *Produkt-Ereignis*

$$A = A_1 \times A_2 \times \dots \times A_K = \{(a_1, a_2, \dots, a_K) \mid a_1 \in A_1, a_2 \in A_2, \dots, a_K \in A_K\} \quad (2)$$

das *simultane* Eintreten aller Einzel-Ereignisse  $A_k$  bei allen Einzel-Vorgängen.

Die Wahrscheinlichkeit für das *simultane* Eintreten aller Einzel-Ereignisse ergibt sich wieder mit der

*Produktregel für unabhängige Ereignisse:* Die Wahrscheinlichkeit für das simultane Eintreten aller Ereignisse ist gleich dem Produkt aller Wahrscheinlichkeiten für das einzelne Eintreten

$$P(A_1 \times A_2 \times \dots \times A_K) = P(A_1) \times P(A_2) \times \dots \times P(A_K). \quad (3)$$

Ist für jedes  $k$  eine interessierende Zufallsvariable  $X_k$  des  $k$ -ten stochastischen Vorgangs gegeben, so gilt entsprechend die

*Produktregel für Ereignisse unabhängiger Zufallsvariablen  $X_1, X_2, \dots, X_K$ :*

$$\begin{aligned} P\{X_1 \in A_1 \text{ und } X_2 \in A_2 \text{ und } \dots \text{ und } X_K \in A_K\} = \\ P\{X_1 \in A_1\} \times P\{X_2 \in A_2\} \times \dots \times P\{X_K \in A_K\}. \end{aligned} \quad (4)$$

Speziell ergibt sich für Ereignisse  $A_k = \{a_k\}$  für alle  $k$

$$\begin{aligned} P\{X_1 = a_1 \text{ und } X_2 = a_2 \text{ und } \dots \text{ und } X_K = a_K\} = \\ P\{X_1 = a_1\} \times P\{X_2 = a_2\} \times \dots \times P\{X_K = a_K\}. \end{aligned} \quad (5)$$

**Werfen mit 3 Würfeln:** Beim gleichzeitigen Werfen von drei homogenen Würfeln sind ihre Augenzahlen  $X_1, X_2$  und  $X_3$  unabhängig und daher gilt z. B.

$$P\{X_1 = 6, X_2 = 6, X_3 = 6\} = P\{X_1 = 6\} \times P\{X_2 = 6\} \times P\{X_3 = 6\} = \left(\frac{1}{6}\right)^3 \approx 0,46\%.$$

**Warnung:** Die Produktregeln (3), (4) und (5) gelten *nur bei unabhängigen stochastischen Vorgängen*. Genauer gesagt ist die (*stochastische*) *Unabhängigkeit* mehrerer Vorgänge bzw. Zufallsvariablen dadurch *charakterisiert* (d.h. definiert), daß die Bedingung (3) bzw. (4) für *alle* Ereignisse  $A_1, A_2, \dots, A_K$  gilt.

### **Spezialfall: Unabhängige Wiederholungen**

In der Statistik spielen *unabhängige Wiederholungen desselben* stochastischen Vorgangs (unter gleichen Bedingungen) eine wichtige Rolle, z.B. die Befragung einer zufällig ausgewählten Person bei einer Umfrage. Ist  $X$  eine interessierende Zufallsvariable dieses stochastischen Vorgangs (z.B. das Alter der befragten Person), so sind die  $K$  Wiederholungen von  $X$  *unabhängige* Zufallsvariablen  $X_1, X_2, \dots, X_K$  und die Produktregeln (4) und (5) gelten.

## 2.2 Diskrete Zufallsvariablen

Bei der Modellierung und Berechnung von Wahrscheinlichkeiten für Zufallsvariablen ergeben sich unterschiedliche Aspekte, je nachdem ob es sich um *diskrete* oder *stetige* Zufallsvariablen handelt. Wir beginnen zunächst mit den in vielerlei Hinsicht einfacheren *diskreten* Variablen.

Ausgangspunkt ist eine einzelne *diskrete* Zufallsvariable  $X$  die nur isolierte Werte aus einer *Menge von Zahlen*, dem sogenannten *Träger*  $T = \{x_1, x_2, \dots\}$  annimmt. Die konkreten Ergebnisse für  $X$ , also die Elemente von  $T$ , werden auch als *Realisierungen* von  $X$  bezeichnet. Typischerweise (aber nicht immer) sind die Realisierungen der hier betrachteten Zufallsvariablen Elemente aus der Menge  $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$  aller *nicht-negativen ganzen* Zahlen. Wir werden auch Beispiele kennenlernen, wo der Träger mit der Menge  $\mathbb{N}_0$  übereinstimmt und somit *unendlich* viele Realisierungen enthält.

Man muß streng unterscheiden zwischen der *Zufallsvariablen*  $X$  und einer beobachteten Realisierung  $x \in T$ . So ist z.B. die *Augensumme*  $X$  zweier Würfel eine *Rechenvorschrift* (beide Augenzahlen zu addieren), während die Realisierung  $x$  bei einem konkreten Wurf eine *Zahl* ist, etwa  $x = 7$ . Üblicherweise werden Zufallsvariablen mit Großbuchstaben (z.B.  $X$ ) und ihre Realisierungen mit dem entsprechenden Kleinbuchstaben (z.B.  $x$  oder  $x_1, x_2$  etc.) bezeichnet.

Die Wahrscheinlichkeit für eine Realisierung  $x \in T$  von  $X$

$$P\{X = x\} \quad (\text{Wahrscheinlichkeit für das Ergebnis } x) \quad (1)$$

wird auch als *Ergebnis-Wahrscheinlichkeit* oder *Elementar-Wahrscheinlichkeit* (für  $x$ ) bezeichnet. Die Wahrscheinlichkeit für ein *beliebiges Ereignis*  $A = \{a_1, a_2, \dots\}$  ergibt sich dann aus den Elementar-Wahrscheinlichkeiten mit der

*Summenregel für eine diskrete Zufallsvariable  $X$ : Die Wahrscheinlichkeit eines Ereignisses ist die Summe der zugehörigen Ergebnis-Wahrscheinlichkeiten*

$$\begin{aligned} P\{X \in A\} &= \sum_{x \in A} P\{X = x\} \\ &= P\{X = a_1\} + P\{X = a_2\} + \dots \quad \text{für } A = \{a_1, a_2, \dots\}. \end{aligned} \quad (2)$$

Die Zuordnung  $A \mapsto P\{X \in A\}$ , die jedem Ereignis  $A$  die Wahrscheinlichkeit  $P\{X \in A\}$  zuordnet, wird als *Verteilung von  $X$*  bezeichnet. Nach der Summenregel (2) ist die Verteilung durch alle Ergebnis-Wahrscheinlichkeiten der Form (1) bereits vollständig festgelegt, d.h. wenn man die Wahrscheinlichkeiten für jede Realisierung  $x$  von  $X$  kennt, so läßt sich daraus jede Wahrscheinlichkeit für  $X$  berechnen.

Um eine graphische Übersicht der Verteilung zu erhalten, stellt man alle Ergebniswahrscheinlichkeiten (analog zu beobachteten relativen Häufigkeiten) in einem *Histogramm (Säulen-Diagramm)* dar, wobei die jeweilige *Fläche* der Wahrscheinlichkeit entspricht (vgl. Abb. 1 links). Eine weitere Darstellungsart ist das *Stabdiagramm*, bei dem die jeweilige *Länge* die Wahrscheinlichkeit repräsentiert (vgl. Abb. 1 rechts).

**Augensumme zweier Würfel:** Beim Werfen eines *roten* und eines *grünen* Würfels tritt bei jedem der beiden (homogenen) Würfel jede Augenzahl mit Wahrscheinlichkeit  $\frac{1}{6}$  auf. Jede der 36 verschiedenen Kombinationen beider Würfelergebnisse (vgl. 2.1.4 Abb. 5) tritt nach der Produktregel (für unabhängige Zufallsvariablen) mit gleicher Wahrscheinlichkeit  $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$  auf. Die Augensumme  $X$  beider Würfel kann nur die Werte aus dem Träger  $T = \{2, 3, \dots, 12\}$  annehmen. Die Verteilung von  $X$  ist in Abb. 1 dargestellt.

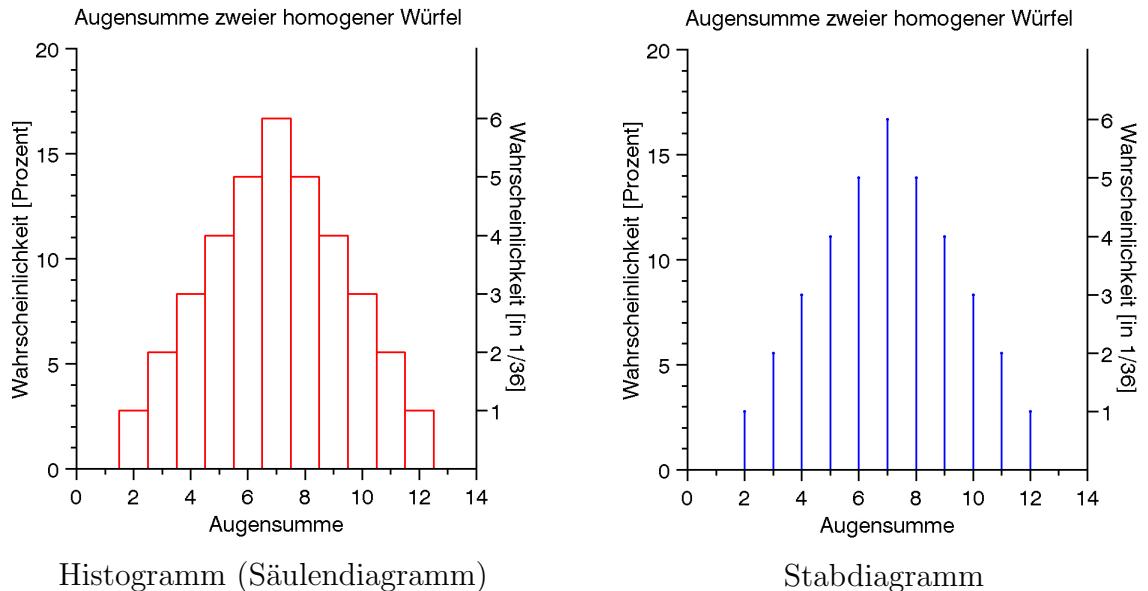


Abb. 1: Grafische Darstellung der Verteilung der Augensumme zweier homogener Würfel.

### 2.2.1 Die diskrete Gleichverteilung (*Laplace-Verteilung*)

Wenn die  $N$  verschiedenen Realisierungen aus  $T = \{x_1, \dots, x_N\}$  einer Zufallsvariablen  $X$  alle *gleichwahrscheinlich* sind, so heißt  $X$  *diskret gleichverteilt* oder *Laplace-verteilt* auf dem Träger  $T$ . Die Ergebnis-Wahrscheinlichkeiten sind dann alle gleich (vgl. Abb. 2)

$$P\{X = x_i\} = \frac{1}{N} \quad \text{für alle } i = 1, \dots, N. \quad (\text{diskrete Gleichverteilung von } X). \quad (1)$$

Die Wahrscheinlichkeit eines beliebigen Ereignisses  $A$  ist der Quotient der Anzahl  $\#A$  (*lies: Anzahl der Elemente von  $A$* ) aller Ergebnisse aus  $A$  und der Anzahl  $N = \#T$  aller Ergebnisse:

$$P\{X \in A\} = \frac{\text{Anzahl der Ergebnisse aus } A}{\text{Anzahl aller Ergebnisse}} = \frac{\#A}{\#T} = \frac{\#A}{N} \quad (2)$$

(für gleichverteiltes  $X$ ).

**Warnung:** Die Darstellungen (1) und (2) sind *nicht allgemein gültig*, sondern gelten nur, wenn  $X$  *gleichverteilt* ist.

**Auftreten der Gleichverteilung:** Gleichverteilungen treten bei Anwendungen extrem selten auf, typischerweise nur im Zusammenhang mit reinen *Glücksspielen* wie z.B.

- die Augenzahl  $X$  eines homogenen Würfels mit  $N = 6$  (vgl. Abb. 2),
- die Gewinnzahl beim idealen Roulette mit  $N = 37$  (vgl. Abb. 2),
- die *erste* gezogene Zahl beim Zahlen-Lotto „6 aus 49“ mit  $N = 49$ .

Dagegen ist die Augensumme zweier (homogener) Würfeln bereits nicht mehr gleichverteilt (vgl. Abb. 1). Auch biologische Merkmale wie z.B. die *ABO*-Blutgruppen ( $N = 4$ ) oder Rhesus-Faktor ( $N = 2$ ) einer zufällig ausgewählten Person sind *nicht gleichverteilt*. Auch das Geschlecht ( $N = 2$ ) eines zufälligen Neugeborenen ist *nicht gleichverteilt* (die Wahrscheinlichkeit für Jungen ist größer als  $\frac{1}{2}$ ).

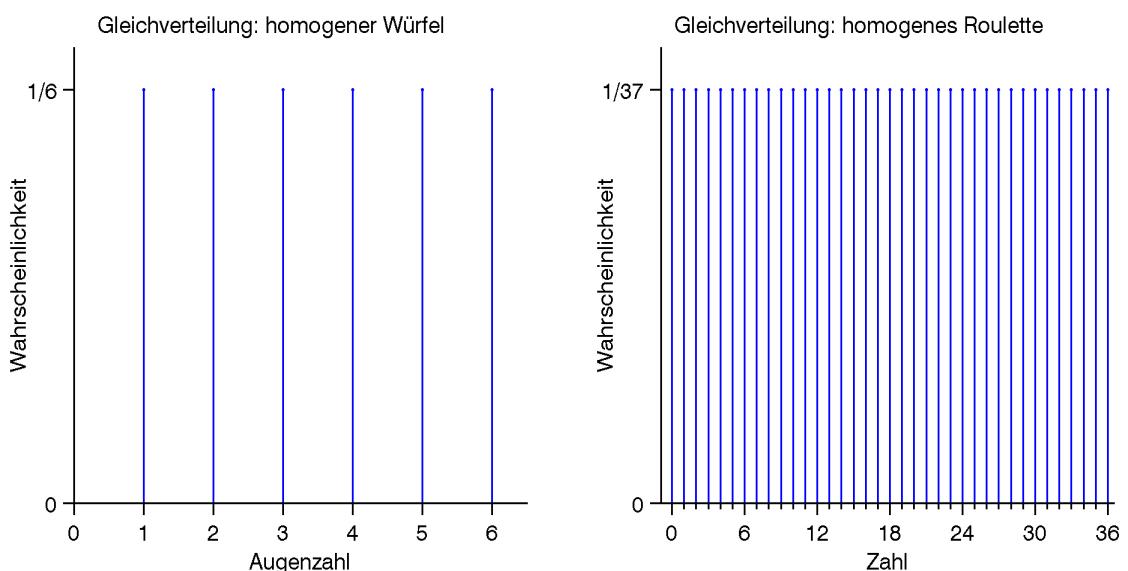


Abb. 2: Stabdiagramme zweier Gleichverteilungen mit  $N = 6$  (links) und  $N = 37$  (rechts).

## 2.3 Die Binomial-Verteilung

Die Binomial-Verteilung ist eine der wichtigsten diskreten Verteilungen. Sie tritt immer dann auf, wenn man sich bei einem stochastischen Vorgang nur dafür interessiert ob ein spezielles *Ziel-Ereignis A* (oft als *Erfolg* bezeichnet) eingetreten ist oder nicht, z.B. das Ereignis *1. Dutzend* beim Roulette (wenn man darauf gesetzt hat) oder der *Heilungserfolg* bei einer speziellen Therapie. Da wir uns bei dem stochastischen Vorgang (Zufallsexperiment) nur dafür interessieren, ob das Ziel-Ereignis *A eintritt* (*Erfolg*) oder *nicht* (*Mißerfolg*), reduzieren wir das Ergebnis unter Verwendung der *Indikator-Variablen Y* für das Ereignis *A* auf die Zahlen 1 und 0, d.h.

$$Y = \begin{cases} 1 & \text{falls Ziel-Ereignis } A \text{ eintritt: } (\text{Erfolg}) \\ 0 & \text{falls Ziel-Ereignis } A \text{ nicht eintritt: } (\text{Mißerfolg}). \end{cases} \quad (1)$$

Die zugehörigen Wahrscheinlichkeiten bezeichnen wir mit

$$p = P\{Y = 1\} = P(A) \quad \text{Eintritts-Wahrscheinlichkeit} \quad (2)$$

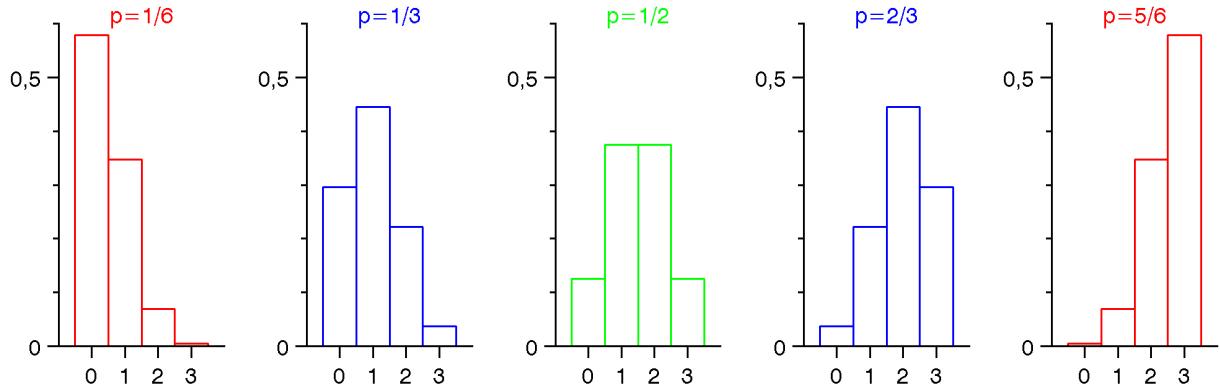
$$q = P\{Y = 0\} = P(\complement A) = 1 - p \quad (\text{Nicht-Eintritts-Wahrscheinlichkeit}). \quad (3)$$

Wird nun der Vorgang  $n$  mal (unabhängig) wiederholt, so ist die Anzahl  $X$  der Eintritte des Ziel-Ereignisses eine Zufallsvariable mit den möglichen Realisierungen  $0, 1, \dots, n$ . Die Verteilung von  $X$  ist eine *Binomial-Verteilung*. Bevor wir diese Binomial-Verteilung für *beliebiges n* betrachten, wollen wir uns den Spezialfall mit  $n = 3$  Wiederholungen ansehen.

Bei drei unabhängigen Wiederholungen seien jetzt  $Y_1, Y_2, Y_3$  die zugehörigen Indikatoren für das Eintreten von *A*. Die (absolute) Häufigkeit mit der das Ereignis *A* unter den 3 Wiederholungen aufgetreten ist, ist dann gegeben durch die *Summe der Indikatoren*  $X = Y_1 + Y_2 + Y_3$ . Die Berechnung der Wahrscheinlichkeit  $P\{X = x\}$  für  $x \in \{0, 1, 2, 3\}$  erfolgt nun (vgl. Tab. 1 und Abb. 1) indem man die Wahrscheinlichkeiten  $P\{Y_1 = y_1, Y_2 = y_2, Y_3 = y_3\}$  für alle 8 Kombinationen von  $y_1, y_2, y_3 \in \{0, 1\}$  bestimmt und dann diejenigen aufaddiert, bei denen  $y_1 + y_2 + y_3 = x$  ist.

$y_1$	$y_2$	$y_3$	$P\{Y_1 = y_1, Y_2 = y_2, Y_3 = y_3\}$	$x = y_1 + y_2 + y_3$	$P\{X = x\}$
0	0	0	$q \cdot q \cdot q$	0	$q^3$
1	0	0	$p \cdot q \cdot q$	1	$3p q^2$
0	1	0	$q \cdot p \cdot q$		
0	0	1	$q \cdot q \cdot p$	2	$3q^2 p$
.....	.....	.....	.....		
1	1	0	$p \cdot p \cdot q$	3	$q^3$
1	0	1	$p \cdot q \cdot p$		
0	1	1	$q \cdot p \cdot p$	.....	.....
.....	.....	.....	.....	.....	.....
1	1	1	$p \cdot p \cdot p$	3	$q^3$

Tabelle 1: Die Verteilung der Indikatoren  $Y_1, Y_2, Y_3$  und deren Summe  $X = Y_1 + Y_2 + Y_3$ .



**Abb. 1:** Histogramme der  $B(3, p)$  Verteilung nach Tab. 1 für verschiedene  $p = P(A)$ , die z.B. folgende Ereignisse beim homogenen Würfel repräsentieren:

$$P\{6\} = \frac{1}{6}, \quad P\{5, 6\} = \frac{1}{3}, \quad P\{4, 5, 6\} = \frac{1}{2}, \quad P\{3, \dots, 6\} = \frac{2}{3}, \quad P\{2, \dots, 6\} = \frac{5}{6}.$$

### 2.3.1 Definition der Binomial-Verteilung

Wir betrachten jetzt *eine beliebige Anzahl  $n$  von unabhängigen Wiederholungen* des Vorgangs (Experiments) und  $Y_i$  bezeichne die Indikator-Variable für das Eintreten von  $A$  bei der  $i$ -ten Wiederholung. Die (absolute) Häufigkeit mit der das Ziel-Ereignis  $A$  unter allen  $n$  Wiederholungen eingetreten ist, ist genau die Summe der Indikatoren

$$X = Y_1 + Y_2 + \dots + Y_n \quad (\text{Anzahl der Eintritte}) . \quad (1)$$

Diese Anzahl  $X$  ist eine *Zufallsvariable* mit Werten im Träger  $\{0, 1, 2, \dots, n\}$ . Die Elementar-Wahrscheinlichkeiten  $P\{X = x\}$  lassen sich (analog Tabelle 1) bestimmen zu

$$\begin{aligned} P\{X = x\} &= \binom{n}{x} p^x q^{n-x}, \quad \text{für } x = 0, 1, \dots, n \quad \text{mit} \quad q = 1 - p, \\ &= b(x \mid n, p) \quad (\text{Abkürzung}) \quad (\text{Binomial-Verteilung}) \end{aligned} \quad (2)$$

Hierbei entspricht der *Binomialkoeffizient*  $\binom{n}{x}$  der Anzahl aller Möglichkeiten, die  $x$  Eintritte auf die  $n$  Wiederholungen zu verteilen (vgl. 2.3.4). Und das Produkt  $p^x q^{n-x}$  ergibt sich daraus, daß das Ereignis  $A$  mit Wahrscheinlichkeit  $p$  genau  $x$ -mal und das komplementäre Ereignis  $\complement A$  mit Wahrscheinlichkeit  $q$  genau  $(n - x)$ -mal (unabhängig voneinander) eingetreten sind.

Die durch (2) gegebene Verteilung von  $X$  auf  $\{0, 1, 2, \dots, n\}$  heißt *Binomial-Verteilung vom Umfang  $n$  mit Eintrittswahrscheinlichkeit  $p$* , kurz:  $B(n, p)$ -Verteilung. Sie ist für eine jede *natürliche Zahl  $n$*  und jede Wahrscheinlichkeit  $0 < p < 1$  definiert. Die Grenzfälle  $p = 0$  bzw.  $p = 1$  schließt man hier aus, weil unter ihnen  $X$  nur jeweils *einen* Wert mit Wahrscheinlichkeit 1 annehmen könnte, und zwar den Wert 0 für  $p = 0$  bzw. den Wert  $n$  für  $p = 1$ .

Die Elementar-Wahrscheinlichkeiten lassen sich *schrittweise vorwärts* (d.h. für  $x = 0, 1, \dots, n$ ) oder *rückwärts* (d.h. für  $x = n, n-1, \dots, 0$ ) berechnen mit

$$P\{X = 0\} = q^n \quad (3)$$

$$P\{X = x\} = \frac{n-x+1}{x} \cdot \frac{p}{q} \cdot P\{X = x-1\} \quad \text{für } x = 1, \dots, n \quad (4)$$

$$P\{X = n\} = p^n. \quad (5)$$

Die Wahrscheinlichkeiten  $P\{X = x\}$  sind für  $x = 0, 1, \dots$  zuerst *anwachsend* – solange  $x \leq np$  gilt – und dann für  $x \geq np$  wieder *abfallend* (vgl. Abb. 1 und Tab. 2-3).

Die Anzahl  $n - X$  der *Nicht-Eintritte* von  $A$  ist  $B(n, q)$ -verteilt.

Die Binomial-Verteilung  $B(1, p)$  vom Umfang 1 mit dem Träger  $\{0, 1\}$  wird auch als *Bernoulli-Verteilung* bezeichnet. Die Indikatorvariable  $Y$  ist  $B(1, p)$ -verteilt.

### 2.3.2 Binomial-Wahrscheinlichkeiten

Bei einer  $B(n, p)$ -verteilten Zufallsvariablen  $X$  sind neben den Elementar-Wahrscheinlichkeiten  $P\{X = x\}$  folgende *kumulierte Wahrscheinlichkeiten* von Interesse.

Wahrscheinlichkeit für *höchstens*  $x$  Eintritte:

$$\begin{aligned} P\{X \leq x\} &= P\{X = 0\} + P\{X = 1\} + \dots + P\{X = x\} \\ &= \sum_{k=0}^x P\{X = k\} \end{aligned} \quad (1)$$

Wahrscheinlichkeit für *mindestens*  $x$  Eintritte:

$$\begin{aligned} P\{X \geq x\} &= P\{X = x\} + P\{X = x+1\} + \dots + P\{X = n\} \\ &= \sum_{k=x}^n P\{X = k\} \end{aligned} \quad (2)$$

Wahrscheinlichkeit für *mindestens*  $x_1$  und *höchstens*  $x_2$  Eintritte:

$$\begin{aligned} P\{x_1 \leq X \leq x_2\} &= P\{X = x_1\} + P\{X = x_1 + 1\} + \dots + P\{X = x_2\} \\ &= \sum_{k=x_1}^{x_2} P\{X = k\} \end{aligned} \quad (3)$$

Die Wahrscheinlichkeiten der Form (2) und (3) lassen sich auch unter Verwendung der Wahrscheinlichkeiten  $P\{X \leq x\}$  wie folgt bestimmen

$$P\{X \geq x\} = 1 - P\{X \leq x - 1\}, \quad (4)$$

$$P\{x_1 \leq X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1 - 1\}. \quad (5)$$

Faßt man die Wahrscheinlichkeit  $P\{X \leq x\}$  als eine Funktion von  $x$  auf, so erhält man die sogenannte *Verteilungsfunktion*  $F$  von  $X$

$$F(x) = P\{X \leq x\} \quad (\text{Verteilungsfunktion von } X). \quad (6)$$

Mit der Verteilungsfunktion  $F$  lassen sich alle Wahrscheinlichkeiten von  $X$  berechnen. Die Elementar-Wahrscheinlichkeiten lassen sich z.B. nach (5) wie folgt durch die Verteilungsfunktion darstellen:

$$P\{X = x\} = P\{x \leq X \leq x\} = F(x) - F(x - 1). \quad (7)$$

Die Verteilungsfunktion der Binomial-Verteilung, d.h. die kumulierten Wahrscheinlichkeiten  $F(x) = P\{X \leq x\}$  sind in Tabellenkalkulations- und Statistik-Programmen für Computer bereits implementiert. Wenn man sie ohne solche Programme ermitteln will, so berechnet man sie am besten *schrittweise* für  $x = 0, \dots, n$  zusammen mit den Elementar-Wahrscheinlichkeiten  $P\{X = x\}$  aus **2.3.1 (3)(4)** und

$$P\{X \leq 0\} = P\{X = 0\}, \quad (8)$$

$$P\{X \leq x\} = P\{X \leq x - 1\} + P\{X = x\} \quad \text{für } x = 1, \dots, n. \quad (9)$$

Analog kann man die kumulierten Wahrscheinlichkeiten  $P\{X \leq x\}$  *schrittweise absteigend* für  $x = n, \dots, 0$  bestimmen aus

$$P\{X \geq n\} = P\{X = n\}, \quad (10)$$

$$P\{X \geq x\} = P\{X \geq x + 1\} + P\{X = x\} \quad \text{für } x = n - 1, \dots, 0. \quad (11)$$

### 2.3.3 Anwendung: Wahlumfragen

Um den unbekannten Stimmanteil  $p$  einer bestimmten Partei zu schätzen, werden  $n$  unabhängige Befragungen von Wahlberechtigten durchgeführt. Da es sich um *unabhängige Wiederholungen* einer Einzelbefragung handeln soll, ist die Anzahl  $X$  der Befürworter dieser Partei unter den  $n$  Befragten dann  $B(n, p)$ -verteilt. Man beachte, daß hierbei eine bereits befragte Person erneut ausgewählt und befragt werden kann (die Wahrscheinlichkeit hierfür ist natürlich extrem gering). – Wir betrachten jetzt eine „Blitzumfrage“ mit nur  $n = 100$  telefonischen Befragungen.

**Umfrage für eine kleine Partei:** Für eine kleine Partei (z.B. *Die Grünen*) mit Stimmanteil  $p = 8\%$  sind die Elementar- und kumulierten Wahrscheinlichkeiten in Tabelle 2 zusammengestellt. Die dort fehlenden Werte  $x \geq 21$  sind uninteressant, weil für diese die Elementar-Wahrscheinlichkeiten im Rahmen der hier relevanten Genauigkeit zu gering sind. Die Wahrscheinlichkeit, daß *in der Umfrage* die Partei *höchstens* 5% erhält (d.h.  $X \leq 5$  ist), obwohl der *wahre Stimmanteil*  $p = 8\%$  beträgt, ergibt sich aus der Tabelle 2 zu  $P\{X \leq 5\} = 17,99\%$  (dies entspricht der markierten Fläche im Histogramm). Folglich ist die Wahrscheinlichkeit nicht gerade gering, daß die Umfrage ein Scheitern an der 5%-Hürde prognostiziert, obwohl der wahre Stimmanteil mit  $p = 8\%$  deutlich darüber liegt. Wie wir später noch sehen werden, liegt dies in erster Linie daran, daß die Anzahl  $n = 100$  nicht groß genug ist, um zuverlässige Schlüsse ziehen zu können.

$x$	$P\{X \leq x\}$	$P\{X \geq x\}$	$P\{X = x\}$	Histogramm
0	0,02%	100,00%	<b>0,02%</b>	
1	0,23%	99,98%	<b>0,21%</b>	
2	1,13%	99,77%	<b>0,90%</b>	
3	3,67%	98,87%	<b>2,54%</b>	
4	9,03%	96,33%	<b>5,36%</b>	
5	<b>17,99%</b>	90,97%	<b>8,95%</b>	
6	30,32%	82,01%	12,33%	
7	44,71%	69,68%	14,40%	
8	59,26%	55,29%	14,55%	
9	72,20%	40,74%	12,93%	
10	82,43%	27,80%	10,24%	
11	89,72%	17,57%	7,28%	
12	94,41%	10,28%	4,70%	
13	97,18%	5,59%	2,76%	
14	98,67%	2,82%	1,49%	
15	99,42%	1,33%	0,74%	
16	99,76%	0,58%	0,34%	
17	99,91%	0,24%	0,15%	
18	99,97%	0,09%	0,06%	
19	99,99%	0,03%	0,02%	
20	100,00%	0,01%	0,01%	

Tabelle 2: Wahrscheinlichkeiten einer  $B(n, p)$ -verteilten Zufallsvariablen  $X$  mit  $n = 100$  und  $p = 8\%$ . Die markierte Fläche entspricht der Wahrscheinlichkeit  $P\{X \leq 5\}$ .

**Umfrage für eine große Partei:** Für eine große Partei (z.B. die CDU) mit einem Stimmanteil  $p = 40\%$  sind die Elementar- und kumulierten Wahrscheinlichkeiten für den interessierenden Bereich in Tabelle 3 angegeben. Die Wahrscheinlichkeit, daß *in der Umfrage* die Partei *mindestens* 50% erhält (d.h.  $X \geq 50$  ist), obwohl der *wahre Stimmanteil* nur  $p = 40\%$  beträgt, ergibt sich aus Tabelle 3 zu  $P\{X \geq 50\} = 2,71\%$  (dies entspricht markierter Fläche im Histogramm) und ist relativ gering.

X	$P\{X \leq x\}$	$P\{X \geq x\}$	$P\{X=x\}$	Histogramm
22	0,01%	100,00%	0,01%	
23	0,03%	99,99%	0,01%	
24	0,06%	99,97%	0,03%	
25	0,12%	99,94%	0,06%	
26	0,24%	99,88%	0,12%	
27	0,46%	99,76%	0,22%	
28	0,84%	99,54%	0,38%	
29	1,48%	99,16%	0,63%	
30	2,48%	98,52%	1,00%	
31	3,98%	97,52%	1,51%	
32	6,15%	96,02%	2,17%	
33	9,13%	93,85%	2,97%	
34	13,03%	90,87%	3,91%	
35	17,95%	86,97%	4,91%	
36	23,86%	82,05%	5,91%	
37	30,68%	76,14%	6,82%	
38	38,22%	69,32%	7,54%	
39	46,21%	61,78%	7,99%	
40	54,33%	53,79%	8,12%	
41	62,25%	45,67%	7,92%	
42	69,67%	37,75%	7,42%	
43	76,35%	30,33%	6,67%	
44	82,11%	23,65%	5,76%	
45	86,89%	17,89%	4,78%	
46	90,70%	13,11%	3,81%	
47	93,62%	9,30%	2,92%	
48	95,77%	6,38%	2,15%	
49	97,29%	4,23%	1,52%	
50	98,32%	<b>2,71%</b>	<b>1,03%</b>	
51	99,00%	1,68%	<b>0,68%</b>	
52	99,42%	1,00%	<b>0,42%</b>	
53	99,68%	0,58%	<b>0,26%</b>	
54	99,83%	0,32%	<b>0,15%</b>	
55	99,91%	0,17%	<b>0,08%</b>	
56	99,96%	0,09%	<b>0,04%</b>	
57	99,98%	0,04%	<b>0,02%</b>	
58	99,99%	0,02%	<b>0,01%</b>	
59	100,00%	0,01%	<b>0,01%</b>	

Tabelle 3: Wahrscheinlichkeiten einer  $B(n, p)$ -verteilten Zufallsvariablen  $X$  mit  $n = 100$  und  $p = 40\%$ . Die markierte Fläche entspricht der Wahrscheinlichkeit  $P\{X \geq 50\}$ .

 **Tipps für Microsoft Excel Funktionen:** Für  $B(n, p)$ -verteiltes  $X$  stehen in *Excel* folgende *Statistik-Funktionen* direkt zur Verfügung:

$$\text{BINOMVERT}(x; n; p; 0) = P\{X = x\} = b(x | n, p)$$

$$\text{BINOMVERT}(x; n; p; 1) = P\{X \leq x\} = \sum_{k=0}^x b(k | n, p)$$

Das *letzte* Argument der Funktion **BINOMVERT** ist der Indikator dafür, ob es sich um kumulierte Wahrscheinlichkeiten handelt (1) oder nicht (0).

Da die Anzahl  $n - X$  der *Nicht-Eintritte*  $B(n, q)$ -verteilt ist mit  $q = 1 - p$ , lässt sich die Wahrscheinlichkeit  $P\{X \geq x\} = P\{n - X \leq n - x\}$  auch wie folgt berechnen

$$\text{BINOMVERT}(n - x; n; q; 1) = P\{X \geq x\} = \sum_{k=x}^n b(k | n, q), \quad q = 1 - p.$$

### 2.3.4 Fakultät und Binomialkoeffizient

**Fakultät:** Die *Fakultät* einer natürlichen Zahl  $n$  ist das Produkt aller natürlichen Zahlen von 1 bis  $n$ :

$$n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n \quad (\text{lies: } n \text{ Fakultät}) \quad (1)$$

*Bedeutung:*  $n!$  ist die Anzahl aller verschiedenen Permutationen (Anordnungen) der Zahlen von 1 bis  $n$ . Für  $n = 3$  gibt es z.B.  $3! = 6$  verschiedene Anordnungen:

$$123, \quad 132, \quad 213, \quad 231, \quad 312, \quad 321.$$

Für  $n = 0$  ist *definiert*:

$$0! = 1 \quad (\text{Definition, Konvention}) \quad (2)$$

Die Fakultäten lassen sich *schrittweise* berechnen aus

$$n! = (n - 1)! \cdot n \quad \text{für } n = 1, 2, 3, \dots \quad (3)$$

**Binomialkoeffizient:** Für *ganze* Zahlen  $0 \leq k \leq n$  ist der *Binomialkoeffizient* definiert als

$$\binom{n}{k} = \frac{n!}{k! \cdot (n - k)!} \quad (\text{lies: } n \text{ über } k) \quad (4)$$

Zur Berechnung des Binomialkoeffizienten ist folgende Darstellung günstiger

$$\binom{n}{k} = \frac{n}{1} \cdot \frac{(n-1)}{2} \cdot \frac{(n-2)}{3} \cdots \frac{(n-k+1)}{k}. \quad (5)$$

*Bedeutung:* Der Binomialkoeffizient  $n$  über  $k$  ist die Anzahl aller verschiedenen Auswahlen (Ziehungen) von  $k$  verschiedenen Zahlen aus den Zahlen von 1 bis  $n$ .

*Beispiel:* Die Anzahl der möglichen Ziehungen beim Lotto „6 aus 49“ ist daher  $\binom{49}{6} \approx 14 \cdot 10^6$ .

Unter Verwendung der Beziehung

$$\binom{n}{k} = \binom{n}{n-k} \quad (6)$$

lässt sich die Berechnung nach (5) von  $k$  auf den Fall  $n - k$  zurückführen, was speziell für  $k \geq \frac{n}{2}$  vorteilhaft ist, weil dann  $n - k \leq k$  ist und somit weniger Faktoren in (5) zu multiplizieren sind.

Für festes  $n$  kann man alle Binomialkoeffizienten *schrittweise* für  $k = 0, 1, \dots, n$  (oder auch rückwärts) berechnen mit

$$\binom{n}{0} = \binom{n}{n} = 1, \quad (7)$$

$$\binom{n}{k} = \binom{n}{k-1} \cdot \frac{(n-k+1)}{k} \quad \text{für } k \geq 1. \quad (8)$$

### Tipps für Microsoft Excel

**Funktionen:** Folgende Funktionen stehen in Excel direkt zur Verfügung:

$$\begin{aligned} \text{FAKULTÄT}(n) &= n! \\ \text{KOMBINATIONEN}(n, k) &= \binom{n}{k}. \end{aligned}$$

## 2.4 Stetige Zufallsvariablen mit Dichten

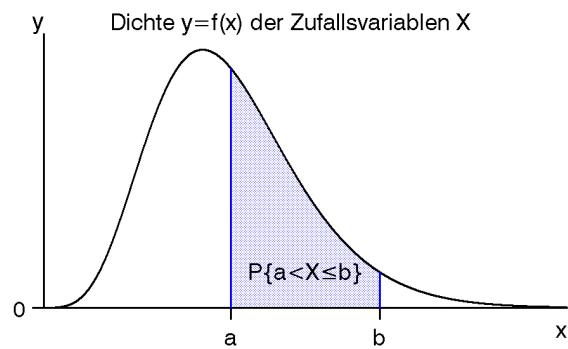
Wir betrachten jetzt eine *stetige* Zufallsvariable  $X$ , die kontinuierlich alle Werte aus einem Intervall  $T$  reeller Zahlen (d.h. einem endlichen oder unendlichen Abschnitt auf der Zahlengeraden) annehmen kann. Der *Träger*  $T$  von  $X$  ist meistens die Menge  $\mathbb{R}$  aller reeller Zahlen oder das Intervall  $(0, \infty) = \{x \in \mathbb{R} \mid x > 0\}$  der *positiven* reellen Zahlen. Bei konkreten Anwendungen ist dies oft nur eine *Idealisierung* (man spricht auch von *Modellierung*), weil z.B. das Geburtsgewicht  $X$  einerseits nicht beliebig große Werte und andererseits als Folge der Meßungenauigkeit auch nicht *jeden* Wert innerhalb eines Intervalls (z.B. von 2 bis 3 kg) annehmen kann. Wie sich noch zeigen wird, sind trotz dieser Einschränkungen solche idealisierten Modelle extrem nützlich bei praktischen Anwendungen.

Bei einer stetigen Zufallsvariablen  $X$  lassen sich den einzelnen *Realisierungen*  $x$  aus dem Träger  $T$  keine positiven Wahrscheinlichkeiten mehr zuordnen. Dies liegt grob gesprochen daran, daß es zu *viele* (genauer: *überabzählbar viele*) solcher Ergebnisse  $x$  im Intervall  $T$  gibt und die Summe aller zugehörigen Wahrscheinlichkeiten daher nicht gleich 1 sein kann (sondern *unendlich* wäre). Aus diesem Grunde werden wir nur Wahrscheinlichkeiten für *Intervalle* betrachten, also z.B.  $P\{a < X \leq b\}$ . Zur Bestimmung solcher Intervall-Wahrscheinlichkeiten benutzt man eine sogenannte *Dichte-Funktion* (kurz: *Dichte*). Diese Dichte ist interpretierbar als Grenzwert von *Histogrammen* für eine wachsende Anzahl  $n$  unabhängiger Wiederholungen von  $X$ , wenn für  $n \rightarrow \infty$  gleichzeitig die Klassenbreite gegen Null strebt (vgl. Abb. 1 und 2).

**Dichte:** Formal ist die Dichte  $f$  von  $X$  eine nicht-negative Funktion, d.h.  $f(x) \geq 0$  für alle  $x$ , mit der Eigenschaft, daß sich Intervall-Wahrscheinlichkeiten wie folgt berechnen lassen:

$$P\{a < x \leq b\} = \int_a^b f(x)dx. \quad (1)$$

Die Wahrscheinlichkeit für das Intervall-Ereignis  $\{a < X \leq b\}$  entspricht also der *Fläche* zwischen der  $x$ -Achse und der Dichtefunktion  $f$  über diesem Intervall.



Obwohl es reicht, die Dichte  $f$  auf dem *Träger-Intervall*  $T$  zu definieren, ist es oft sinnvoll, die Dichte außerhalb von  $T$  auf Null zu setzen (d.h.  $f(x) = 0$  für  $x \notin T$ ), damit die Funktion  $f$  auf der gesamten Zahlengeraden  $\mathbb{R}$  definiert ist.

Die *Gesamtfläche* zwischen  $x$ -Achse und Dichtefunktion  $f$  ist auf  $1 = 100\%$  normiert:

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

Abb. 1: Verteilung des Lebensalters  $X$  der 1974 gestorbenen **Männer** der Bundesrepublik Deutschland mit markierter Fläche  $P\{60 \leq X < 80\}$ . Durch Verringerung der Klassenbreite gelangt man schrittweise zur (idealisierten) Dichtefunktion (unten).

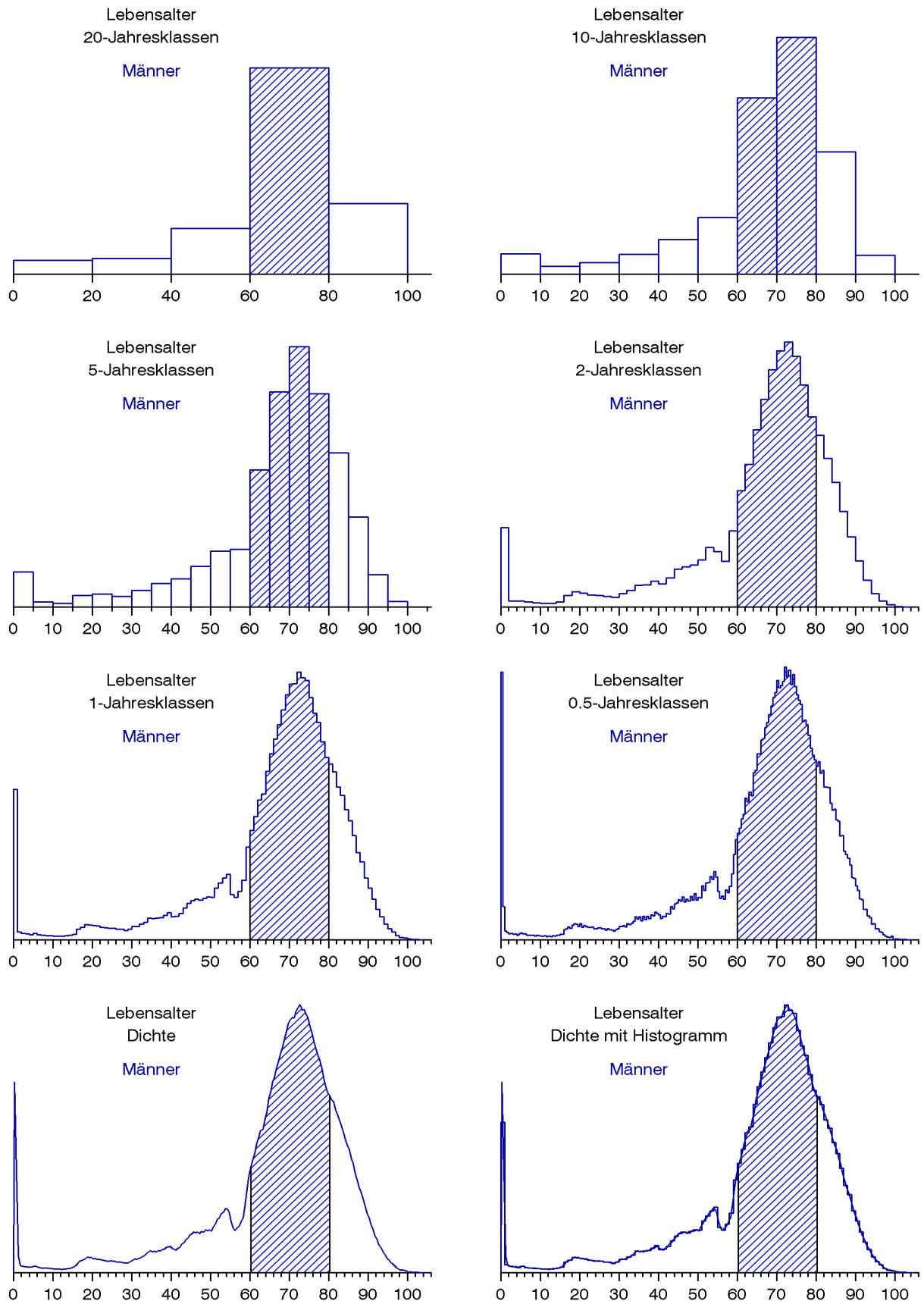
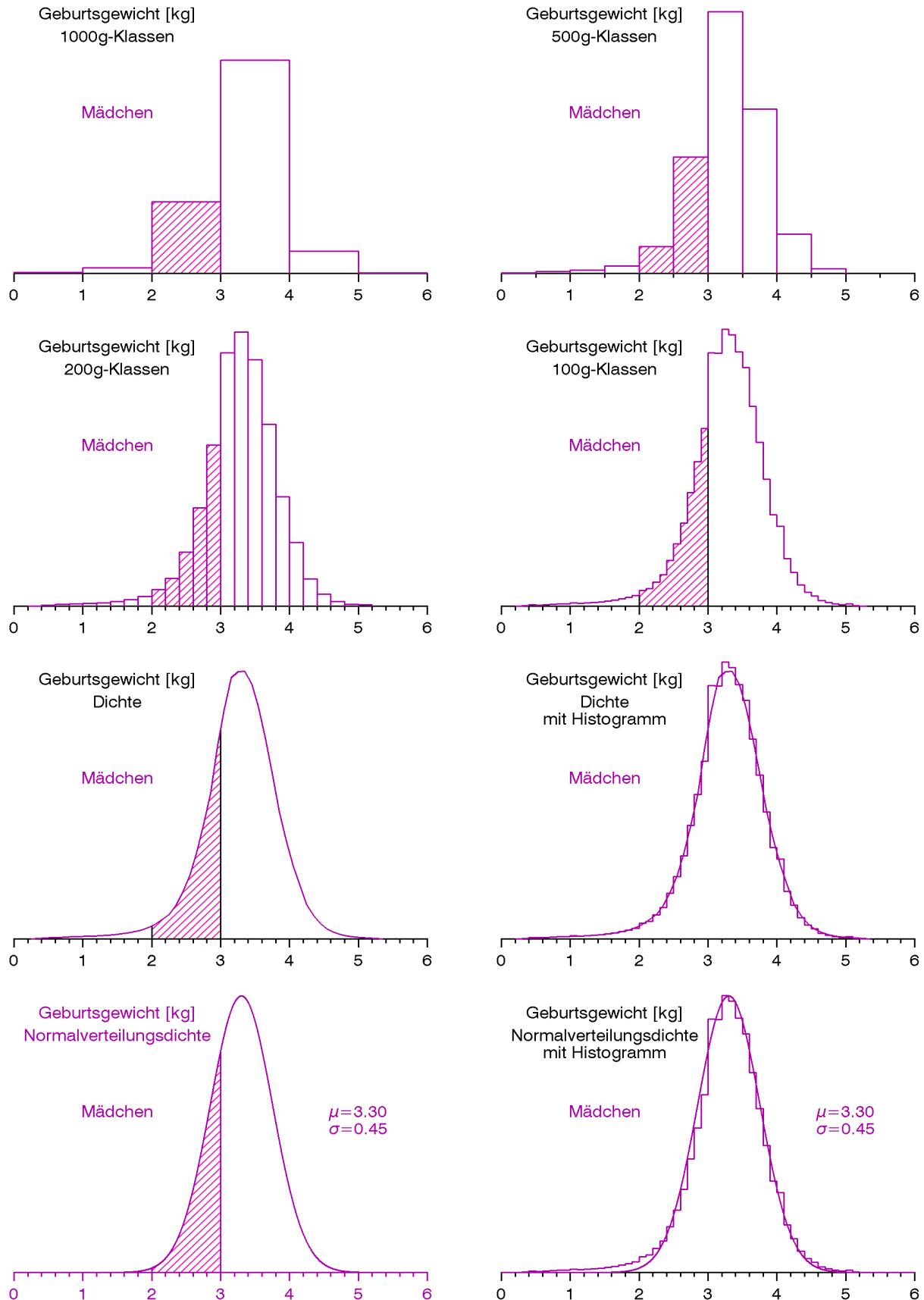


Abb. 2: Verteilung des Geburtsgewichts  $X$  der 1974 geborenen Mädelchen der Bundesrepublik Deutschland mit markierter Fläche  $P\{2 \leq X < 3\}$ . Verringerung der Klassenbreite führt wie in Abb. 1 zur Dichte. Unten: Normalverteilungsdichte aus 2.5.1 zum Vergleich.



Wir werden hier nur Dichten  $f$  betrachten, die *stetige* Funktionen auf dem Träger  $T$  darstellen, d.h. daß die Funktion  $f$  dort keine *Sprungstellen* hat bzw. sich über  $T$  *zeichnen lässt, ohne den Stift abzusetzen*. Für eine solche Dichte  $f$  ist die Wahrscheinlichkeit, daß  $X$  einen festen Wert  $a$  annimmt, immer gleich Null:

$$P\{X = a\} = 0 \quad \text{für jedes } a. \quad (3)$$

Im Gegensatz zu den *diskreten* Zufallsvariablen ist bei einer *stetigen* Zufallsvariablen  $X$  also jede Ergebnis-Wahrscheinlichkeit gleich Null. Dies ist zunächst irritierend, klärt sich aber auf, weil in der Praxis z.B. eine Gewichtsangabe  $X = 75$  [kg] immer *gerundet* ist und somit ein *Intervall-Ereignis*  $\{74,5 \leq X < 75,5\}$  mit positiver Wahrscheinlichkeit darstellt. Als Folgerung aus (3) ergibt sich, daß man bei Intervall-Wahrscheinlichkeiten über  $X$  die Beziehung „<“ (*echt kleiner*) stets durch „≤“ (*kleiner oder gleich*) ersetzen kann (und umgekehrt) also z.B.

$$P\{a < X \leq b\} = P\{a \leq X \leq b\} = P\{a \leq X < b\} = P\{a < X < b\} \quad (\text{für stetiges } X). \quad (4)$$

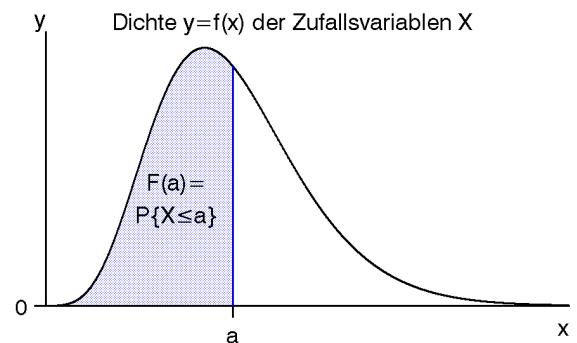
*Anmerkung für Fortgeschrittene:* Die Wahrscheinlichkeit  $P\{X \in A\}$  für ein *beliebiges* Ereignis (also z.B. wenn  $A$  *kein* Intervall ist) erhält man ebenfalls als *Fläche* unter der Dichtefunktion  $f$  über dem Abschnitt  $A$ , d.h. als *Integral* von  $f$  über  $A$ . Wir werden hier aber für stetiges  $X$  nur Wahrscheinlichkeiten für *Intervalle* oder für *Vereinigungen von Intervallen* betrachten, weil dies für die Anwendungen hier ausreicht.

**Verteilungsfunktion:** Ein wichtiges Werkzeug zur Berechnung von Wahrscheinlichkeiten ist die *Verteilungsfunktion*  $F$  von  $X$ , definiert durch

$$F(a) = P\{X \leq a\} = \int_{-\infty}^a f(x)dx \quad (5)$$

Die *Ableitung* der Verteilungsfunktion  $F$  auf dem Träger  $T$  ist die Dichte  $f$ :

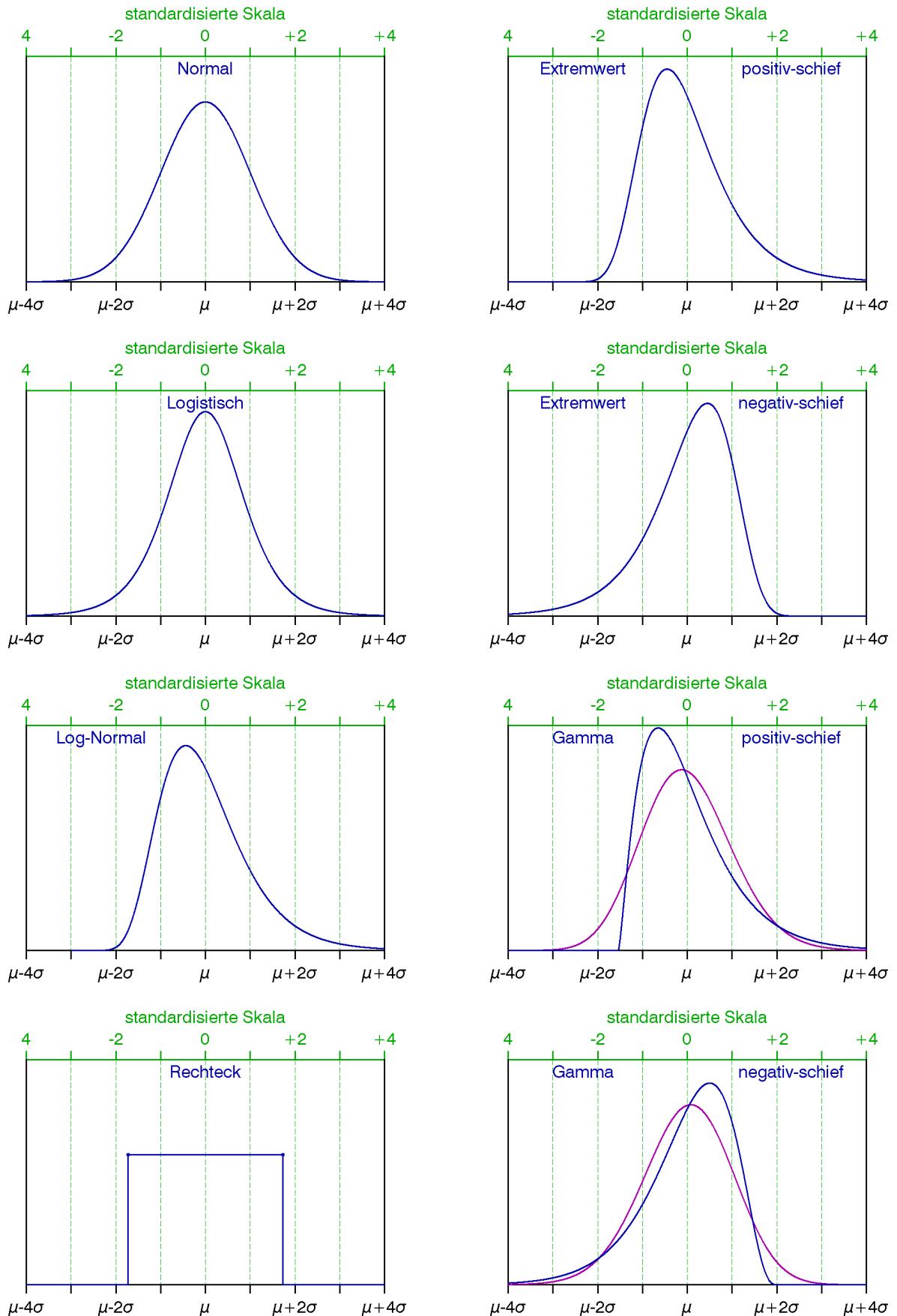
$$F'(x) = f(x) \quad \text{für } x \in T. \quad (6)$$



Mit der Verteilungsfunktion  $F$  lassen sich Wahrscheinlichkeiten für  $X$  berechnen:

$$P\{a < X \leq b\} = F(b) - F(a), \quad P\{X > a\} = 1 - F(a). \quad (7)$$

Abb. 3: Einige typische und mathematisch einfache Dichtefunktionen, die nur teilweise später näher besprochen werden (z.B. die Normalverteilung in 2.5).



## 2.5 Die Normalverteilung

Die wichtigste stetige Verteilung ist die Normalverteilung  $N(\mu, \sigma^2)$ . Sie hat den Träger  $T = \mathbb{R}$  und wird durch *zwei* Parameter beschrieben

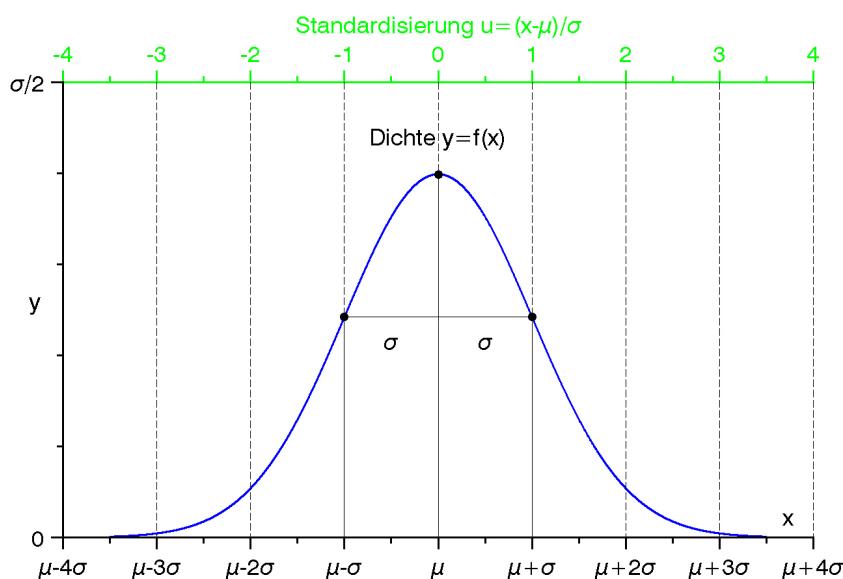
- einen *Lageparameter*, den sogenannten *Erwartungswert*  $\mu$
- einen *positiven Skalenparameter*, die sogenannte *Standardabweichung*  $\sigma > 0$ .

Die Bedeutung der Normalverteilung liegt vor allem darin, daß man zahlreiche Verteilungen unter gewissen Umständen hinreichend genau durch eine Normalverteilung *approximieren* kann, z.B. die *Binomial-Verteilung* (vgl. 2.5.6) oder die Verteilung eines *Mittelwerts* von unabhängigen Wiederholungen einer Zufallsvariablen (vgl. 2.6.5). Die Ursache hierfür ist ein tiefliegendes Resultat der Wahrscheinlichkeitstheorie, der *Zentrale Grenzwertsatz*, auf den hier nicht näher eingegangen wird (vgl. aber 2.6.5).

### 2.5.1 Die Dichte der Normalverteilung

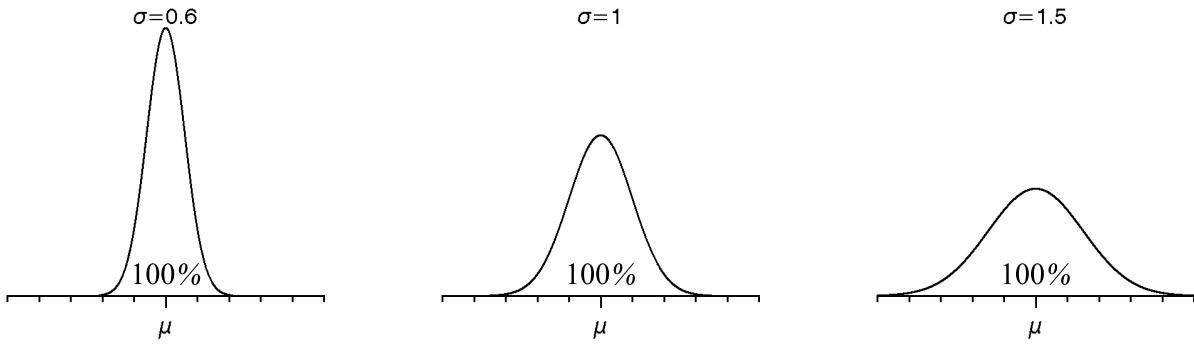
Die *Dichte*  $f$  der Normalverteilung  $N(\mu, \sigma^2)$  ist für *jede reelle Zahl*  $x \in \mathbb{R}$  gegeben durch (vgl. Abb. 1)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \quad \text{Dichte von } N(\mu, \sigma^2). \quad (1)$$



**Abb 1:** Dichte  $f(x)$  der Normalverteilung  $N(\mu, \sigma^2)$  als Funktion von  $x$  (*untere Achse*) und als Funktion  $f(u)$  der *Standardisierung*  $u = \frac{1}{\sigma}(x - \mu)$  (*obere Achse*).

Die Dichtefunktion (Abb. 1) stellt eine *Gauß'sche Glockenkurve* dar mit einem *Maximum* in  $x_m = \mu$  und den *Wendepunkten*  $x_w = \mu \pm \sigma$ . Die Kurve ist *symmetrisch* um den *Lageparameter*  $\mu$ . Der *Skalenparameter*  $\sigma$  ist ein Maß für die *Konzentration* um den Wert  $\mu$  (vgl. auch 2.5.4). Je *kleiner*  $\sigma$  ist, desto *stärker* ist die Konzentration um  $\mu$  (vgl. Abb. 2). Oder umgekehrt: je *größer*  $\sigma$  ist, desto *stärker* ist die *Streuung* um den Wert  $\mu$ .



**Abb. 2:** Streuung bzw. Konzentration der Dichte von  $N(\mu, \sigma^2)$  um den Wert  $\mu$  für verschiedene Werte von  $\sigma$ . Die Fläche unter der Kurve ist stets gleich groß (100%).

Eine Zufallsvariable  $X$  ist  $N(\mu, \sigma^2)$ -verteilt, wenn sie die Dichte  $f$  aus (1) besitzt. Betrachtet man die *Abweichung* der Zufallsvariablen  $X$  vom *Erwartungswert*  $\mu$  in *Vielfachen* der *Standardabweichung*  $\sigma$  (vgl. auch Abb. 1), so erhält man die *Standardisierung*:

$$U = \frac{X - \mu}{\sigma} \quad (\text{Standardisierung von } X). \quad (2)$$

Die Standardisierung  $U$  hat wieder eine Normalverteilung und zwar die sogenannte *Standard-Normalverteilung* (die wir im nächsten Abschnitt näher betrachten). Dieser Sachverhalt erlaubt es, Wahrscheinlichkeiten über  $X$  unter Verwendung der Standard-Normalverteilung zu bestimmen (vgl. 2.5.3).

Löst man (2) nach  $X$  auf, so erhält man  $X = \mu + \sigma U$ . Folglich lässt sich eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable  $X$  auch stets als eine lineare Transformation einer standardnormalverteilten Zufallsvariablen  $U$  darstellen.

### 2.5.2 Die Standard-Normalverteilung

Für  $\mu = 0$  und  $\sigma = 1$  liegt die *Standard-Normalverteilung*  $N(0, 1)$  vor, deren Dichte hier mit  $\varphi$  bezeichnet wird (vgl. Abb. 3)

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \quad (1)$$

Die Verteilungsfunktion  $\Phi$  der Standard-Normalverteilung ist (vgl. Abb. 3)

$$\Phi(a) = \int_{-\infty}^a \varphi(x) dx. \quad (2)$$

Dieses Integral lässt sich *nicht* explizit lösen. Die Funktion  $\Phi$  liegt aber im Anhang **T tabelliert** vor und kann in Tabellenkalkulations- oder statistischen Programmen auf dem Computer (und auch auf leistungsfähigen Taschenrechnern) direkt aufgerufen werden.

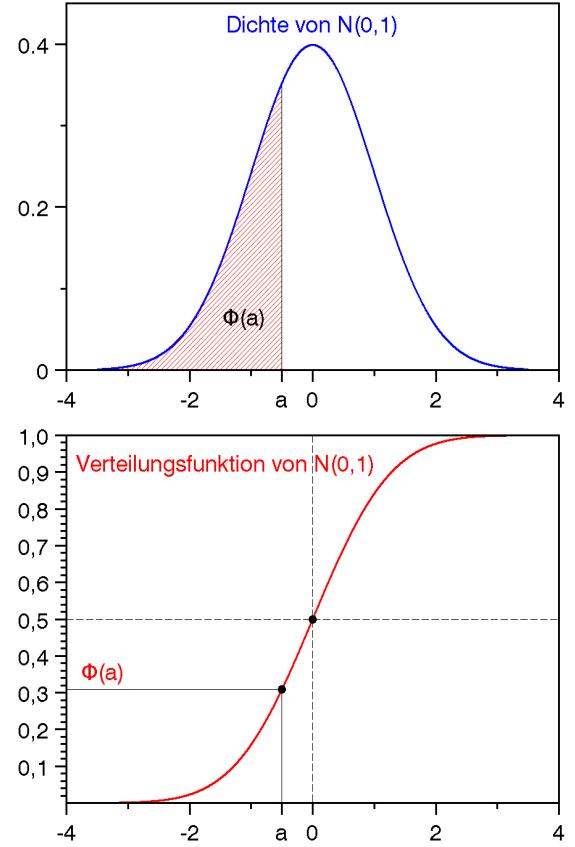


Abb. 3: Dichte  $\varphi$  und Verteilungsfunktion  $\Phi$  der Standard-Normalverteilung

Die Verteilungsfunktion  $\Phi$  ist eine *sigmoide* Kurve, die um den Punkt  $(0, \frac{1}{2})$  *drehsymmetrisch* ist, d.h.

$$\Phi(-b) = 1 - \Phi(+b). \quad (3)$$

### 2.5.3 Berechnung von Wahrscheinlichkeiten bei Normalverteilung

Für eine Zufallsvariable  $X$  mit  $N(\mu, \sigma^2)$ -Verteilung lassen sich die Ereignisse für  $X$  auch äquivalent durch die Standardisierung  $U = \frac{1}{\sigma}(X - \mu)$  formulieren:

$$\{X \leq x\} = \{U \leq u\}, \quad \{X \geq x\} = \{U \geq u\} \quad \text{mit} \quad u = \frac{1}{\sigma}(x - \mu).$$

Da  $U$  eine Standard-Normalverteilung hat, können Wahrscheinlichkeiten für  $X$  unter Verwendung der Verteilungsfunktion  $\Phi$  von  $N(0, 1)$  bestimmt werden:

$$P\{X \leq x\} = \Phi(+u) \quad \text{mit} \quad u = \frac{1}{\sigma}(x - \mu). \quad (1)$$

$$P\{X \geq x\} = \Phi(-u) = 1 - \Phi(u) \quad (2)$$

$$P\{x_1 \leq X \leq x_2\} = \Phi(u_2) - \Phi(u_1) \quad \text{mit} \quad u_{1,2} = \frac{1}{\sigma}(x_{1,2} - \mu). \quad (3)$$

**Wirkstoffgehalt einer Tablette:** Bei der Tablettenproduktion variiert der Anteil  $X$  [in mg] des schmerzstillenden Wirkstoffs Acetylsalicylsäure pro Tablette geringfügig und ist (in guter Näherung)  $N(\mu, \sigma^2)$ -verteilt. Für eine konkrete Produktion mit  $\mu = 300$  mg (Sollwert) und  $\sigma = 20$  mg ergeben sich aus (1) - (3) die Wahrscheinlichkeiten für folgende Ereignisse:

$$\begin{aligned} \text{stark unterdosiert: } P\{X \leq 250\} &= \Phi(-2,5) &= 0,62 \% \\ \text{nicht unterdosiert: } P\{X \geq 300\} &= \Phi(0) &= 50,00 \% \\ \text{nicht stark fehdosiert: } P\{260 \leq X \leq 340\} &= \Phi(+2) - \Phi(-2) &= 95,44 \% \end{aligned}$$

#### 2.5.4 Eigenschaften der Normalverteilung

Bei einer  $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen  $X$  ist der *Erwartungswert*  $\mu$  das *Symmetrie-Zentrum* der Verteilung von  $X$ , d.h. es gilt:

$$P\{X \leq \mu - a\} = P\{X \geq \mu + a\} \quad \text{für jedes } a. \quad (1)$$

Die *Standardabweichung*  $\sigma$  beschreibt die *Streuung* um den Erwartungswert  $\mu$ . Die Wahrscheinlichkeit für den  $1\sigma$ -Bereich um  $\mu$  ist (vgl. Abb.4) bereits

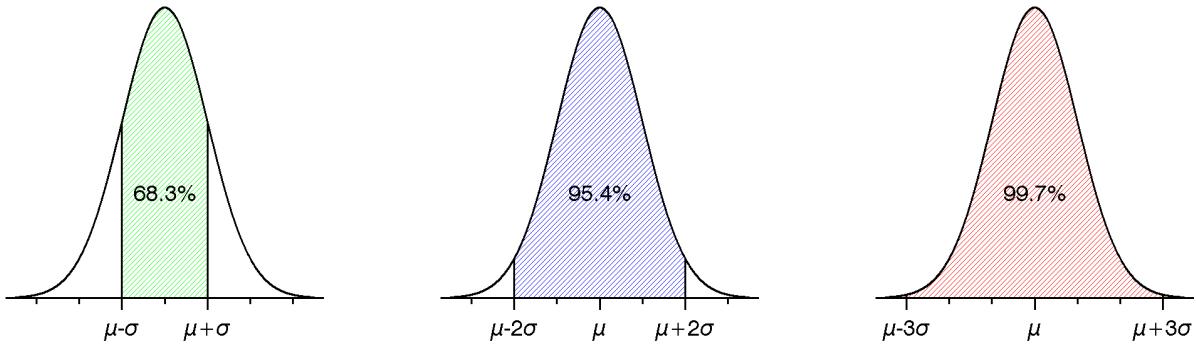
$$P\{|X - \mu| \leq 1\sigma\} = P\{\mu - 1\sigma \leq X \leq \mu + 1\sigma\} = 68,3\% \quad (1\sigma\text{-Bereich}), \quad (2)$$

und für den  $2\sigma$ -,  $3\sigma$ - und  $4\sigma$ -Bereich ergeben sich:

$$P\{|X - \mu| \leq 2\sigma\} = P\{\mu - 2\sigma \leq X \leq \mu + 2\sigma\} = 95,4\% \quad (2\sigma\text{-Bereich}) \quad (3)$$

$$P\{|X - \mu| \leq 3\sigma\} = P\{\mu - 3\sigma \leq X \leq \mu + 3\sigma\} = 99,7\% \quad (3\sigma\text{-Bereich}) \quad (4)$$

$$P\{|X - \mu| \leq 4\sigma\} = P\{\mu - 4\sigma \leq X \leq \mu + 4\sigma\} = 99,9937\% \quad (4\sigma\text{-Bereich}) \quad (5)$$



**Abb. 4:** Die Wahrscheinlichkeiten für den  $1\sigma$ -Bereich (links),  $2\sigma$ -Bereich (mitte) und  $3\sigma$ -Bereich bei einer  $N(\mu, \sigma^2)$ -Verteilung.

Die Wahrscheinlichkeit, daß  $X$  außerhalb des  $5\sigma$ -Bereichs liegt ist mit 0,000057% so gering, daß sie in konkreten Anwendungen als Null angesehen werden kann.

Wenn man eine normalverteilte Zufallsvariable *linear transformiert* (z.B. bei der Umrechnung einer Temperatur von Celsius auf Fahrenheit) so ergibt sich wieder eine Normalverteilung (das logische Symbol „ $\Rightarrow$ “ bedeutet: *impliziert*):

### **Lineare Transformation von Normalverteilungen**

$$X \text{ ist } N(\mu_X, \sigma_X^2)\text{-verteilt} \quad \Rightarrow \quad Y = a + bX \text{ ist } N(\mu_Y, \sigma_Y^2)\text{-verteilt mit}$$

$$\mu_X = a + b\mu_X, \quad \sigma_Y^2 = b^2\sigma_X^2.$$

Eine weitere wichtige Eigenschaft von Normalverteilungen ist, daß die Summe von *unabhängigen normalverteilten Zufallsvariablen* wieder *normalverteilt* ist:

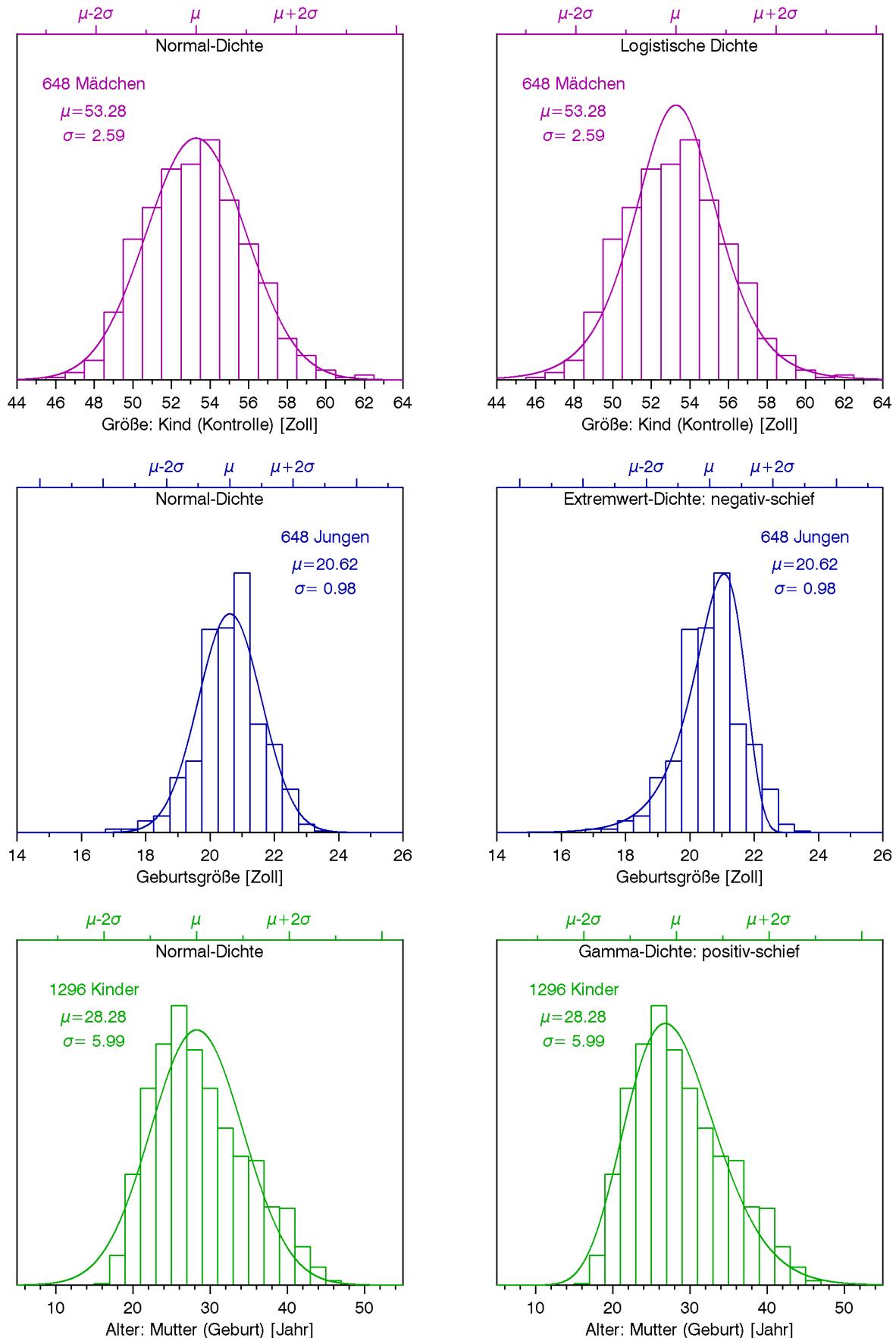
### **Summe zweier unabhängiger Normalverteilungen**

Ist  $X_1$  mit  $N(\mu_1, \sigma_1^2)$ -Verteilung unabhängig von  $X_2$  mit  $N(\mu_2, \sigma_2^2)$ -Verteilung, so hat die Summe  $X_1 + X_2$  eine  $N((\mu_1 + \mu_2), (\sigma_1^2 + \sigma_2^2))$ -Verteilung.

## **2.5.5 Auftreten der Normalverteilung**

Eine Zufallsvariable  $X$ , die sich *additiv* aus *vielen kleinen unabhängigen* Effekten zusammensetzt, ist näherungsweise normalverteilt (eine mathematische Präzisierung hierfür liefert der sogenannte *Zentraler Grenzwertsatz* der Wahrscheinlichkeitstheorie). Dies erklärt, warum wichtige Zufallsvariable der *deduktiven Statistik* (wie z.B. *Schätzer und Teststatistiken* in Kap. 3) approximativ normalverteilt sind. Dagegen kann man bei Zufallsvariablen  $X$ , die in konkreten Anwendungen auftreten, im allgemeinen *nicht* von einer Normalverteilung ausgehen. Optisch erkennt man eine Normalverteilung daran, daß das *Histogramm* der klassifizierten Beobachtungswerte  $x_1, x_2, \dots, x_n$  (aus unabhängigen Wiederholungen) nicht zu stark von einer Gauß'schen Glockenkurve abweicht, also *symmetrisch* ist und vom Symmetrie-Zentrum zu den Rändern *schnell* abfällt. Die *deduktive Statistik* liefert auch *formale Tests* auf Normalverteilung.

Abb. 5 zeigt die Histogramme von drei verschiedenen Merkmalen der *StatLab-Daten* zusammen mit der Dichte der zugehörigen optimalen  $N(\mu, \sigma^2)$ -Dichte (mit  $\mu = \bar{x}$  als Mittelwert und  $\sigma = s_x$  als Standardabweichung der  $x$ -Werte, vgl. 2.6.5) und einer jeweils alternativen Dichte. Eine relativ gute Übereinstimmung mit der Normalverteilung liegt bei der Größe des Kindes (zum Kontrollzeitpunkt) vor.



**Abb. 5:** Histogramme für Merkmale der StatLab-Daten zusammen mit der optimalen  $N(\mu, \sigma^2)$ -Dichte (links) sowie einer alternativen Dichte (rechts), vgl. auch den Text oben.

### 2.5.6 Die Normal-Approximation der Binomial-Verteilung

Nach dem *Grenzwertsatz von de Moivre und Laplace* strebt die Binomial-Verteilung  $B(n, p)$  für *wachsenden Umfang*  $n$  gegen die Normalverteilung  $N(\mu, \sigma^2)$  (vgl. Abb. 6), wobei die Parameter dem Erwartungswert  $\mu$  und der Standardabweichung  $\sigma$  der  $B(n, p)$ -Verteilung entsprechen (vgl. hierzu auch die Berechnung in 2.6.4):

$$\mu = np, \quad \sigma = \sqrt{npq} \quad \text{mit} \quad q = 1 - p. \quad (1)$$

Für *großes*  $n$  ergeben sich hieraus unter Verwendung der Verteilungsfunktion  $\Phi$  von  $N(0, 1)$  folgende

**Normalapproximationen für  $B(n,p)$ -verteiltes  $X$ :**

$$P\{X \leq k\} \approx \Phi(u) \quad \text{mit} \quad u = \frac{1}{\sigma}(k + \frac{1}{2} - \mu), \quad (2)$$

$$P\{X \geq m\} \approx \Phi(-v) = 1 - \Phi(v) \quad \text{mit} \quad v = \frac{1}{\sigma}(m - \frac{1}{2} - \mu), \quad (3)$$

$$P\{m \leq X \leq k\} \approx \Phi(u) - \Phi(v) \quad \text{mit} \quad u, v \text{ wie oben.} \quad (4)$$

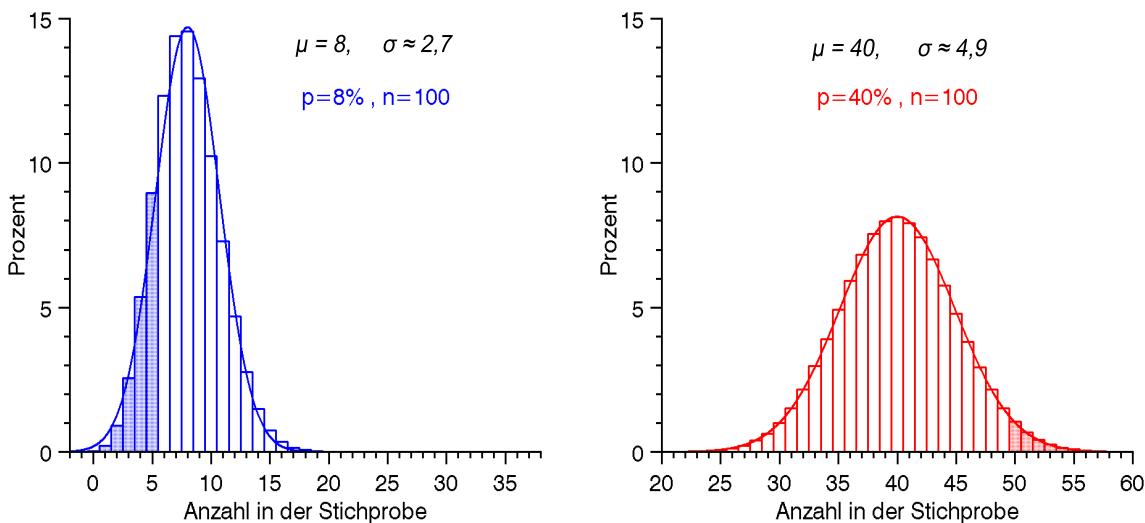
Man beachte, daß in  $u$  bzw.  $v$  jeweils noch die sogenannte *Stetigkeitskorrektur*  $\pm \frac{1}{2}$  auf-taucht, deren Sinn man in Abb. 6 erkennt. Die Wahrscheinlichkeit  $P\{X \leq 5\}$  entspricht der schraffierten Fläche in Abb. 6 *links* und diese Fläche soll durch die zugehörige Fläche unter der Normalverteilungs-Dichte approximiert werden, d.h. durch die Fläche bis zum Wert 5,5 auf der  $x$ -Achse (weil die *Säule* für die Realisierung 5 sich *von 4,5 bis 5,5* erstreckt). Wenn man die Fläche unter der Dichtefunktion nur bis zu  $x = 5$  bestimmt, so verwendet man systematisch eine zu kleine Fläche als Approximation der schraffierten Fläche. Entsprechend wird bei der Approximation der Wahrscheinlichkeit  $P\{X \geq 50\}$ , die der schraffierten Fläche in Abb. 6 *rechts* entspricht, die zugehörige Fläche unter der Kurve bereits ab dem Wert 49,5 auf der  $x$ -Achse verwendet. Die exakten Werte (aus 2.3.3) und ihre Approximationen stimmen hier gut überein:

$$\begin{aligned} 17,99\% &= P\{X \leq 5\} \approx \Phi(-0,92) = 17,88\% && \text{für } n = 100, p = 8\%, \\ 2,71\% &= P\{X \geq 50\} \approx \Phi(-1,94) = 2,62\% && \text{für } n = 100, p = 40\%. \end{aligned}$$

Obwohl der Einfluß der Stetigkeitskorrektur  $\pm \frac{1}{2}$  bei wachsenden Umfang  $n$  immer geringer wird, sollte die Stetigkeitskorrektur *immer* verwendet werden, um eine größtmögliche Genauigkeit der Approximation zu erreichen. *Ohne* Verwendung dieser Korrektur ist die Approximation *systematisch kleiner* als die wahre Wahrscheinlichkeit. Für obige Werte ergeben sich z.B. *ohne Stetigkeitskorrektur* die deutlich schlechteren Approximationen:

$$\begin{aligned} 17,99\% &= P\{X \leq 5\} \approx \Phi(-1,11) = 13,44\% \quad \text{für } n = 100, p = 8\%, \\ 2,71\% &= P\{X \geq 50\} \approx \Phi(-2,04) = 2,02\% \quad \text{für } n = 100, p = 40\%. \end{aligned}$$

Die Normal-Approximationen (2)-(4) werden für *wachsendes*  $n$  immer genauer. Bei *festem* Umfang  $n$  ist die Genauigkeit der Approximation für  $p = 50\%$  am besten und wird schlechter, je weiter  $p$  von 50% entfernt ist. Man sollte die Approximationen daher nur dann verwenden, wenn der Umfang  $n$  *nicht zu klein* ist, und die Wahrscheinlichkeit  $p$  *weder zu klein noch zu groß* ist. Eine *Faustregel* besagt, daß man die Approximationen nur anwenden soll, wenn  $\sigma^2 = npq \geq 5$  ist.



**Abb. 6:** Histogramme der  $B(n, p)$ -Verteilungen zu den Wahlumfragen aus 2.3.3 und Dichte der approximierenden  $N(\mu, \sigma^2)$ -Verteilung.  
Die markierte Fläche entspricht der Wahrscheinlichkeit  $P\{X \leq 5\}$  (links) bzw.  $P\{X \geq 50\}$  (rechts).

### ☒ Tipps für Microsoft Excel

Funktionen: Die Verteilungsfunktion  $\Phi$  der Standard-Normalverteilung heißt dort

$$\text{STANDNORMVERT}(x) = \Phi(x)$$

Für  $N(\mu, \sigma^2)$ -verteiltes  $X$  läßt sich die *Verteilungsfunktion* direkt berechnen als

$$\text{NORMVERT}(x; \mu; \sigma; 1) = P\{X \leq x\} \quad (\text{Verteilungsfunktion}).$$

Und die *Dichte*  $f$  von  $N(\mu, \sigma^2)$  erhält man (z.B. für grafische Darstellungen) wie folgt

$$\text{NORMVERT}(x; \mu; \sigma; 0) = f(x) \quad (\text{Dichte}).$$

Man beachte, daß das *letzte* Argument bei **NORMVERT** ein *Indikator* ist, der angibt ob die *Verteilungsfunktion* oder die *Dichte* berechnet wird.

## 2.6 Erwartungswert, Varianz und Standardabweichung von Zufallsvariablen

Im Abschnitt 1.3 haben wir den *Mittelwert* und die *Standardabweichung* bzw. *Varianz* von *beobachteten Werten* eingeführt, um die *Lage* und *Streuung* der Werte zu beschreiben. Wir wollen jetzt entsprechende *Lage- und Streuungs-Parameter* für die *Verteilung* einer Zufallsvariablen  $X$  einführen. Bei der Normalverteilung  $N(\mu, \sigma^2)$  haben wir den Lageparameter  $\mu$  bereits als *Erwartungswert* und den Skalenparameter  $\sigma$  als *Standardabweichung* bezeichnet und diese Begriffe sollen jetzt allgemein für eine beliebige Verteilung definiert werden. Zunächst werden wir eine *heuristische Interpretation* dieser Parameter für die Verteilung von  $X$  geben und danach aufzeigen, wie man sie für konkrete Verteilungen *berechnen* kann.

Ausgangspunkt der *heuristischen Interpretation* sind  $n$  unabhängige Wiederholungen der Zufallsvariablen  $X$  mit den zugehörigen Realisierungen  $x_1, \dots, x_n$ . Für diese *beobachteten* Werte können wir die folgenden Maßzahlen aus 1.3 berechnen

$$\text{empirischer Mittelwert von } x_1, \dots, x_n : \bar{x} = \frac{1}{n} \sum_i x_i$$

$$\text{empirische Varianz von } x_1, \dots, x_n : s_x^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

$$\text{empirische Standardabweichung von } x_1, \dots, x_n : s_x = \sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2}$$

Lassen wir nun die Anzahl  $n$  der Wiederholungen *beliebig anwachsen* (Schreibweise:  $n \rightarrow \infty$ ), so kommen immer neue  $x$ -Werte hinzu und die obigen Maßzahlen *stabilisieren* sich wie folgt.

Der *empirische* Mittelwert  $\bar{x}$  strebt gegen eine Zahl  $E(X)$ , den *Erwartungswert* oder *theoretischen Mittelwert* von  $X$ .

Die *empirische* Varianz  $s_x^2$  strebt gegen eine Zahl  $Var(X) \geq 0$ , die *Varianz von  $X$* .

Die *empirische* Standardabweichung  $s_x$  strebt gegen die *Standardabweichung*  $SD(X) = \sqrt{Var(X)}$  von  $X$ .

<b>Bezeichnungen:</b>	$\mu = \mu_X = \mu(X) = E(X)$	(Erwartungswert von $X$ ),
	$\sigma^2 = \sigma_X^2 = \sigma^2(X) = Var(X)$	(Varianz von $X$ ),
	$\sigma = \sigma_X = \sigma(X) = SD(X)$	(Standardabweichung von $X$ ).

Der Erwartungswert  $\mu$  und die Varianz  $\sigma^2$  bzw. Standardabweichung  $\sigma$  lassen sich auch *formal definieren*, wobei für *diskrete* und *stetige* Zufallsvariablen jeweils unterschiedliche Methoden benutzt werden.

### 2.6.1 Der Erwartungswert einer *diskreten* Zufallsvariablen

Für eine *diskrete* Zufallsvariable  $X$ , die Werte aus ihrem *Träger*  $T = \{x_1, x_2, \dots\}$  annimmt, ist ihr Erwartungswert gegeben durch die Summe von *Produkten* aller *Ergebnisse* mit *ihrer Ergebnis-Wahrscheinlichkeit*:

$$\begin{aligned} E(X) &= \sum_{x \in T} x \cdot P\{X = x\} && \text{Erwartungswert für diskretes } (1) \\ &= x_1 \cdot P\{X = x_1\} + x_2 \cdot P\{X = x_2\} + \dots && \text{X mit Träger } T = \{x_1, x_2, \dots\} \end{aligned}$$

Der Erwartungswert  $\mu = E(X)$  repräsentiert den *Durchschnittswert* der Zufallsvariablen  $X$  und entspricht geometrisch dem *Schwerpunkt der Wahrscheinlichkeitsmasse*, d.h. dem *Schwerpunkt (auf der x-Achse) des Stabdiagramms bzw. Histogramms* der Verteilung von  $X$  (vgl. z.B. **2.2 Abb. 1 und 2**).

*Für Fortgeschrittene:* Wenn der Träger  $T$  unendlich viele Ergebnisse enthält (wie z.B. bei der Poisson-Verteilung in **2.7**), so handelt es sich in (1) um eine *unendliche Reihe*. Streng genommen ist der Erwartungswert  $E(X)$  nur dann definiert, wenn die Reihe der Absolutbeträge  $|x \cdot P\{X = x\}|$  konvergent ist.

**Homogener Würfel:** Der Erwartungswert für die Augenzahl  $X$  eines homogenen Würfels mit dem Träger  $T = \{1, 2, 3, 4, 5, 6\}$  ergibt sich wie folgt

$$E(X) = \sum_{x=1}^6 x \cdot P\{X = x\} = \sum_{x=1}^6 x \cdot \frac{1}{6} = \frac{21}{6} = 3,5.$$

Man beachte, daß der Erwartungswert 3,5 nicht *gewürfelt* werden kann, weil der Wert 3,5 nicht zu den möglichen Ergebnissen von  $X$  gehört. Der *Erwartungswert* ist daher *nicht* die bei *einem* Werfen zu erwartende Augenzahl, sondern der (zu erwartende) *Mittelwert* in einer wachsenden Anzahl unabhängiger Würfe.  $\square$

Der Erwartungswert der *quadrierten* Zufallsvariablen  $X^2$  oder allgemeiner der Erwartungswert einer beliebigen *Funktion*  $g(X)$  kann wie folgt berechnet werden:

$$\begin{aligned} E(g(X)) &= \sum_{x \in T} g(x) \cdot P\{X = x\} && (2) \\ &= g(x_1) \cdot P\{X = x_1\} + g(x_2) \cdot P\{X = x_2\} + \dots && \text{für } T = \{x_1, x_2, \dots\}. \end{aligned}$$

**Homogener Würfel (Fortsetzung):** Der Erwartungswert für das Quadrat  $g(X) = X^2$  der Augenzahl  $X$  eines homogenen Würfels ergibt sich wie folgt

$$E(X^2) = \sum_{x=1}^6 x^2 \cdot P\{X = x\} = \sum_{x=1}^6 x^2 \cdot \frac{1}{6} = \frac{91}{6} \approx 15,17.$$

### 2.6.2 Der Erwartungswert einer *stetigen* Zufallsvariablen

Für eine *stetige* Zufallsvariable  $X$ , die Werte aus einem Intervall  $T$  reeller Zahlen annimmt und dort die Dichtefunktion  $f$  hat, ist ihr Erwartungswert als das folgende Integral über dem Intervall  $T$  gegeben:

$$E(X) = \int_T x \cdot f(x) dx \quad (\text{Erwartungswert für stetiges } X \text{ mit Dichte } f) \quad (1)$$

Der Erwartungswert  $\mu = E(X)$  repräsentiert den *Durchschnittswert* der Zufallsvariablen  $X$  und entspricht geometrisch dem *Schwerpunkt der Wahrscheinlichkeitsmasse*, d.h. dem *Schwerpunkt der Dichte  $f$  auf der  $x$ -Achse* (vgl. z.B. **2.4 Abb. 3** und **2.5 Abb. 5**).

*Für Fortgeschrittene:* Genau genommen ist der Erwartungswert nur dann durch (1) definiert, wenn das Integral des Absolutbetrages  $|x \cdot f(x)|$  über  $T$  *endlich* ist. Es gibt Verteilungen, bei denen das nicht der Fall ist, z.B. die Student'sche  $t_1$ -Verteilung mit einem Freiheitsgrad.

Der Erwartungswert der *quadrierten* Zufallsvariablen  $X^2$  oder allgemeiner der Erwartungswert einer beliebigen *Funktion  $g(X)$*  kann wie folgt berechnen werden:

$$E(g(X)) = \int_T g(x) \cdot f(x) dx. \quad (2)$$

**Standard-Normalverteilung:** Für eine  $N(0, 1)$ -verteilte Zufallsvariable  $X$  ist der Träger  $T = \mathbb{R}$  und die Dichte ist die Funktion  $\varphi$ . Unter Berücksichtigung der Ableitung  $\varphi'(x) = x \cdot \varphi(x)$  erhält man (mit den üblichen Integrationstechniken):

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot \varphi(x) dx = \int_{-\infty}^0 x \cdot \varphi(x) dx + \int_0^{+\infty} x \cdot \varphi(x) dx = -\varphi(0) + \varphi(0) = 0 \\ E(X^2) &= \int_{-\infty}^{+\infty} x^2 \cdot \varphi(x) dx = [-x \cdot \varphi(x)]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} -x \cdot \varphi(x) dx = 0 + 1 = 1. \quad \square \end{aligned}$$

### 2.6.3 Varianz und Standardabweichung einer Zufallsvariablen

Die *Varianz* einer Zufallsvariablen  $X$  lässt sich auch beschreiben als die *durchschnittliche quadratische Abweichung*  $g(X) = [X - \mu_x]^2$  der Zufallsvariablen  $X$  von ihrem Erwartungswert  $\mu_X$ , d.h. als Erwartungswert von  $g(X)$ :

$$\sigma_X^2 = \text{Var}(X) = E([X - \mu_X]^2) \quad (\text{Varianz von } X) \quad \text{mit} \quad \mu_X = E(X). \quad (1)$$

Die *Standardabweichung* von  $X$  ist die *Wurzel* aus der Varianz (und hat somit die gleiche Dimension wie  $X$ ):

$$\sigma_X = SD(X) = \sqrt{Var(X)} \quad (\text{Standardabweichung von } X). \quad (2)$$

Prinzipiell läßt sich die *Varianz von  $X$  als Erwartungswert von  $g(X) = [X - \mu_X]^2$*  für diskretes bzw. stetiges  $X$  nach (2) in **2.6.1** bzw. **2.6.2** bestimmen. Oft ist es jedoch einfacher, erst den Erwartungswert des *Quadrats  $X^2$*  zu bestimmen, und dann die Varianz zu bestimmen aus

$$\sigma_X^2 = Var(X) = E(X^2) - \mu_X^2 \quad \text{mit} \quad \mu_X = E(X). \quad (3)$$

Da die Varianz  $\sigma_X^2$  nicht *negativ* sein kann, ergibt sich aus (3), daß der Erwartungswert  $E(X^2)$  immer mindestens so groß sein muß wie das Quadrat  $\mu_X^2 = [E(X)]^2$ . Für  $\sigma_X^2 > 0$  ist sogar  $E(X^2) > [E(X)]^2$ .

**Homogener Würfel (Fortsetzung):** Die Varianz und Standardabweichung der Augenzahl  $X$  eines homogenen Würfels ergibt sich mit den Ergebnissen in **2.6.1** aus (3) wie folgt:

$$\sigma_X^2 = Var(X) = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{35}{12} \approx 2,92, \quad \sigma_X = SD(X) \approx 1,71. \quad \square$$

**Standard-Normalverteilung (Fortsetzung):** Für eine  $N(0, 1)$ -verteilte Zufallsvariable  $X$  haben wir in **2.6.2** bereits den Erwartungswert  $\mu_X = 0$  und  $E(X^2) = 1$  bestimmt. Hieraus ergibt sich die Varianz und Standardabweichung mit (3) wie folgt

$$\sigma_X^2 = Var(X) = 1 - 0^2 = 1, \quad \sigma_X = SD(X) = 1. \quad \square$$

#### 2.6.4 Eigenschaften von Erwartungswert, Varianz und Standardabweichung

Erwartungswerte und Varianzen lassen sich manchmal einfacher unter Verwendung der folgenden Rechenregeln bestimmen. Wenn für eine Zufallsvariable  $X$  ihr Erwartungswert und ihre Varianz bereits bekannt sind, so ergibt sich der Erwartungswert und die Varianz für die *lineare Transformation*  $g(X) = a + bX$  aus:

$$E(a + bX) = a + bE(X), \quad (1)$$

$$Var(a + bX) = b^2 \cdot Var(X), \quad SD(a + bX) = |b| \cdot SD(X).$$

Speziell für  $a = 0$  und  $b = -1$  ergibt sich, daß  $X$  und  $-X$  zwar den *entgegengesetzten* Erwartungswert aber die *gleiche* Standardabweichung haben:

$$E(-X) = -E(X), \quad SD(-X) = SD(X). \quad (2)$$

**Normalverteilung:** Für eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable  $X$  hat ihre Standardisierung  $U = \frac{1}{\sigma}(X - \mu)$  eine  $N(0, 1)$ -Verteilung. Wie wir bereits gezeigt haben, ist  $E(U) = 0$  und  $Var(U) = 1$ . Aus der Darstellung  $X = \mu + \sigma U$  ergibt sich mit (1)

$$E(X) = \mu + \sigma E(U) = \mu, \quad Var(X) = \sigma^2 Var(U) = \sigma^2. \quad (3)$$

Somit ist der *Lageparameter*  $\mu$  auch tatsächlich der *Erwartungswert* von  $X$  im Sinne der Definition aus **2.6.2**, und der *Skalenparameter*  $\sigma$  ist die *Standardabweichung* von  $X$  gemäß **2.6.3**.  $\square$

Der *Erwartungswert einer Summe* von Zufallsvariablen ist die *Summe ihrer Erwartungswerte*:

$$E(X + Y) = E(X) + E(Y). \quad (4)$$

Für (stochastisch) *unabhängige* Zufallsvariablen (vgl. **2.1.4-5**) ist auch die *Varianz der Summe* gleich der *Summe der Varianzen* und auch der *Erwartungswert des Produktes* gleich dem *Produkt der Erwartungswerte*:

Für *unabhängige* Zufallsvariablen  $X$  und  $Y$  gilt:

$$Var(X + Y) = Var(X) + Var(Y), \quad E(X \cdot Y) = E(X) \cdot E(Y). \quad (5)$$

Wenn  $X$  und  $Y$  *nicht* unabhängig sind, so gilt (5) im allgemeinen *nicht*, aber (4) gilt stets! In der Tat sind beide Gleichungen in (6) falsch z.B. für  $X$  als Augenzahl eines Würfels und  $Y = 2X$  als eine von  $X$  abhängige Zufallsvariable.

**Augensumme homogener Würfel:** Für die Augensumme  $X$  und  $Y$  zweier homogener Würfel (vgl. **2.2 Abb. 1**) gilt bei *gleichzeitigem Werfen* die Unabhängigkeit, also

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) &= 3,5 + 3,5 &= 7 \\ Var(X + Y) &= Var(X) + Var(Y) &= \frac{35}{12} + \frac{35}{12} &\approx 5,83 \end{aligned} \quad \square$$

**Bernoulli-Verteilung:** Für eine Zufallsvariable  $Y$  mit Bernoulli-Verteilung  $B(1, p)$  und Träger  $T = \{0, 1\}$  erhält man

$$\begin{aligned}\mu_Y &= E(Y) = 0 \cdot P\{Y = 0\} + 1 \cdot P\{Y = 1\} = p \\ E(Y^2) &= 0^2 \cdot P\{Y = 0\} + 1^2 \cdot P\{Y = 1\} = p \\ \sigma_Y^2 &= \text{Var}(Y) = E(Y^2) - [E(Y)]^2 = p - p^2 = pq \quad \text{mit } p = 1 - p.\end{aligned}\quad \square$$

**Binomial-Verteilung:** Sind  $Y_1, Y_2, \dots, Y_n$  unabhängige Zufallsvariablen jeweils mit  $B(1, p)$ -Verteilung, so ergibt sich für die Summe  $X = Y_1 + Y_2 + \dots + Y_n$  aus (4) (5)

$$\mu_X = E(X) = np, \quad \sigma_X^2 = \text{Var}(X) = npq \quad \text{mit} \quad q = 1 - p.$$

Da  $X$  eine  $B(n, p)$ -Verteilung hat, haben wir damit auch den *Erwartungswert* und die *Varianz* der *Binomial-Verteilung* bestimmt. Die Werte  $\mu_X$  und  $\sigma_X$  haben wir bei der Normalapproximation der Binomial-Verteilung in 2.5.6(1) bereits verwendet.  $\square$

Für eine  $N(\mu, \sigma^2)$ -verteilte Zufallsvariable  $X$  haben wir in 2.5.4 die Wahrscheinlichkeiten für Bereiche der Form  $\{|X - \mu| \leq k\sigma\}$  mit  $k = 1, 2, 3, 4$  berechnet und dadurch die *Konzentration der Verteilung um ihren Erwartungswert  $\mu$*  beschrieben. Für eine Zufallsvariable  $X$  mit beliebiger Verteilung lassen sich die Wahrscheinlichkeiten solcher Bereiche nach unten abschätzen durch die

**Ungleichung von Chebychev:**

Ist  $X$  eine Zufallsvariable mit  $\mu = E(X)$  und  $\sigma^2 = \text{Var}(X)$ , so gilt für jede Zahl  $c > 0$  die Abschätzung

$$P\{|X - \mu| \leq c\} \geq 1 - \frac{\sigma^2}{c^2} \quad \text{bzw.} \quad P\{|X - \mu| \geq c\} \leq \frac{\sigma^2}{c^2}. \quad (6)$$

Man beachte, daß die Ungleichung nur für solche Werte von Nutzen ist, für die die untere Schranke  $> 0$  bzw. die obere  $< 1$  ist. Für  $c = 2\sigma, 3\sigma, 4\sigma$  ergeben sich konkret die Abschätzungen

$$P\{|X - \mu| \geq 2\sigma\} \geq 1 - \frac{1}{4} = 75,00\% \quad (2\sigma - \text{Bereich}), \quad (7)$$

$$P\{|X - \mu| \geq 3\sigma\} \geq 1 - \frac{1}{9} = 88,89\% \quad (3\sigma - \text{Bereich}), \quad (8)$$

$$P\{|X - \mu| \geq 4\sigma\} \geq 1 - \frac{1}{16} = 93,750\% \quad (4\sigma - \text{Bereich}). \quad (9)$$

Die zugehörigen Wahrscheinlichkeiten bei *normalverteiltem*  $X$  sind zwar wesentlich größer (vgl. 2.5.4). Der Vorteil der Ungleichungen (7)-(9) liegt aber darin, daß sie für *jede beliebige* Verteilung von  $X$  gelten und deshalb für eine *konkrete* Verteilung keine besonders *scharfe* Abschätzung liefern (vgl. hierzu auch 2.4 Abb. 3, 2.5 Abb. 5).

Wenn  $\sigma = 0$  ist, so folgt aus der Chebychev-Ungleichung  $P\{|X - \mu| \geq c\} \leq 0$  für jedes  $c \leq 0$ , und daher nimmt  $X$  seinen Erwartungswert  $\mu$  mit Wahrscheinlichkeit 1 an:

$$\mu = E(X) \quad \text{und} \quad \text{Var}(X) = 0 \quad \Rightarrow \quad P\{X = \mu\} = 1. \quad (10)$$

Wenn  $\text{Var}(X) = 0$  gilt, so bezeichnet man die Verteilung von  $X$  auch als eine *Ein-Punkt-Verteilung in  $\mu = E(X)$* , weil  $X$  dann nicht „streut“, sondern mit Sicherheit (d.h. mit Wahrscheinlichkeit von 100%) den Wert  $\mu$  annimmt.

### 2.6.5 Schätzung von Erwartungswert und Standardabweichung

In Anwendungen ist bei einer konkreten Zufallsvariablen  $X$  (z.B. das *Gewicht eines Individuums* einer Population oder die *Lebensdauer eines Produktes* einer Fertigung) typischerweise weder der Erwartungswert  $\mu = E(X)$  noch die Varianz  $\sigma^2 = \text{Var}(X)$  bekannt. Da diese Parameter jedoch von primären Interesse sind, möchte man sie aus *beobachteten Daten* schätzen. Der Ausgangspunkt hierzu sind  $n$  *unabhängige Wiederholungen*  $X_1, X_2, \dots, X_n$  der Zufallsvariablen  $X$ . Die zugehörigen Realisierungen (beobachtete Ergebnisse)  $x_1, x_2, \dots, x_n$  bilden dann eine Stichprobe vom Umfang  $n$ .

Entsprechend der heuristischen Interpretation zu Beginn dieses Unterkapitels wird der Erwartungswert  $\mu$  von  $X$  durch den *Mittelwert der Stichprobe* geschätzt, und die Varianz  $\sigma^2$  bzw. Standardabweichung  $\sigma$  von  $X$  wird durch die (*empirische*) *Varianz* bzw. *Standardabweichung der Stichprobe* geschätzt (vgl. auch *Rechen-Box 1*):

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_i x_i \quad (\text{Schätzung von } \mu) \quad (1)$$

$$\hat{\sigma}^2 = s_x^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 = \frac{1}{n-1} S_{xx} \quad (\text{Schätzung von } \sigma^2) \quad (2)$$

$$\hat{\sigma} = s_x = \sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} S_{xx}} \quad (\text{Schätzung von } \sigma) \quad (3)$$

Wir haben hier auch gleich neue Bezeichnungen eingeführt. Und zwar haben wir – wie in der Statistik allgemein üblich – die *Schätzung* des Parameters  $\mu$  bzw.  $\sigma$  durch ein zusätzliches  $\hat{\cdot}$  in  $\hat{\mu}$  bzw.  $\hat{\sigma}$  gekennzeichnet, um anzudeuten, daß es sich hier nicht um den *wahren Wert*, sondern um eine *Schätzung* für den wahren Wert handelt.

### Eigenschaften der Schätzung des Erwartungswertes

Der *geschätzte* Wert  $\hat{\mu} = \bar{x}$  ist nicht der *wahre* Erwartungswert  $\mu$ , sondern ist wieder die Realisierung der *Zufallsvariablen*

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n} \sum_i X_i \quad (\text{Stichproben-Mittelwert}). \quad (4)$$

Aus den Eigenschaften in **2.6.4** ergibt sich, daß der Stichproben-Mittelwert  $\bar{X}$  denselben *Erwartungswert* wie die Zufallsvariable  $X$  hat, aber seine Varianz ist um den Faktor  $\frac{1}{n}$  geringer:

$$E(\bar{X}) = E(X) = \mu \quad (5)$$

$$Var(\bar{X}) = \frac{1}{n} \cdot Var(X) = \frac{1}{n} \cdot \sigma^2 \quad \text{bzw.} \quad (6)$$

$$SD(\bar{X}) = \sqrt{\frac{1}{n}} \cdot SD(X) = \frac{1}{\sqrt{n}} \cdot \sigma .$$

Die Bildung des Mittelwerts  $\bar{X}$  *verringert die Streuung* bzw. *erhöht die Genauigkeit* gegenüber dem Einzelwert  $X$ . Dies ist der Grund, warum man *Mittelwerte* über unabhängige Wiederholungen bildet anstatt nur einen *Einzelwert* zu bestimmen. Für *wachsenden* Stichprobenumfang  $n$  strebt die Standardabweichung  $SD(\bar{X})$  des Stichproben-Mittelwerts  $\bar{X}$  sogar gegen Null (vgl. Abb. 1), d.h.  $\bar{X}$  streut immer geringer um der wahren Wert  $\mu$ .

### Approximative Normalverteilung des Stichproben-Mittelwerts

Wenn die Zufallsvariable  $X$  eine *Normalverteilung*  $N(\mu, \sigma^2)$  besitzt, so ergibt sich aus den Eigenschaften der Normalverteilung (in **2.5.4**), daß der Mittelwert  $\bar{X}$  die Normalverteilung  $N(\mu, \frac{1}{n}\sigma^2)$  besitzt. Folglich hat seine Standardisierung

$$U = \frac{\bar{X} - E(\bar{X})}{SD(\bar{X})} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \quad (7)$$

eine Standard-Normalverteilung  $N(0, 1)$ . Das überraschende ist, daß auch bei *nicht normalverteiltem*  $X$  der Mittelwert  $\bar{X}$  stets *approximativ* (d.h. in guter Näherung) eine *Normalverteilung*  $N(\mu, \frac{1}{n}\sigma^2)$  hat, wobei die Genauigkeit der Approximation mit *wachsendem* Stichprobenumfang  $n$  besser wird. Präziser gesagt strebt die Verteilung der Standardisierung  $U$  (nach dem *Zentralen Grenzwertsatz der Wahrscheinlichkeitstheorie*) bei wachsendem Stichprobenumfang  $n \rightarrow \infty$  gegen die  $N(0, 1)$ -Verteilung, und zwar *unabhängig davon, welche Verteilung die ursprüngliche Zufallsvariable X hat* (vgl. Abb. 2). Dies demonstriert die Bedeutung der Normalverteilung und bildet eine wichtige Grundlage für viele statistische Verfahren.

### Eigenschaften der Varianz-Schätzung

Der geschätzte Wert  $\hat{\sigma}^2 = s_x^2$  ist nicht die wahre Varianz  $\sigma^2$ , sondern ist die Realisierung folgender Zufallsvariablen

$$S_X^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 \quad (\text{mittlere quadratische Abweichung vom Mittel } \bar{X}). \quad (8)$$

Der Erwartungswert dieser mittleren quadratischen Abweichung ist gerade die Varianz von  $X$ :

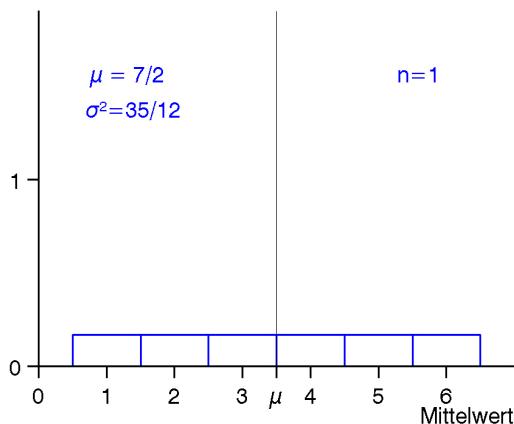
$$E(S_X^2) = Var(X) = \sigma^2. \quad (9)$$

Das bedeutet, daß die Schätzung  $\hat{\sigma}^2 = s_x^2$  im Durchschnitt den wahren Wert  $\sigma^2$  liefert. Diese wichtige Eigenschaft ist der Grund dafür, daß man in der Definition der Stichproben-Varianz  $s_x^2$  durch  $n-1$  teilt (würde man statt dessen durch  $n$  teilen, so wäre diese Schätzung im Durchschnitt *systematisch kleiner* als das wahre  $\sigma$ ).

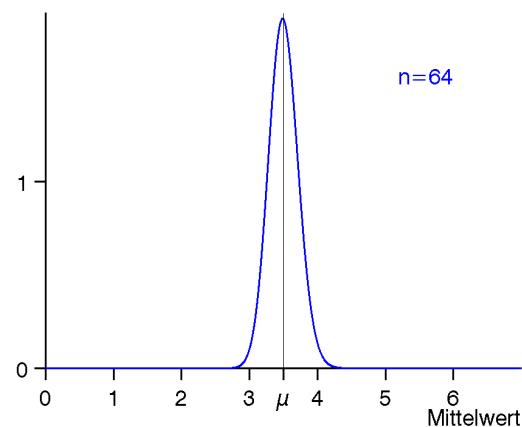
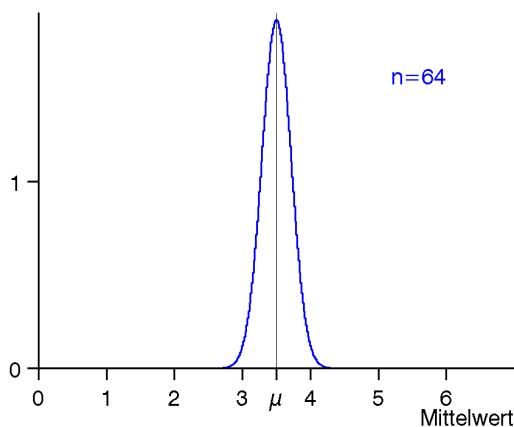
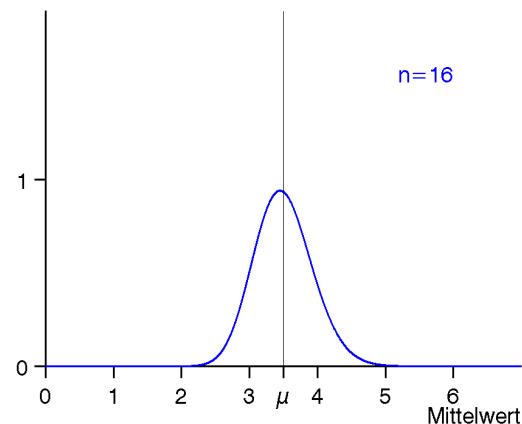
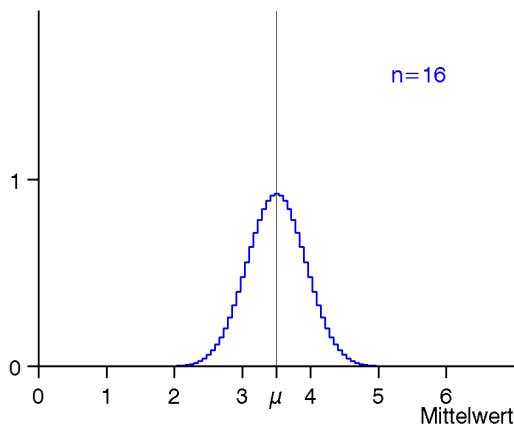
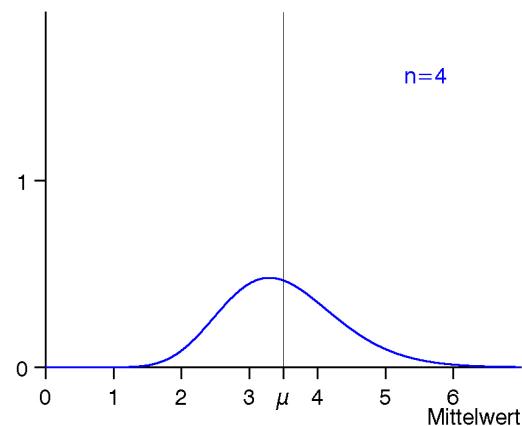
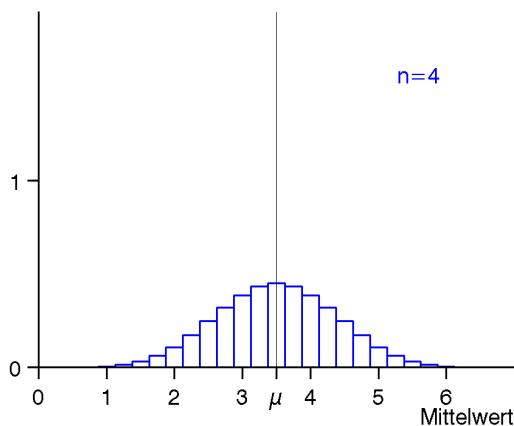
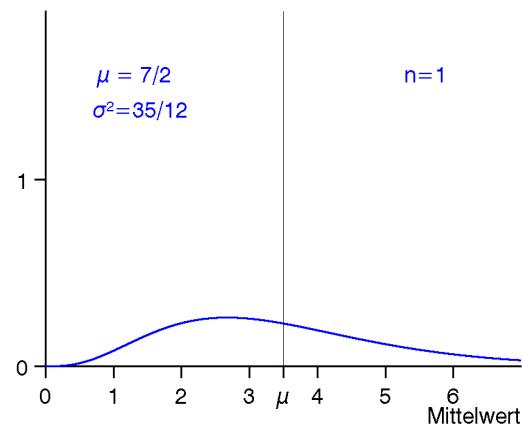
Für wachsendes  $n$  strebt auch die Standardabweichung von  $S_X^2$  gegen Null, d.h.  $S_X^2$  streut immer weniger um die wahre Varianz  $\sigma^2$ .

**Abb. 1:** Die Verteilung des Mittelwerts  $\bar{X}$  streut mit wachsendem Umfang  $n$  immer weniger um den Erwartungswert  $\mu$  und geht in eine Normalverteilung über.

$X$  ist **diskret** (z.B. homogener Würfel)

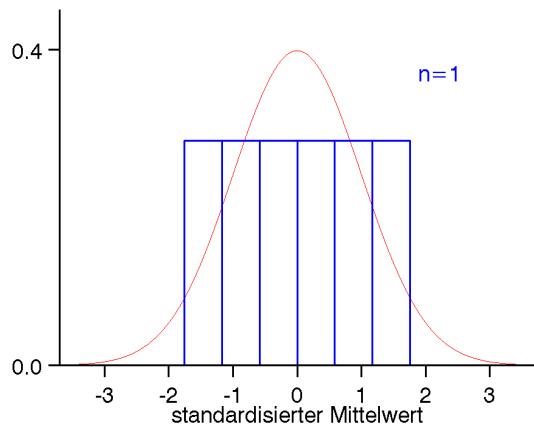


$X$  ist **stetig** (z.B. Gamma-verteilt)

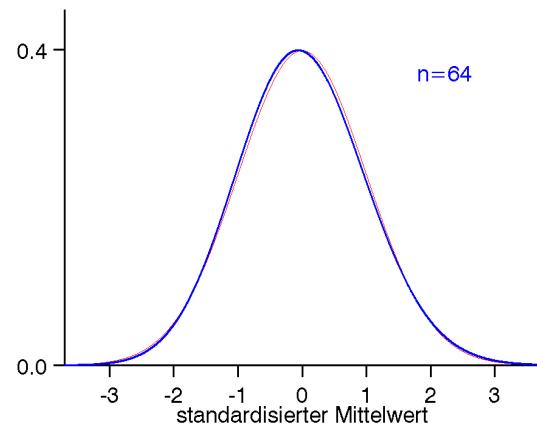
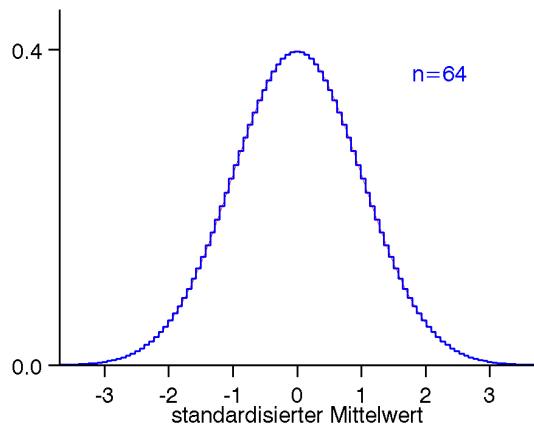
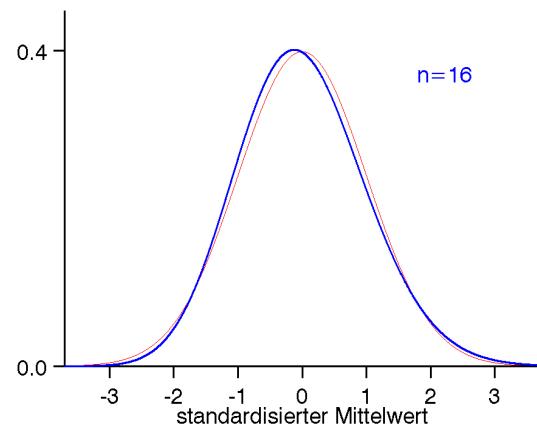
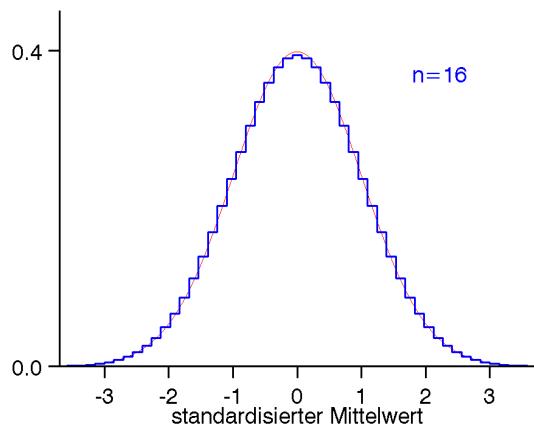
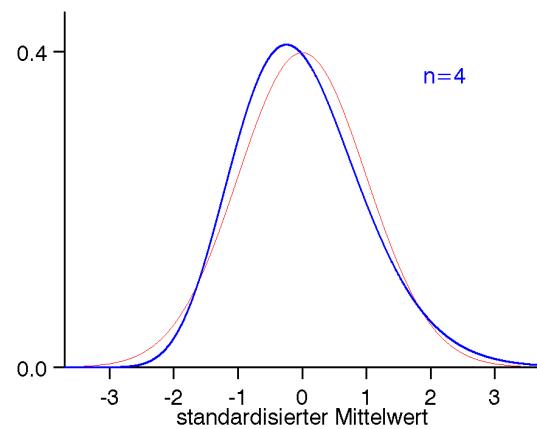
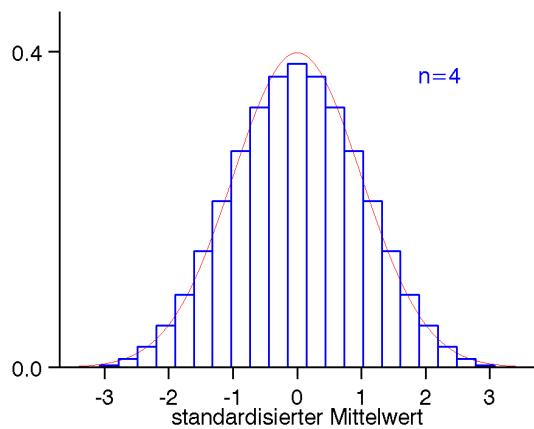
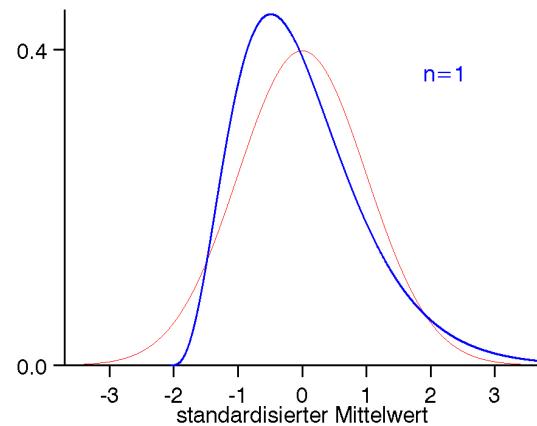


**Abb. 2:** Die Verteilung der Standardisierung  $U = \frac{1}{\sigma}\sqrt{n}(\bar{X} - \mu)$  des Mittelwerts  $\bar{X}$  (vgl. Abb. 1) im Vergleich mit der Standard-Normalverteilung  $N(0, 1)$  (dünne Linie).

$X$  ist **diskret** (z.B. homogener Würfel)



$X$  ist **stetig** (z.B. Gamma-verteilt)



## 2.7 Die Poisson-Verteilung

### Definition

Wir wollen jetzt noch eine weitere wichtige diskrete Verteilung kennenlernen, die *Poisson-Verteilung*  $\text{Pois}(\mu)$  mit Parameter  $\mu > 0$ . Eine Zufallsvariable  $X$ , deren Werte im Bereich  $T = \{0, 1, 2, \dots\}$  aller nicht-negativen ganzen Zahlen liegen, heißt  $\text{Pois}(\mu)$ -verteilt, wenn sie ihre Werte mit folgenden Wahrscheinlichkeiten annimmt

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

Man beachte, daß der Träger  $T$  der Poisson-Verteilung *unendlich* viele Ergebnisse enthält. Der Parameter  $\mu$  ist sowohl der *Erwartungswert* als auch die Varianz von  $X$

$$E(X) = \mu, \quad \text{Var}(X) = \mu. \quad (2)$$

Die Poisson-Wahrscheinlichkeiten (1) lassen sich schrittweise für  $k = 0, 1, 2, \dots$  wie folgt berechnen

$$P\{X = 0\} = e^{-\mu} \quad (3)$$

$$P\{X = x\} = \frac{\mu}{x} \cdot P\{X = x - 1\} \quad \text{für } x \geq 1. \quad (4)$$

Die Folge dieser Wahrscheinlichkeiten  $P\{X = x\}$  ist für  $x = 0, 1, 2, \dots$  zuerst *anwachsend*, solange  $x \leq \mu$  gilt, und dann für  $x \geq \mu$  wieder *abfallend* (vgl. Abb. 1, 2). Wenn  $x$  gegenüber  $\mu$  anwächst, so fällt die Wahrscheinlichkeit  $P\{X = x\}$  extrem schnell gegen Null ab (obwohl sie niemals *exakt* gleich Null ist).

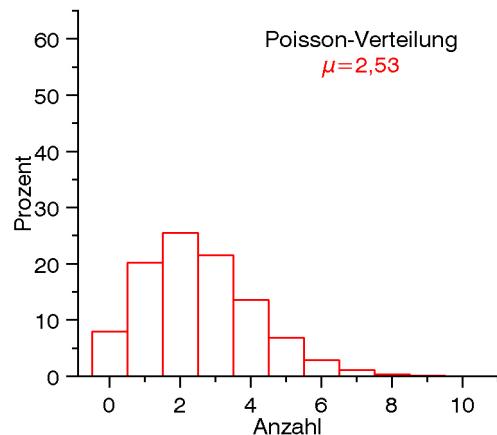
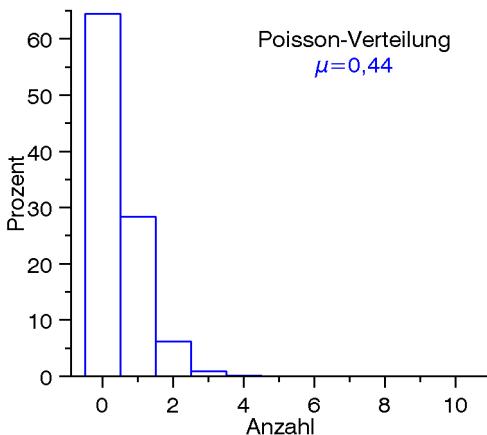


Abb.. 1: Histogramme der Poisson-Verteilung für zwei verschiedene Erwartungswerte

## Auftreten

Die Poisson-Verteilung wird als Modell verwendet für die Verteilung einer *Anzahl*  $X$  von *Eintritten* eines interessierenden Ereignisses in einer Grundgesamtheit (und innerhalb eines Zeitraums), wie z.B. die Anzahl von Mutationen in einer Zellkultur oder die Anzahl der Leukämieerkrankungen von Kindern einer Region (innerhalb eines Jahres). Hierbei müssen allerdings die einzelnen Ereignisse *zufällig* und *unabhängig voneinander* eintreten, was z.B. für Grippe-Erkrankungen in einer Region nicht zutrifft, weil diese sich durch Ansteckung ausbreiten und somit nicht *unabhängig voneinander* auftreten.

In der Praxis können solche (zufälligen) Anzahlen  $X$  von Ereignissen nicht beliebig große Zahlen annehmen, weil die Grundgesamtheit (z.B. die Anzahl aller Zellen einer Kultur oder die Anzahl aller Kinder einer Region) *beschränkt* ist. Dennoch kann man hier in guter Näherung von einer Poisson-Verteilung ausgehen, weil die Poisson-Wahrscheinlichkeiten  $P\{X = x\}$  extrem klein werden, wenn  $x$  groß wird (im Verhältnis zu  $\mu$ ), so daß man sie bei praktischen Berechnungen vernachlässigen kann.

**Leukämiefälle im Umkreis des Kernkraftwerks Krümmel:** Der Kinderarzt M. Demuth berichtet in dem Preprint *Leukämiemorbidität bei Kindern in der direkten Umgebung des Kernkraftwerks Krümmel* (Kassel 1991, Tabelle 1) von Leukämie-Erkrankungsfällen bei Kindern von 0 bis 14 Jahren in der Umgebung des Kerkraftwerks Krümmel und vergleicht die *beobachteten* Leukämiefälle mit den (unter Berücksichtigung der Bevölkerungsdichte) nach bundesdeutschem Durschnitt zu *erwartenden* Fällen (vgl. Tab. 1).

Die Anzahl  $X$  der Leukämiefälle einer Region in einem festen Zeitraum sind in guter Näherung *Pois( $\mu$ )-verteilt*, wobei  $\mu = E(X)$  die *erwartete* Zahl der Leukämiefälle ist. Für den *5-km-Radius um das Kraftwerk Krümmel* aus Tab. 1 ist die zugehörige Poisson-Verteilung in Abb. 1 dargestellt, und zwar *links* für den Zeitraum 1990 – 1991 mit  $\mu = 0,44$  und *rechts* für 1980 – 1991 mit  $\mu = 2,53$ . Man erkennt daran, daß die beobachteten Fälle (5 bzw. 4) unter der Annahme des bundesdurchschnittlichen Erwartungswerts  $\mu$  relativ unwahrscheinlich sind.

Region	Zeitraum	Läukemiefälle	
		beobachtet	erwartet
Samtgemeinde Elbmarsch	1990	3	0.06
	1990-1991	4	0.13
	1980-1990	4	0.68
	1980-1991	5	0.75
5-km-Radius um das Kernkraftwerk Krümmel	1990	3	0.21
	1990-1991	5	0.44
	1980-1990	4	2.30
	1980-1991	6	2.53

Tabelle 1: Leukämieerkrankungsfälle bei Kindern von 0–14 Jahren in der Umgebung des Kernkraftwerks Krümmel (nach Demuth 1991).

### Normal-Approximation der Poisson-Verteilung

Für *wachsenden* Erwartungswert  $\mu$  strebt die Poisson-Verteilung gegen die Normalverteilung  $N(\mu, \sigma^2)$  mit der Varianz  $\sigma^2 = \mu$  (vgl. Abb. 2). Für *großes*  $\mu$  ergeben sich hieraus unter Verwendung der Verteilungsfunktion  $\Phi$  von  $N(0, 1)$  folgende

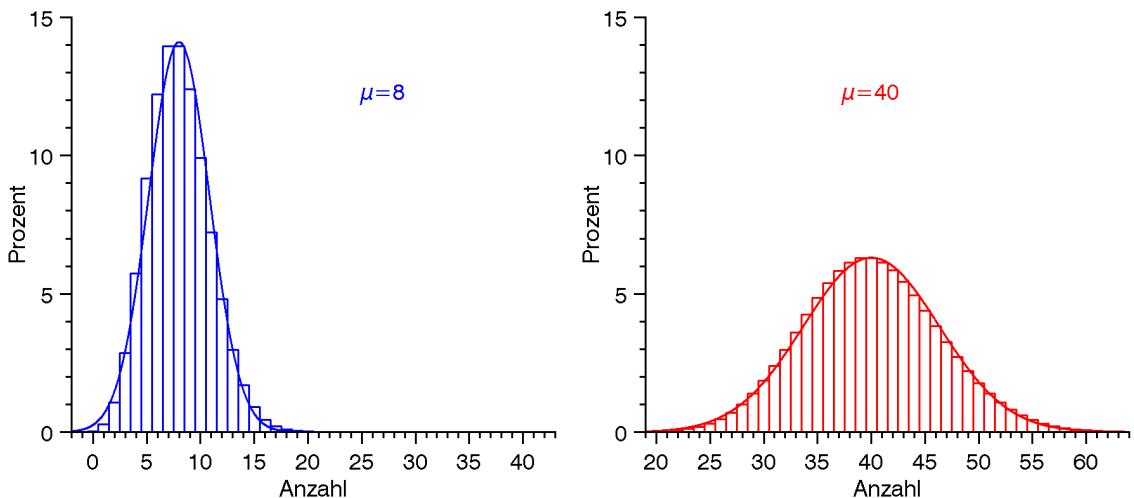
**Normal-Approximationen für  $Pois(\mu)$ -verteiltes  $X$ :**

$$P\{X \leq k\} \approx \Phi(u) \quad \text{mit } u = \frac{1}{\sqrt{\mu}} \left( k + \frac{1}{2} - \mu \right), \quad (5)$$

$$P\{X \geq m\} \approx \Phi(-v) = 1 - \Phi(v) \quad \text{mit } v = \frac{1}{\sqrt{\mu}} \left( m - \frac{1}{2} - \mu \right), \quad (6)$$

$$P\{m \leq X \leq k\} \approx \Phi(u) - \Phi(v) \quad \text{mit } u, v \text{ wie oben.} \quad (7)$$

Aus dem gleichen Grund wie bei der Normal-Approximation der Binomialverteilung in 2.5.6 taucht hier in  $u$  bzw.  $v$  wieder die *Stetigkeitskorrektur* von  $\pm \frac{1}{2}$  auf. Diese Normal-Approximationen werden für *wachsendes*  $\mu$  immer genauer.



**Abb. 2:** Histogramme der  $Pois(\mu)$ -Verteilungen und zugehörige Dichte der approximierenden  $N(\mu, \sigma^2)$ -Verteilung mit  $\sigma^2 = \mu$ .

### Zusammenhang zur Binomial-Verteilung

Es besteht ein enger Zusammenhang zwischen der Poisson-Verteilung  $Pois(\mu)$  und der Binomialverteilung  $B(n, p)$ , die ja ebenfalls ein Modell für eine zufällige Anzahl  $X$  von Eintritten eines Ziel-Ereignisses (*innerhalb von n unabhängigen Wiederholungen*) ist. Bei *großem* Umfang  $n$  und *kleiner* Wahrscheinlichkeit  $p$  (für das Eintreten des Ziel-Ereignisses) lässt sich die  $B(n, p)$ -Verteilung durch die Poisson-Verteilung  $Pois(\mu)$  mit dem Erwartungswert  $\mu = np$  der Binomialverteilung approximieren

$$B(n, p) \approx Pois(\mu) \quad \text{mit } \mu = np, \quad \text{falls } n \text{ groß und } p \text{ klein ist.} \quad (8)$$

(Man beachte hierzu auch die Ähnlichkeit von Abb. 1 mit **2.5 Abb. 6.**) Deshalb wird in solchen Situationen auch die Poisson-Verteilung anstelle der Binomial-Verteilung verwendet, insbesondere dann wenn man sich primär für die *erwartete* Anzahl  $\mu$  der Eintritte interessiert und weniger für die Eintritts-Wahrscheinlichkeit  $p$ .

### Tipps für *Microsoft Excel*

*Funktionen:* Für  $Pois(\mu)$ -verteiltes  $X$  stehen in *Excel* folgende *Statistik-Funktionen* direkt zur Verfügung:

$$\begin{aligned}\text{POISSON}(x; \mu; 0) &= P\{X = x\} \\ \text{POISSON}(x; \mu; 1) &= P\{X \leq x\}.\end{aligned}$$

Das *letzte* Argument der Funktion **POISSON** ist der Indikator dafür, ob es sich um *kumulierte* Wahrscheinlichkeiten handelt (1) oder nicht (0).

### 3 Deduktive Statistik

In der *deduktiven (schließenden) Statistik* geht man von der Modell-Vorstellung aus, daß die beobachteten Daten *Realisierungen von Zufallsvariablen* sind, und will dann Rückschlüsse über die *unbekannten Parameter der Verteilung* dieser Zufallsvariablen ziehen. Wir werden uns hier in erster Linie mit der statistischen Analyse von *Erwartungswerten* (als wichtigstem Parameter einer Verteilung) und *Wahrscheinlichkeiten* (für interessierende Ziel-Ereignisse) beschäftigen. Wir beginnen dabei mit dem *Schätzen* dieser Parameter aus beobachteten Daten und der Konstruktion von *Konfidenzgrenzen (Vertrauensgrenzen)*. Im Anschluß daran behandeln wir das *Testen von Hypothesen* über Modell-Parameter. Da es hier nicht möglich ist alle relevanten Verfahren vorzustellen, werden *exemplarisch* einige wichtige Methoden behandelt und wir verweisen ansonsten auf die Fachliteratur. Konkret behandeln wir ausführlich das *Schätzen* und *Testen* von *Wahrscheinlichkeiten* und *Erwartungswerten*, inklusive dem *Vergleich* zweier Wahrscheinlichkeiten bzw. Erwartungswerte. Die aus Kapitel 1.5 bekannte lineare Regression wird hier unter dem Aspekt des Schätzens und Testens der unbekannten Parameter der Regressionsgeraden erneut aufgegriffen. Weiter werden für diskrete Zufallsvariablen *Anpassungstests* und *Vergleiche (zweier Verteilungen)* sowie *Tests auf Unabhängigkeit* behandelt.

### 3.1 Schätzung einer Wahrscheinlichkeit mit Konfidenzgrenzen

Bei Anwendungen gilt das Hauptinteresse oft einem speziellen *Ziel-Ereignis A* (z.B. dem *Heilungs-Erfolg* einer neuen Therapie), dessen unbekannte *Eintritts-Wahrscheinlichkeit*  $p = P(A)$  *geschätzt* werden soll. Zur formalen Beschreibung betrachten wir die Indikatorvariable  $Y$  für das Ereignis  $A$

$$Y = \begin{cases} 1 & \text{falls Ziel-Ereignis } A \text{ eintritt} \\ 0 & \text{falls Ziel-Ereignis } A \text{ nicht eintritt:} \end{cases} \quad \begin{array}{l} (\text{z.B. Erfolg}) \\ (\text{z.B. kein Erfolg}). \end{array}$$

Ausgangspunkt sind nun  $n$  *unabhängige* Wiederholungen  $Y_1, \dots, Y_n$  der Zufallsvariablen  $Y$  (die sich z.B. aus  $n$  unabhängigen Behandlungen mit der neuen Therapie ergeben). Die *absolute Häufigkeit* des Eintretens von  $A$  unter den  $n$  Wiederholungen ist die  $B(n, p)$ -verteilte Zufallsvariable

$$X = Y_1 + Y_2 + \dots + Y_n \quad (\text{absolute Häufigkeit, Anzahl der Eintritte}).$$

Die *beobachteten* Daten – die den statistischen Analysen zugrunde liegen – sind die Realisierungen  $y_i$  von  $Y_i$  für  $i = 1, \dots, n$  sowie die *beobachtete Anzahl*  $x = y_1 + \dots + y_n$  (von Eintreten) als Realisierung von  $X$ .

Bei den folgenden Betrachtungen spielt nur noch die  $B(n, p)$ -verteilte Zufallsvariable  $X$  und ihre Realisierung  $x$  eine Rolle (die Indikatorvariablen dienten nur der näheren Erläuterung und man könnte sie jetzt wieder „vergessen“).

#### 3.1.1 Schätzen einer Wahrscheinlichkeit

Aus der beobachteten Realisierung  $x$  einer  $B(n, p)$ -verteilten Zufallsvariablen  $X$  soll zunächst die unbekannte Wahrscheinlichkeit  $p$  *geschätzt* werden. Entsprechend der Interpretation einer Wahrscheinlichkeit (vgl. 2.1.1) verwendet man als Schätzwert für  $p$  die *relative Häufigkeit* der Eintritte unter den  $n$  Wiederholungen

$$\hat{p} = \hat{p}(x) = \frac{1}{n} x \quad (\text{Schätzwert für } p, \text{ beobachtete relative Häufigkeit}). \quad (1)$$

Die Schreibweise  $\hat{p}(x)$  verwenden wir, wenn die Abhängigkeit von der Beobachtung  $x$  betont werden soll. Ansonsten schreiben wir (insbesondere bei Anwendungen) nur kurz  $\hat{p}$ . Der geschätzte Wert  $\hat{p}(x)$  ist nicht die *wahre* Wahrscheinlichkeit  $p$ , sondern die Realisierung der zugehörigen *Zufallsvariablen*

$$\hat{p}(X) = \frac{1}{n} X \quad (\text{Schätzvariable für } p, \text{ zufällige relative Häufigkeit}). \quad (2)$$

Um die Eigenschaften der Schätzung  $\hat{p}(x)$  zu studieren, betrachtet man die *Verteilung* der Schätzvariablen  $\hat{p}(X)$ . Da sich die Zufallsvariable  $\hat{p}(X)$  nur um den Faktor  $\frac{1}{n}$  von  $X$  unterscheidet, ergibt sich die *Verteilung* von  $\hat{p}(X)$  direkt aus der  $B(n, p)$ -Verteilung von  $X$  zu

$$P\{\hat{p}(X) = \frac{x}{n}\} = P\{X = x\} = b(x | n, p) \quad \text{vgl. auch 2.3.1(2).} \quad (3)$$

Das Histogramm der Verteilung von  $\hat{p}(X)$  unterscheidet sich daher vom Histogramm der  $B(n, p)$ -Verteilung nur durch eine *andere Skalierung der x-Achse* (Abb. 1).

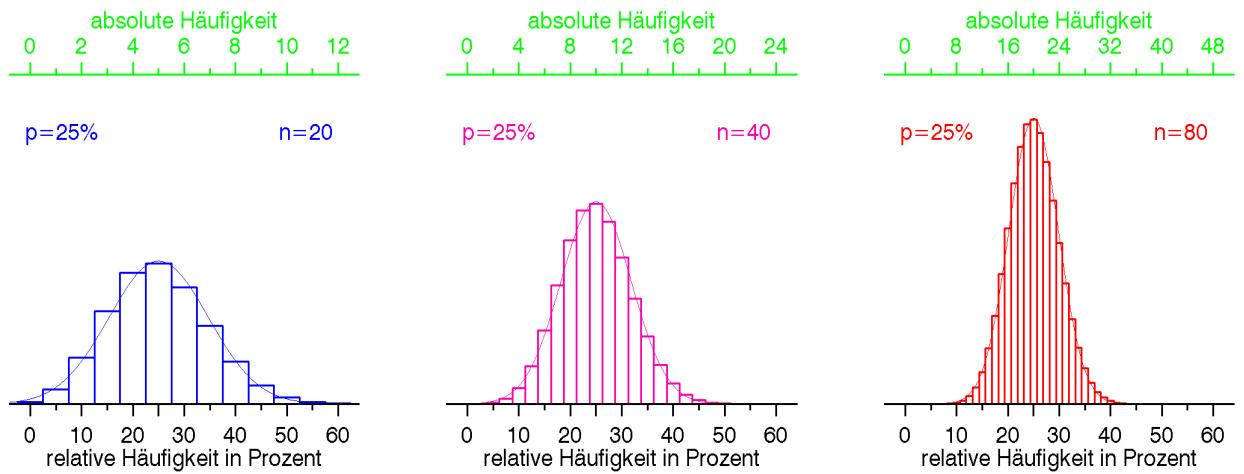


Abb. 1: Gemeinsame Darstellung der  $B(n, p)$ -verteilten absoluten Häufigkeit  $X$  (obere Achse) und der Verteilung der relativen Häufigkeit  $\hat{p}(X)$  (untere Achse) als Histogramm für  $p = 25\%$  und verschiedene Umfänge  $n = 20$  (links),  $40$  (Mitte) und  $80$  (rechts). Für größeres  $n$  streut die Verteilung von  $\hat{p}(X)$  geringer um  $p$ . Die glatte Kurve ist die Dichte der approximierenden Normalverteilung.

Aus dem Erwartungswert und der Varianz von  $X$  (vgl. 2.6.4)

$$E(X) = np, \quad \text{Var}(X) = npq \quad \text{mit} \quad q = 1 - p \quad (4)$$

ergeben sich der Erwartungswert und die Varianz von  $\hat{p}(X)$  zu

$$E(\hat{p}(X)) = p, \quad \text{Var}(\hat{p}(X)) = \frac{1}{n} pq \quad \text{mit} \quad q = 1 - p. \quad (5)$$

Die Zufallsvariable  $\hat{p}(X)$  streut daher „im Mittel“ um die wahre Eintritts-Wahrscheinlichkeit  $p$ , wobei die Streung (Varianz) bei *wachsendem* Umfang  $n$  geringer wird und für  $n \rightarrow \infty$  sogar gegen 0 strebt.

Man beachte, daß die Varianz von  $\hat{p}(X)$  (ebenso wie die von  $X$ ) auch von der Wahrscheinlichkeit  $p$  abhängt. Bei *festem*  $n$  ist die Varianz *maximal* für  $p = \frac{1}{2}$  (d.h. in der Mitte des

Intervalls von 0 bis 1) und fällt an den Rändern (d.h. für  $p \rightarrow 0$  oder  $p \rightarrow 1$ ) sogar bis auf Null ab. Ersetzt man in der Formel für die Varianz von  $\hat{p}(X)$  die unbekannte Wahrscheinlichkeit  $p$  durch ihre Schätzung  $\hat{p}$ , so erhält man die geschätzte Varianz  $\frac{1}{n} \hat{p}(1 - \hat{p})$ . Der Kehrwert dieser Varianz-Schätzung entspricht genau dem im **1.6.7(5)** verwendeten *Gewichtsfaktor* einer relativen Häufigkeit (die dort mit  $y$  bezeichnet wurde und hier  $\hat{p}$  entspricht). Und dies erklärt die Wahl der dortigen Gewichtsfaktoren.

Aus der *Normal-Approximation* der Binomialverteilung (vgl. **2.5.6**) ergibt sich, daß  $\hat{p}(X)$  für nicht zu kleinen Umfang  $n$  näherungsweise  $N(p, \frac{1}{n} pq)$ -verteilt ist (vgl. Abb. 1).

**Für Fortgeschrittene:** Die Indikatorvariable  $Y$  für das Ziel-Ereignis  $A$  ist  $B(1, p)$ -verteilt mit *Erwartungswert*  $\mu_Y = p$ . Die Schätzung  $\hat{\mu}_Y$  dieses Erwartungswerts nach **2.6.5** ist der Mittelwert  $\bar{y} = \frac{1}{n} \sum_i y_i = \frac{1}{n} x$  und stimmt mit der relativen Häufigkeit  $\hat{p}(x)$  überein. In diesem Sinne ist die Schätzung der *Wahrscheinlichkeit*  $p$  nur ein Spezialfall der Schätzung eines *Erwartungswerts* aus **2.6.5**.

### 3.1.2 Einseitige exakte Konfidenzgrenzen für eine *Wahrscheinlichkeit* (nach *Pearson-Clopper*)

Unser Ausgangspunkt ist wieder die Realisierung (Beobachtung)  $x$  einer  $B(n, p)$ -verteilten Zufallsvariablen  $X$ . Zusätzlich zur *Schätzung*  $\hat{p}(x)$  der unbekannten Wahrscheinlichkeit  $p$  wollen wir jetzt noch eine untere bzw. obere *Konfidenzgrenze* (*Vertrauengrenze*)  $\hat{p}_{un}(x)$  bzw.  $\hat{p}_{ob}(x)$  konstruieren. Diese Grenzen sollen die unbekannte Wahrscheinlichkeit  $p$  mit einer *vorgegebenen Sicherheit nach unten* bzw. *oben abschätzen*. Damit ist folgendes gemeint. Die Grenze  $\hat{p}_{un}(x)$  kann *unterhalb* von  $p$  liegen (und damit  $p$  wie gewünscht nach *unten abschätzen*) oder *oberhalb* (was unerwünscht ist). Berücksichtigt man nun, daß die untere Grenze  $\hat{p}_{un}(x)$  mit der Beobachtung  $x$  zufällig variiert – so läßt sich die Forderung an die *untere Grenze* dadurch beschreiben, daß die zugehörige Zufallsvariable  $\hat{p}_{un}(X)$  mit der geforderten *Sicherheit unterhalb* von  $p$  bzw. mit der entsprechenden *Irrtums-Wahrscheinlichkeit oberhalb* von  $p$  liegen soll:

$$\begin{aligned} P\{\hat{p}_{un}(X) < p\} &= \text{Sicherheits-Wahrscheinlichkeit} && \text{bzw.} \\ P\{\hat{p}_{un}(X) \geq p\} &= \text{Irrtums-Wahrscheinlichkeit}. \end{aligned} \quad (1)$$

Analog soll die *obere Grenze* mit der geforderten Sicherheit *oberhalb* von  $p$  liegen:

$$\begin{aligned} P\{p < \hat{p}_{ob}(X)\} &= \text{Sicherheits-Wahrscheinlichkeit} && \text{bzw.} \\ P\{p \geq \hat{p}_{ob}(X)\} &= \text{Irrtums-Wahrscheinlichkeit}. \end{aligned} \quad (2)$$

Die vorgegebene *Irrtumswahrscheinlichkeit* wird generell mit  $\alpha$  bezeichnet und *routinemäßig* auf  $\alpha = 5\%$  gesetzt. In begründeten Situationen werden auch *größere* Werte (maximal jedoch  $\alpha = 10\%$ ) oder *kleinere* Werte (z.B.  $\alpha = 1\%$  oder geringer) verwendet. Die

zugehörige *Sicherheit* ist dann  $1 - \alpha$ . Wir wollen zuerst die sogenannten *exakten* Konfidenzgrenzen konstruieren, die auf *exakten* Berechnungen der Wahrscheinlichkeiten in (1) und (2) basieren und erst danach die *normalen* (sowie *groben*) Konfidenzgrenzen behandeln, die auf der *Normal-Approximation* der Binomialverteilung beruhen.

Die zur Berechnung der jeweiligen Konfidenzgrenzen erforderlichen Schritte sind im Abschnitt **3.1.4** in der *Konfidenz-Box 1* übersichtlich zusammengestellt.

### 3.1.2 a Die *exakte untere einseitige* Konfidenzgrenze

Wir beginnen mit der Konstruktion der unteren Grenze und betrachten hierzu die Binomial-Wahrscheinlichkeit für die Beobachtung  $x$  und alle *größeren Werte*

$$P\{X \geq x \mid p\} = \sum_{i=x}^n b(i \mid n, p), \quad (1)$$

wobei wir die Abhängigkeit von  $p$  explizit mit in die Notation aufnehmen. Da wir den Wert von  $p$  nicht kennen, stellen wir uns verschiedene ‘‘Szenarien’’ für  $p$  vor, wobei wir zuerst den Fall  $x > 0$  betrachten. Die Wahrscheinlichkeit  $P\{X \geq x \mid p\}$  fällt wenn  $p$  fällt bzw. steigt wenn  $p$  steigt (vgl. in Abb. 2 jeweils das dünne mit dem fetten Histogramm). Es gibt nun genau einen Wert für  $p$ , bei dem diese Wahrscheinlichkeit gleich  $\alpha$  ist. Dieser Wert ist die gesuchte *exakte untere Konfidenzgrenze*  $\hat{p}_{un,\alpha}(x)$  zur Sicherheit  $1 - \alpha$  und er ist *implizit* gegeben als Lösung der Gleichung

$$P\{X \geq x \mid \hat{p}_{un,\alpha}(x)\} = \alpha \quad \text{für } x > 0. \quad (2)$$

Auf die konkrete Berechnung von  $\hat{p}_{un,\alpha}(x)$  gehen wir erst im Abschnitt **3.1.2 d** ein.

Für den bisher nicht betrachteten Fall  $x = 0$  ist die Schätzung  $\hat{p}(0) = 0$  und die Wahrscheinlichkeit  $P\{X \geq 0 \mid p\}$  ist unabhängig von  $p$  stets gleich 1. Es ist daher sinnvoll, die untere Konfidenzgrenze ebenfalls auf 0 zu setzen, d.h.

$$\hat{p}_{un,\alpha}(0) = 0. \quad (3)$$

Die untere Grenze  $\hat{p}_{un,\alpha}(x)$  ist damit (auch im Fall  $x = 0$ ) der *kleinste* Wert für  $p$ , unter dem die Wahrscheinlichkeit  $P\{X \geq x \mid p\}$  noch *mindestens*  $\alpha$  beträgt:

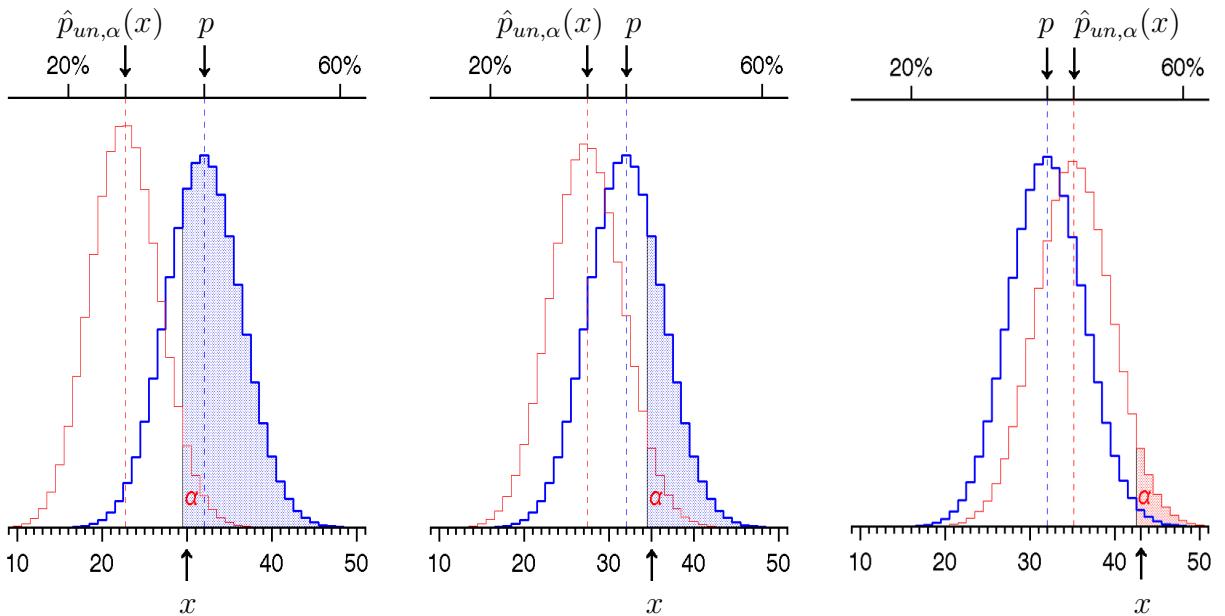
$$\hat{p}_{un,\alpha}(x) = \text{Min } \{p \mid P\{X \geq x \mid p\} \geq \alpha\} \quad (\text{exakte untere einseitige Konfidenzgrenze}) \quad (4)$$

In Worten: *die untere Grenze ist der kleinstmögliche Wert für p, unter dem das (eingetretene) Ereignis {X ≥ x} noch eine Eintritts-Wahrscheinlichkeit von mindestens α hat*. Noch kleinere Werte von  $p$  als  $\hat{p}_{un,\alpha}(x)$  werden also deshalb „abgelehnt“, weil unter ihnen das eingetretene Ereignis  $\{X \geq x\}$  zu „unwahrscheinlich“ wäre, d.h. eine Eintritts-

Wahrscheinlichkeit von *weniger* als  $\alpha$  hätte (entsprechend dem Motto: Das kann kein Zufall mehr sein!).

Die so konstruierte untere Grenze erfüllt unsere Forderung **3.1.2 (1)**, wobei die geforderte Sicherheit zwar nicht *genau erreicht*, aber dafür *mindestens* garantiert ist:

$$P\{\hat{p}_{un,\alpha}(X) < p\} \geq 1 - \alpha \quad \text{bzw.} \quad P\{p \leq \hat{p}_{un,\alpha}(X)\} \leq \alpha. \quad (5)$$



*Abb. 2:* Zur Konstruktion der *exakten unteren* Konfidenzgrenze von  $p$  für eine  $B(n, p)$ -verteilte Zufallsvariable  $X$ , wobei  $\alpha = 5\%$  ist. Das **fette** Histogramm entspricht der Verteilung für  $n = 80$  mit dem *wahren* Wert  $p = 40\%$ . Für drei verschiedene Realisierungen  $x = 30$  (links),  $x = 35$  (Mitte) und  $x = 43$  (rechts) zeigt die *obere* Skala die untere Konfidenzgrenze  $\hat{p}_{un,\alpha}(x)$  und das **dünne** Histogramm entspricht der zugehörigen  $B(n, \hat{p}_{un,\alpha}(x))$ -Verteilung. Die markierte Fläche stellt die Wahrscheinlichkeit  $P\{X \geq x\}$  für das jeweilige Histogramm dar, wobei der Anteil unter dem **dünnen** Histogramm genau gleich  $\alpha$  ist. Man beachte, daß die *untere* Grenze nur in der rechten Grafik *oberhalb* des wahren Wertes  $p$  liegt, weil  $x$  dort so groß ist, daß für den wahren Wert  $p = 40\%$  bereits  $P\{X \geq x | p\} < \alpha$  gilt.

### 3.1.2 b Die *exakte obere einseitige* Konfidenzgrenze

Für die Konstruktion der oberen Grenze gehen wir analog zur unteren Grenze vor und betrachten die Binomial-Wahrscheinlichkeit für die Beobachtung  $x$  *und alle kleineren Werte*

$$P\{X \leq x | p\} = \sum_{i=0}^x b(i | n, p). \quad (1)$$

Wir stellen uns wieder verschiedene „Szenarien“ für  $p$  vor, wobei wir zuerst den Fall  $x < n$  betrachten. Die Wahrscheinlichkeit  $P\{X \leq x | p\}$  fällt wenn  $p$  steigt bzw. steigt wenn  $p$  fällt (vgl. in Abb. 3 jeweils das dünne mit dem fetten Histogramm). Es gibt nun genau einen Wert für  $p$ , bei dem diese Wahrscheinlichkeit gleich  $\alpha$  ist. Dieser Wert ist die gesuchte exakte obere Konfidenzgrenze  $\hat{p}_{ob,\alpha}(x)$  zur Sicherheit  $1 - \alpha$  und er ist implizit gegeben als Lösung der Gleichung

$$P\{X \leq x | \hat{p}_{ob,\alpha}(x)\} = \alpha \quad \text{für } x < n. \quad (2)$$

Auf die konkrete Berechnung von  $\hat{p}_{ob,\alpha}(x)$  gehen wir erst im Abschnitt 3.1.2 d ein.

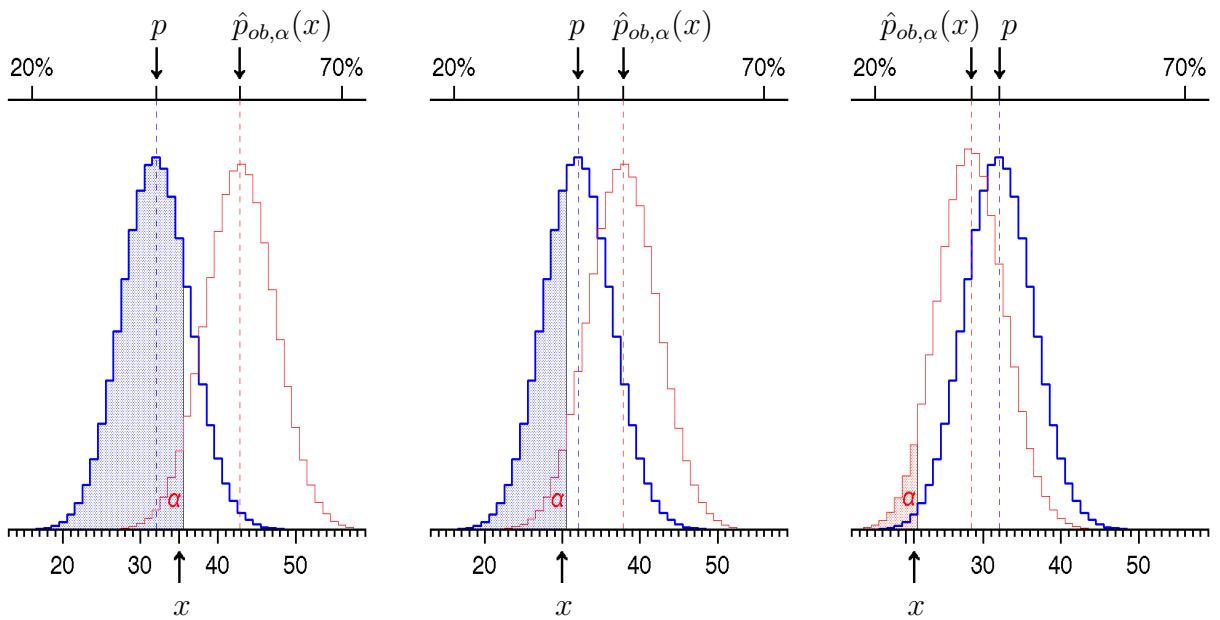


Abb. 3: Zur Konstruktion der exakten oberen Konfidenzgrenze von  $p$  für eine  $B(n,p)$ -verteilte Zufallsvariable  $X$ , wobei  $\alpha = 5\%$  ist. Das **fette** Histogramm entspricht der Verteilung für  $n = 80$  mit dem *wahren* Wert  $p = 40\%$ . Für drei verschiedene Realisierungen  $x = 35$  (links),  $x = 30$  (Mitte) und  $x = 21$  (rechts) zeigt die *obere* Skala die obere Konfidenzgrenze  $\hat{p}_{ob,\alpha}(x)$  und das **dünne** Histogramm entspricht der zugehörigen  $B(n, \hat{p}_{ob,\alpha}(x))$ -Verteilung. Die markierte Fläche stellt die Wahrscheinlichkeit  $P\{X \leq x\}$  für das jeweilige Histogramm dar, wobei der Anteil unter dem **dünnen** Histogramm genau gleich  $\alpha$  ist. Man beachte, daß die *obere* Grenze nur in der rechten Grafik *unterhalb* des wahren Wertes  $p$  liegt, weil  $x$  dort so klein ist, daß für den wahren Wert  $p = 40\%$  bereits  $P\{X \leq x | p\} < \alpha$  gilt.

Für den bisher nicht betrachteten Fall  $x = n$  ist die Schätzung  $\hat{p}(n) = 1$  und die Wahrscheinlichkeit  $P\{X \leq n | p\}$  ist unabhängig von  $p$  stets gleich 1. Es ist daher sinnvoll, die obere Konfidenzgrenze ebenfalls auf 1 zu setzen, d.h.

$$\hat{p}_{ob,\alpha}(n) = 1. \quad (3)$$

Die obere Grenze  $\hat{p}_{ob,\alpha}(x)$  ist damit (auch im Fall  $x = n$ ) der *größte* Wert für  $p$ , unter dem die Wahrscheinlichkeit  $P\{X \leq x | p\}$  noch *mindestens*  $\alpha$  beträgt:

$$\hat{p}_{ob,\alpha}(x) = \text{Max}\{p \mid P\{X \leq x \mid p\} \geq \alpha\} \quad (\text{exakte obere einseitige Konfidenzgrenze}). \quad (4)$$

In Worten: *die obere Grenze ist der größtmögliche Wert für  $p$ , unter dem das (eingetretene) Ereignis  $\{X \leq x\}$  noch eine Eintritts-Wahrscheinlichkeit von mindestens  $\alpha$  hat.* Noch größere Werte von  $p$  als  $\hat{p}_{ob,\alpha}(x)$  werden also deshalb „abgelehnt“, weil unter ihnen das eingetretene Ereignis  $\{X \leq x\}$  zu „unwahrscheinlich“ wäre, d.h. eine Eintritts-Wahrscheinlichkeit von *weniger* als  $\alpha$  hätte.

Die so konstruierte obere Grenze erfüllt unsere primäre Forderung 3.1.2 (2), wobei die geforderte Sicherheit nicht *genau erreicht*, aber dafür *mindestens* garantiert ist

$$P\{p < \hat{p}_{ob,\alpha}(X)\} \geq 1 - \alpha \quad \text{bzw.} \quad P\{\hat{p}_{ob,\alpha}(X) \leq p\} \leq \alpha. \quad (5)$$

**Für Fortgeschrittene:** Man kann die Konstruktion der *oberen Grenze* auch direkt auf die der *unteren Grenze* zurückführen. Hierzu betrachtet man das *komplementäre Ziel-Ereignis*  $\complement A$  und die zugehörige Anzahl  $Y = n - X$  der *Eintritte* von  $\complement A$ , d.h. der *Nicht-Eintritte* von  $A$ . Dann hat  $Y$  eine  $B(n, q)$ -Verteilung und die *komplementäre Wahrscheinlichkeit* der *unteren Grenze für  $q$*  ist die *obere Grenze für  $p$*

$$1 - \hat{q}_{un,\alpha}(y) = \hat{p}_{ob,\alpha}(x) \quad \text{mit} \quad y = n - x. \quad (6)$$

### 3.1.2 c Die exakten zweiseitigen Konfidenzgrenzen

Will man die Wahrscheinlichkeit  $p$  sowohl nach *unten* als auch noch *oben* zuverlässig abschätzen, so wird man hierzu das *Intervall* von der unteren bis zur oberen Grenze betrachten. Die *Irrtums-Wahrscheinlichkeit* dieses Intervalls ist allerdings die *Summe der Irrtums-Wahrscheinlichkeiten* der beiden einzelnen Grenzen:



*Intervall zwischen beiden Grenzen*

$$P\{p \leq \hat{p}_{un}(X) \text{ oder } \hat{p}_{ob}(X) \leq p\} = P\{p \leq \hat{p}_{un}(X)\} + P\{\hat{p}_{ob}(X) \leq p\}. \quad (1)$$

Damit dieses Intervall die vorgegebene Irrtums-Wahrscheinlichkeit  $\alpha$  nicht überschreitet, muß man die untere und obere Grenze jeweils zur *halben Irrtums-Wahrscheinlichkeit*  $\frac{\alpha}{2}$  bestimmen. Dies führt zum *exakten zweiseitigen Konfidenzintervall*

$$(\hat{p}_{un,\frac{\alpha}{2}}(x), \hat{p}_{ob,\frac{\alpha}{2}}(x)) = \{p \mid \hat{p}_{un,\frac{\alpha}{2}}(x) < p < \hat{p}_{ob,\frac{\alpha}{2}}(x)\} \quad (2)$$

*(exaktes zweiseitiges Konfidenzintervall).*

Das (zufällige) exakte zweiseitige Konfidenz-Intervall hat dann *mindestens* die geforderte Sicherheit  $1 - \alpha$  bzw. *höchstens* die Irrtums-Wahrscheinlichkeit  $\alpha$

$$P\{\hat{p}_{un,\frac{\alpha}{2}}(X) < p < \hat{p}_{ob,\frac{\alpha}{2}}(X)\} \geq 1 - \alpha \quad \text{bzw.} \quad (3)$$

$$P\{p \leq \hat{p}_{un,\frac{\alpha}{2}}(X) \quad \text{oder} \quad \hat{p}_{ob,\frac{\alpha}{2}}(X) \leq p\} \leq \alpha.$$

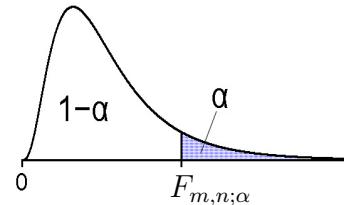
Das hier verwendete Prinzip zur Konstruktion des *zweiseitigen* Konfidenz-Intervalls gilt universell:

*Das zweiseitige Konfidenz-Intervall zur Irrtums-Wahrscheinlichkeit  $\alpha$  bzw. Sicherheit  $1 - \alpha$  wird gebildet aus der einseitigen unteren und einseitig oberen Konfidenzgrenze jeweils zur halben Irrtums-Wahrscheinlichkeit  $\frac{\alpha}{2}$ .*

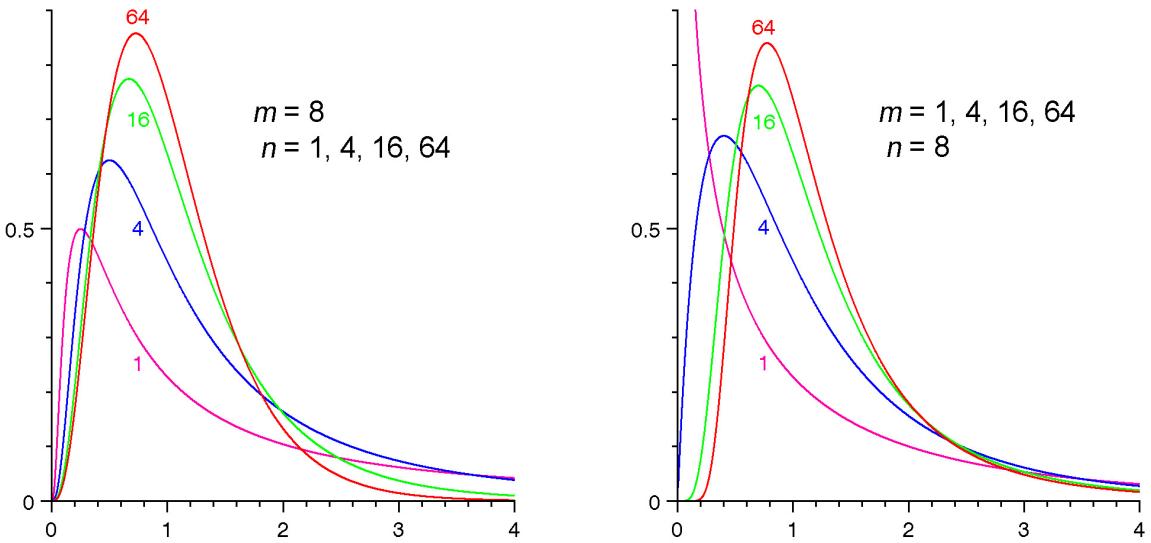
### 3.1.2 d Die Berechnung der *exakten* Konfidenzgrenzen

Wir wollen jetzt die Konfidenzgrenzen  $\hat{p}_{un,\alpha}(x)$  und  $\hat{p}_{ob,\alpha}(x)$  für vorgegebene Werte von  $\alpha, n$  und  $x$  konkret berechnen. Als Hilfsmittel hierzu dienen die sogenannten *F-Verteilungen* von Fisher. Die  $F_{m,n}$ -Verteilung ist eine *stetige* Verteilung auf der *positiven Zahlengeraden* (d.h. auf dem Intervall der positiven reellen Zahlen), die von *zwei ganzzahligen* Parametern abhängt, den *Freiheitsgraden*  $m \geq 1$  und  $n \geq 1$ . Die Dichte der  $F_{m,n}$ -Verteilung ist für verschiedene Freiheitsgrade in Abb. 4 dargestellt. Man beachte, daß die Reihenfolge der Freiheitsgrade eine Rolle spielt: die  $F_{8,1}$ -Verteilung (Abb. 4 links) stimmt *nicht* mit der  $F_{1,8}$ -Verteilung (Abb. 4 rechts) überein.

Das (*obere*)  $\alpha$ -Quantil  $F_{m,n;\alpha}$  der  $F_{m,n}$ -Verteilung ist dadurch gegeben, daß die Fläche der Dichte über dem Intervall von 0 bis zum  $F_{m,n;\alpha}$  genau  $1 - \alpha$  ist, d.h. die Fläche über dem Intervall von  $F_{m,n;\alpha}$  bis  $\infty$  ist  $\alpha$ .



Diese Quantile sind im Anhang T für gängige Argumente tabelliert und sie lassen sich in Tabellenkalkulationen (z.B. Excel) und statistischen Computer-Programmen direkt aufrufen.

Abb. 4: Dichten der  $F_{m,n}$ -Verteilung für verschiedene Freiheitsgrade  $m$  und  $n$ .

Unter Verwendung des oberen  $\alpha$ -Quantils  $F_{k,l;\alpha}$  der  $F_{k,l}$ -Verteilung erhält man folgende Darstellung der Konfidenzgrenzen

$$\hat{p}_{un,\alpha}(x) = \frac{1}{1+a} \quad (\text{exakte untere Grenze}) \quad \text{für } 0 < x \text{ mit} \quad (1)$$

$$a = \frac{k}{l} \cdot F_{k,l;\alpha}, \quad k = 2(n-x+1), \quad l = 2x.$$

$$\hat{p}_{ob,\alpha}(x) = \frac{a}{1+a} \quad (\text{exakte obere Grenze}) \quad \text{für } x < n \text{ mit} \quad (2)$$

$$a = \frac{k}{l} \cdot F_{k,l;\alpha}, \quad k = 2(x+1), \quad l = 2(n-x).$$

In den beiden Grenzfällen  $x = 0$  (d.h.  $\hat{p} = 0$ ) bzw.  $x = n$  (d.h.  $\hat{p} = 1$ ) lassen sich die exakten Grenzen sogar direkt angeben

$$\hat{p}_{un,\alpha}(0) = 0, \quad \hat{p}_{un,\alpha}(n) = \alpha^{\frac{1}{n}} \quad (\text{exakte untere Grenze}), \quad (3)$$

$$\hat{p}_{ob,\alpha}(0) = 1 - \alpha^{\frac{1}{n}}, \quad \hat{p}_{ob,\alpha}(n) = 1 \quad (\text{exakte obere Grenze}). \quad (4)$$

**Erfolg einer Therapie:** In der Tagespresse wird berichtet, daß bei einer neuen Therapie in nur  $\hat{p} = 12\%$  der Anwendungen ein Versagen beobachtet wurde, wobei die Anzahl  $n = 25$  der Anwendungen und die beobachte Zahl  $x = n\hat{p} = 3$  des Versagens nicht genannt wird. Die exakte obere 95%-Konfidenzgrenze (d.h.  $\alpha = 5\%$ ) der Versagenswahrscheinlichkeit  $p$  ergibt sich aus (2) (oder der Konfidenz-Box 1 in 3.1.4) mit  $k = 8$ ,  $l = 44$ ,  $F_{kl;5\%} = 2,157$  und  $a = 0,3922$  zu (vgl. auch Abb. 7)

$$\hat{p}_{ob}(3) = 28,2\% \quad \text{für} \quad n = 25, \quad \hat{p} = 12\%.$$

Man beachte, daß diese oberen Grenzen – bedingt durch den kleinen Stichprobenumfang  $n = 25$  – mehr als doppelt so groß sind, wie die *beobachtete* Rate  $\hat{p} = 12\%$ .

Wenn bei dem *vierfachen* Umfang  $n = 100$  auch wieder in  $\hat{p} = 12\%$  Fällen ein Versagen beobachtet wird, d.h.  $x = n\hat{p} = 12$ , so ergibt sich  $k = 26$ ,  $l = 176$ ,  $F_{kl;5\%} = 1,669$  und  $a = 0,2303$  die erheblich geringere obere Grenze (vgl. auch Abb. 7)

$$\hat{p}_{ob}(12) = 18,7\% \quad \text{für} \quad n = 100, \quad \hat{p} = 12\%. \quad \square$$

**Für Fortgeschrittene:** Die Dichte  $\varphi_{m,n}$  der  $F_{m,n}$ -Verteilung ist gegeben durch

$$\varphi_{m,n}(x) = d_{m,n} \cdot x^{\frac{m}{2}-1} \cdot (mx + n)^{-\frac{m+n}{2}}$$

wobei die Konstante  $d_{m,n}$  durch die *Eulersche Gammafunktion*  $\Gamma$

definiert ist

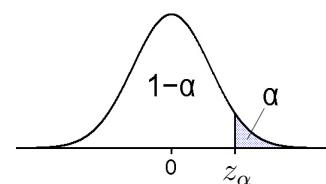
$$d_{m,n} = m^{\frac{m}{2}} n^{\frac{n}{2}} \cdot \Gamma\left(\frac{1}{2}(m+n)\right) / (\Gamma\left(\frac{1}{2}m\right) \cdot \Gamma\left(\frac{1}{2}n\right)). \quad \square$$

### 3.1.3 Asymptotische (approximative) Konfidenzgrenzen für eine Wahrscheinlichkeit

Nachdem wir bisher *exakte* Konfidenzgrenzen für die Wahrscheinlichkeit  $p$  konstruiert haben, wollen wir jetzt die sogenannten *asymptotischen (approximativen) Konfidenzgrenzen* behandeln. Ausgangspunkt ist wieder die beobachtete Realisierung  $x$  einer  $B(n,p)$ -verteilten Zufallsvariablen  $X$ . Der Unterschied bei der Konstruktion dieser *asymptotischen* gegenüber den *exakten* Konfidenzgrenzen besteht darin, daß man die Binomialverteilung  $B(n,p)$  durch die approximierende Normalverteilung  $N(\mu, \sigma^2)$  mit  $\mu = np$  und  $\sigma^2 = np(1-p)$  ersetzt. Graphisch bedeutet dies, daß man in den Abb. 2-3 die *Histogramme* der Binomialverteilungen durch die *glatten Dichten* der zugehörigen Normalverteilungen ersetzt und dann Abb. 5-6 erhält. Die resultierenden approximativen Grenzen lassen sich dann leicht bestimmen. Hierzu benötigt man nur das sogenannte *obere  $\alpha$ -Quantil*  $z_\alpha$  der *Standard-Normalverteilung*  $N(0, 1)$ , welches gegeben ist durch

$$\begin{aligned} \Phi(z_\alpha) &= 1 - \alpha \quad \text{bzw.} \\ z_\alpha &= \Phi^{-1}(1 - \alpha) \end{aligned} \quad (1)$$

mit  $\Phi^{-1}$  als *Umkehrfunktion (Inverse)* von  $\Phi$ .



$N(0,1)$ -Dichte mit  $\alpha$ -Quantil

Die Quantile  $z_\alpha$  sind im Anhang **T** für gängige Werte von  $\alpha$  tabelliert, und sie lassen sich in Tabellenkalkulationen und statistischen Computer-Programmen direkt aufrufen.

### 3.1.3 a Normale Konfidenzgrenzen

Die normale einseitige untere Grenze  $\tilde{p}_{un,\alpha}(x)$  und die obere Grenze  $\tilde{p}_{ob,\alpha}(x)$  sind die beiden Nullstellen der quadratischen Funktion

$$\begin{aligned} f(p) &= n(p - \hat{p})^2 - p(1 - p)z_a^2 \\ &= ap^2 - bp + c \end{aligned} \quad \text{mit} \quad (1)$$

$$a = n + z_\alpha^2, \quad b = 2n\hat{p} + z_\alpha^2 = 2x + z_\alpha^2, \quad c = n\hat{p}^2 = \frac{1}{n}x^2. \quad (2)$$

Mit den Hilfsgrößen

$$p_m = \frac{b}{2a} = \frac{\hat{p} + \frac{1}{2n}z_\alpha^2}{1 + \frac{1}{n}z_\alpha^2} = \frac{x + \frac{1}{2}z_\alpha^2}{n + z_\alpha^2} \quad (3)$$

$$D = p_m^2 - \frac{c}{a} = p_m^2 - \frac{\hat{p}^2}{1 + \frac{1}{n}z_\alpha^2} = p_m^2 - \frac{x^2}{n(n + z_\alpha^2)} \quad (4)$$

erhält man die *normalen Konfidenzgrenzen* als die beiden Nullstellen der Funktion  $f$ :

$$\tilde{p}_{un,\alpha}(x) = p_m - \sqrt{D} \quad (\text{normale untere Konfidenzgrenze}), \quad (5)$$

$$\tilde{p}_{ob,\alpha}(x) = p_m + \sqrt{D} \quad (\text{normale obere Konfidenzgrenze}). \quad (6)$$

Diese normalen Konfidenzgrenzen halten die angestrebte Sicherheit  $1 - \alpha$  nur *approximativ* ein

$$P\{\tilde{p}_{un,\alpha}(X) < p\} \approx 1 - \alpha, \quad P\{p < \tilde{p}_{ob,\alpha}(X)\} \approx 1 - \alpha, \quad (7)$$

wobei die Approximationen für *wachsenden* Umfang  $n$  immer genauer werden. Man bezeichnet die *normalen Konfidenzgrenzen* deshalb auch als *asymptotische Konfidenzgrenzen*, weil sie die angestrebte Sicherheit  $1 - \alpha$  erst *asymptotisch* (d.h. für  $n \rightarrow \infty$ ) erreichen.

Das *normale zweiseitige* Konfidenz-Intervall zur approximativen Sicherheit  $1 - \alpha$  ergibt sich jetzt wieder aus den beiden einseitigen normalen Grenzen mit  $\frac{\alpha}{2}$  statt  $\alpha$ :

$$\begin{aligned} (\tilde{p}_{un,\frac{\alpha}{2}}(x), \tilde{p}_{ob,\frac{\alpha}{2}}(x)) &= \{p \mid \tilde{p}_{un,\frac{\alpha}{2}}(x) < p < \tilde{p}_{ob,\frac{\alpha}{2}}(x)\} \\ &\quad (\text{normales zweiseitiges Konfidenzintervall}). \end{aligned} \quad (8)$$

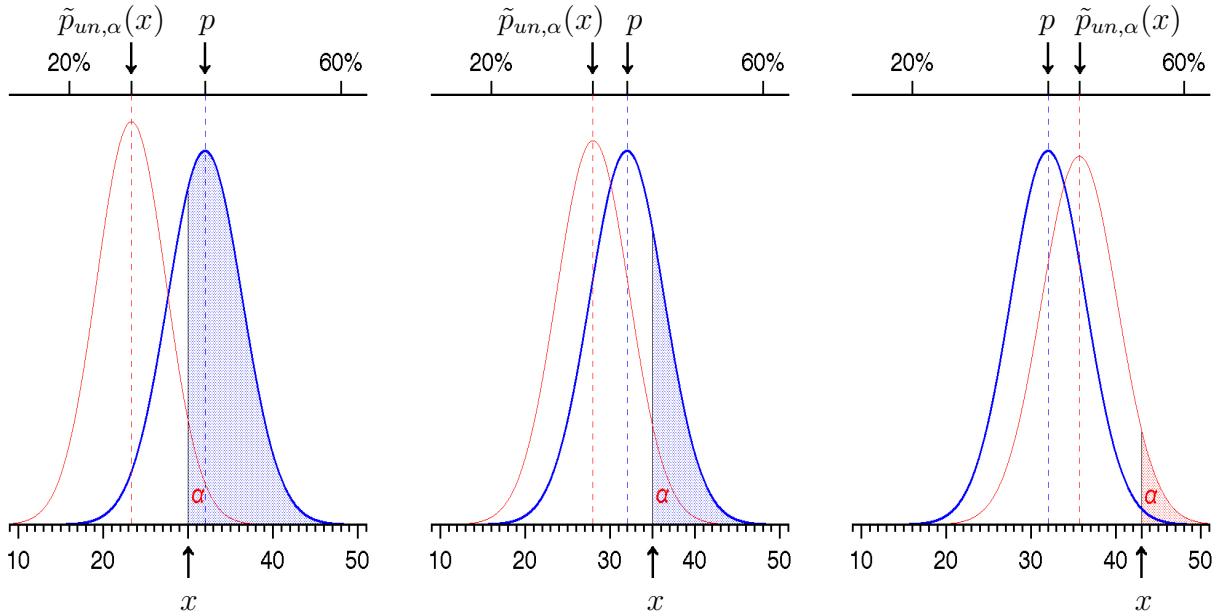


Abb. 5: Konstruktion der *normalen unteren* Konfidenzgrenze von  $p$  für  $\alpha = 5\%$  analog Abb. 2. Die **fette** Dichte entspricht der approximierenden Normalverteilung der Binomialverteilung  $B(n, p)$  für  $n = 80$  mit dem *wahren* Wert  $p = 40\%$ . Für drei Realisierungen  $x$  zeigt die *obere* Skala die untere Konfidenzgrenze  $\tilde{p}_{un,\alpha}(x)$  und die **dünne** Dichte ist die Normalapproximation der zugehörigen  $B(n, \tilde{p}_{un,\alpha}(x))$ -Verteilung. Die markierte Fläche approximiert die Wahrscheinlichkeit  $P\{X \geq x\}$ , wobei der Anteil unter der **dünnen** Dichte genau gleich  $\alpha$  ist.

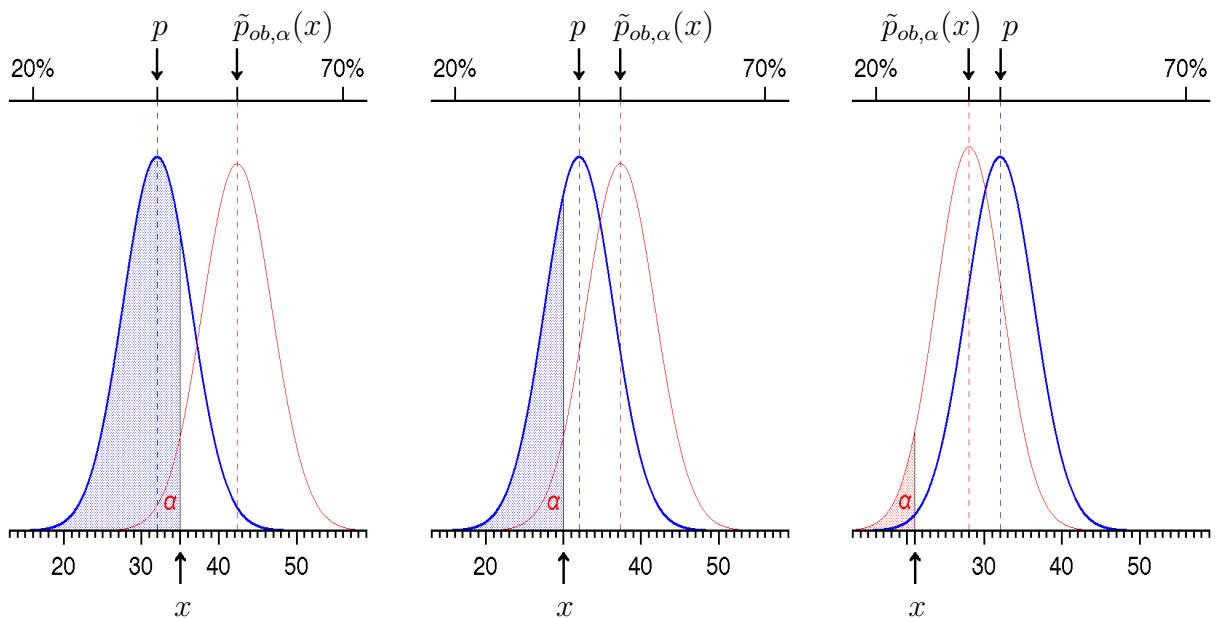
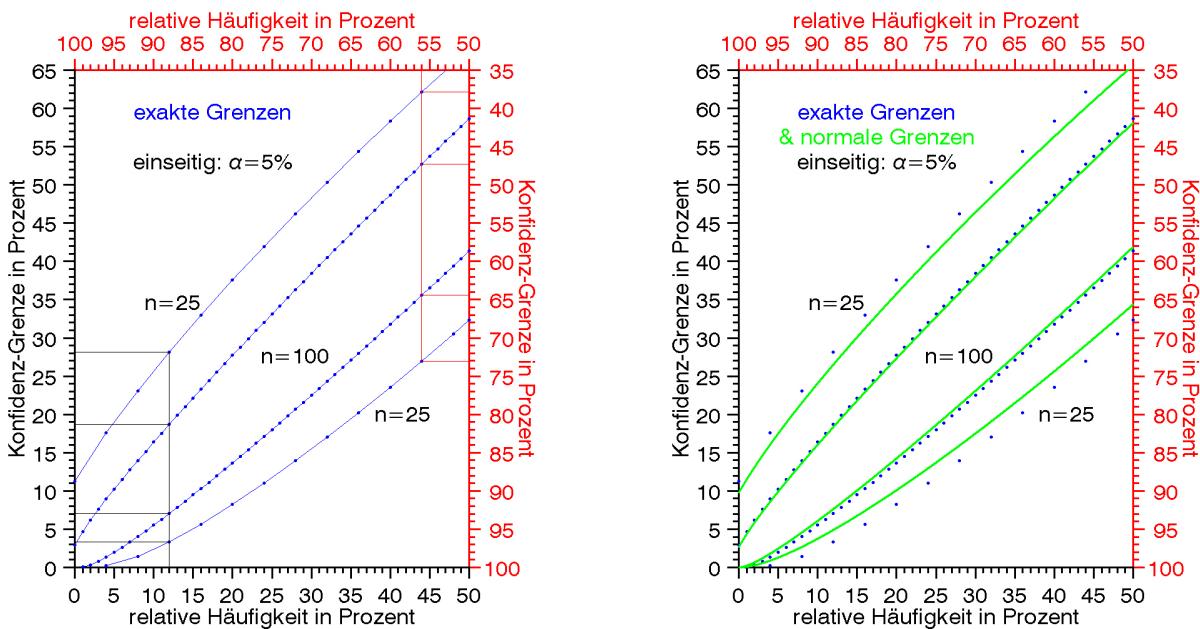


Abb. 6: Konstruktion der *normalen oberen* Konfidenzgrenze von  $p$  für  $\alpha = 5\%$  analog Abb. 3. Die **fette** Dichte entspricht der approximierenden Normalverteilung der Binomialverteilung  $B(n, p)$  für  $n = 80$  mit dem *wahren* Wert  $p = 40\%$ . Für drei Realisierungen zeigt die *obere* Skala die obere Konfidenzgrenze  $\tilde{p}_{ob,\alpha}(x)$  und die **dünne** Dichte ist die Normalapproximation der zugehörigen  $B(n, \tilde{p}_{ob,\alpha}(x))$ -Verteilung. Die markierte Fläche approximiert die Wahrscheinlichkeit  $P\{X \leq x\}$ , wobei der Anteil unter der **dünnen** Dichte genau gleich  $\alpha$  ist.

**Normale (asymptotische) vs. exakte Konfidenzgrenzen:** Ein Vergleich der normalen mit den exakten 95%-Grenzen in Abb. 7 zeigt für  $n = 25$  noch deutliche Unterschiede, aber bereits für  $n = 100$  eine zufriedenstellende Übereinstimmung. Typischerweise sind die exakten Grenzen jeweils *weiter* von der Schätzung (relativen Häufigkeit)  $\hat{p}$  entfernt als die normalen Grenzen. Dies liegt daran, daß die Sicherheit der exakten Grenzen nicht *genau*  $1 - \alpha$ , sondern *mindestens*  $1 - \alpha$  ist. Generell ist man auf der sicheren Seite, wenn man die *exakten* Grenzen verwendet. Die normalen Grenzen sollte man erst bei größerem Umfang  $n$  verwenden, z.B. wenn die erforderlichen Quantile der  $F$ -Verteilung nicht mehr zur Verfügung stehen. Bei den Tabellen im Anhang T4 ist dies ab  $x \geq 15$  oder  $n - x \geq 15$  der Fall.



**Abb. 7:** Exakte und normale untere und obere Grenzen für  $p$  zur (einseitigen) Sicherheit von 95% als Funktion der beobachteten relativen Häufigkeit  $\hat{p}$  für  $n = 25$  und  $n = 100$ . Für  $\hat{p} \leq 50\%$  gilt die *untere* und *linke* Skala, und für  $\hat{p} \geq 50\%$  gilt die *obere* und *rechte* Skala. Das Intervall zwischen unterer und oberer Grenze hat eine Sicherheit von 90%.

**links:** exakte Grenzen mit Ablesebeispielen für  $\hat{p} = 12\%$  und  $\hat{p} = 56\%$ .

**rechts:** Vergleich der *exakten* (Punkte) mit den *normalen* (Linie) Grenzen. Die *exakten* Grenzen sind typischerweise *weiter* von der relativen Häufigkeit entfernt als die *normalen*. Der Unterschied wird bei wachsenden  $n$  geringer.

**Erfolg einer Therapie (Fortsetzung):** Für  $n = 100$  Anwendungen und einer beobachteten Zahl  $x = 12$  haben wir die *exakte* obere 95%-Konfidenzgrenze der Versagenswahrscheinlichkeit  $p$  bereits in 3.1.2 d bestimmt. Zum Vergleich berechnen wir jetzt auch die *normale* obere 95%-Konfidenzgrenze. Für  $\alpha = 5\%$  ist  $z_\alpha = 1,645$  und mit den Hilfsgrößen

$$a = 102,71 \quad b = 26,71 \quad c = 1,44 \quad p_m = 13,0\% \quad \sqrt{D} = 5,4\%$$

ergibt sich die *normale* obere Grenze zu  $\tilde{p}_{ob}(12) = 18,4\%$ , die hier nur unwesentlich von der *exakten* Grenze  $\hat{p}_{ob}(12) = 18,7\%$  abweicht.

Für  $n = 25$  und  $x = 3$  weicht die *normale* obere 95%-Grenze  $\tilde{p}_{ob}(3) = 26,5\%$  deutlich stärker von der *exakten* Grenze  $\hat{p}_{ob}(3) = 28,2\%$  ab. Hier sollte man sich aber wegen des kleinen Umfangs  $n$  in jedem Fall nur auf die *exakte* Grenze verlassen.  $\square$

### 3.1.3 b Grobe Konfidenzgrenzen für große Umfänge

Die normalen Grenzen liegen *symmetrisch* um den Punkt  $p_m$  jeweils im Abstand  $\sqrt{D}$ . Man beachte, daß  $p_m$  nicht *genau* mit der Schätzung  $\hat{p}$  übereinstimmt, aber typischerweise bei großem  $n$  nur gering von  $\hat{p}$  abweicht.

Für *sehr großen Umfang*  $n$  (z.B. ab  $n = 1000$ ) verwendet man vereinfachend die so genannten *groben Konfidenzgrenzen*

$$\tilde{\tilde{p}}_{un,\alpha}(x) = \hat{p} - \hat{d}_\alpha \quad (\text{grobe untere Konfidenzgrenze}), \quad (1)$$

$$\tilde{\tilde{p}}_{ob,\alpha}(x) = \hat{p} + \hat{d}_\alpha \quad (\text{grobe obere Konfidenzgrenze}) \quad \text{mit} \quad (2)$$

$$\hat{d}_\alpha = z_\alpha \cdot \sqrt{\frac{1}{n}\hat{p}(1-\hat{p})} \quad (\text{Bandbreite}) \quad (3)$$

Die groben Konfidenzgrenzen liegen *symmetrisch* zur relativen Häufigkeit  $\hat{p}$  im Abstand  $\pm \hat{d}_\alpha$ . Sie halten die angestrebte Sicherheit  $1 - \alpha$  auch nur *approximativ* ein

$$P\{\tilde{\tilde{p}}_{un,\alpha}(X) < p\} \approx 1 - \alpha, \quad P\{p < \tilde{\tilde{p}}_{ob,\alpha}(X)\} \approx 1 - \alpha, \quad (4)$$

wobei die Approximationen für *wachsenden* Umfang  $n$  immer genauer werden.

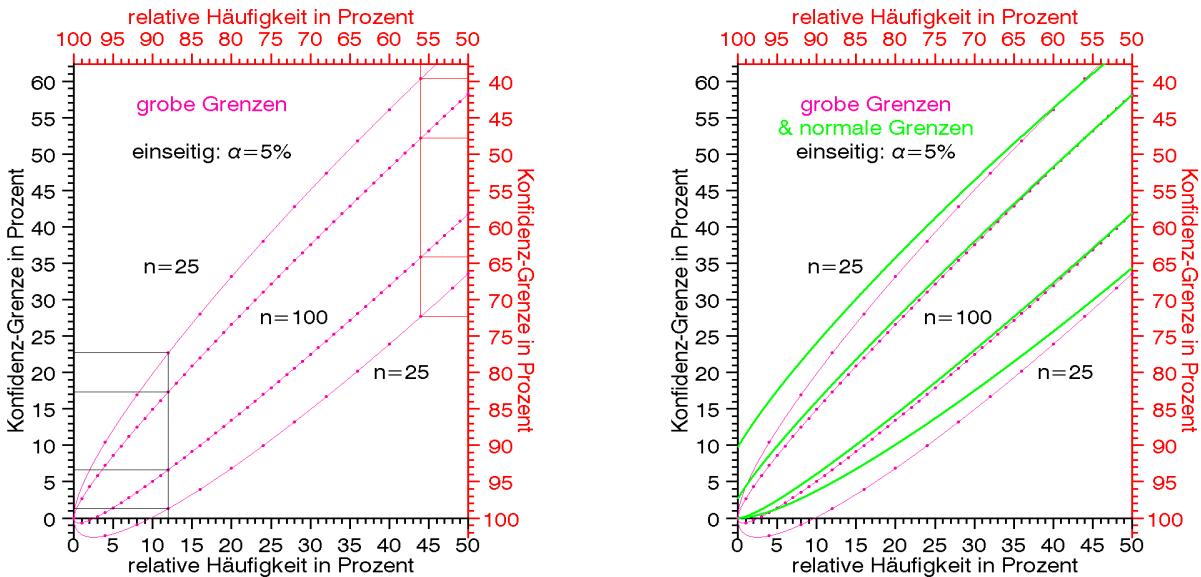
Das zugehörige *grobe zweiseitige* Konfidenz-Intervall zur approximativen Sicherheit  $1 - \alpha$  ergibt sich wieder aus den beiden einseitigen groben Grenzen mit  $\frac{\alpha}{2}$  statt  $\alpha$

$$(\tilde{\tilde{p}}_{un,\frac{\alpha}{2}}(x), \tilde{\tilde{p}}_{ob,\frac{\alpha}{2}}(x)), \quad = \quad \{P \mid \tilde{\tilde{p}}_{un,\frac{\alpha}{2}}(x) < p < \tilde{\tilde{p}}_{ob,\frac{\alpha}{2}}(x)\} \quad (5)$$

*grobes zweiseitiges Konfidenzintervall .*

**Grobe vs. normale (asymptotische) Konfidenzgrenzen:** Wir haben die *groben* Grenzen hier nur aus Gründen der Vollständigkeit erwähnt. Sie haben gegenüber den *normalen* asymptotischen Grenzen mehrere Nachteile. Typischerweise weicht die Sicherheit der groben Grenzen stärker von  $1 - \alpha$  ab als die der normalen Grenzen. Außerdem können die groben Grenzen auch außerhalb des Intervalls  $[0, 1]$  liegen (vgl. Abb. 8), und ergeben in den Extremfällen  $\hat{p} = 0$  oder  $\hat{p} = 1$  wegen  $\hat{d}_\alpha = 0$  keine sinnvollen Werte. Es ist immer sicherer, die *normalen* Grenzen zu verwenden, deren Berechnung auch nur unwesentlich aufwendiger ist. Lediglich bei *extrem großem* Umfang  $n$  oder wenn die

Schätzung  $\hat{p}$  nur *wenig* von 50% abweicht, können die groben Grenzen als hinreichende Approximation der normalen Grenzen verwendet werden. Dies ist typischerweise der Fall, wenn sowohl  $x \geq 50$  als auch  $n - x \geq 50$  gilt (*Faustregel*).



**Abb. 8:** Grobe und normale untere und obere Grenzen für  $p$  zur (einseitigen) Sicherheit von 95% als Funktion der beobachteten relativen Häufigkeit  $\hat{p}$  für  $n = 25$  und  $n = 100$ . Für  $\hat{p} \leq 50\%$  gilt die *untere* und *linke* Skala, und für  $\hat{p} \geq 50\%$  gilt die *obere* und *rechte* Skala.

**links:** Die *grobe* untere (bzw. obere) Grenze ist für kleines (bzw. großes)  $\hat{p}$  sogar *negativ* (bzw. größer als 100%).

**rechts:** Die Abweichung der *groben* (dünn) von den *normalen* (fett) Grenzen wird kleiner, je dichter  $\hat{p}$  bei 50% liegt, und verringert sich bei wachsendem  $n$ .

**Wahlumfrage:** Wir wenden uns wieder Wahlumfragen zu, bei denen der Stimmanteil  $p$  einer Partei aus  $n$  Befragungen geschätzt werden soll (vgl. 2.3.3). Konkret betrachten wir hierzu die Umfrage aus 2.1 Abb. 2 vom Umfang  $n = 1300$ , bei der sich folgende Schätzungen  $\hat{p}$  (relativer Stimmanteil in der Stichprobe) ergaben.

SPD	CDU/CSU	Grüne	FDP	PDS	Sonstige
36%	40%	6%	9%	6%	4%

Wir wollen das *zweiseitige* 95%-Konfidentintervall exemplarisch für eine *kleine* und eine *große* Partei bestimmen, wobei wir wegen des hohen Umfangs  $n$  hier die *groben* Grenzen berechnen. Aus  $\alpha = 5\%$  und  $z_{\alpha/2} = 1,96$  ergibt sich

Partei	$\hat{p}$	$\hat{d}_{\alpha/2}$	$\tilde{p}_{un,\alpha/2}$	$\tilde{p}_{ob,\alpha/2}$
Grüne	6%	1,29%	4,7%	7,3%
CDU/CSU	40%	2,66%	37,3%	42,7%

Die Abweichung der *groben* von den *normalen* bzw. *exakten* Konfidenzgrenzen ist hier übrigens geringer als 0,1%. – Man beachte, daß die Bandbreite  $\hat{d}_{\alpha/2}$  (in **2.1 Abb. 2** als *Fehlertoleranz* bezeichnet) sich mit der Schätzung  $\hat{p}$  ändert. Sie ist für  $\hat{p} = 50\%$  am größten (2,72%) und fällt, wenn  $\hat{p}$  nach unten (oder oben) von 50% abweicht: für  $\hat{p} = 5\%$  ergibt sich z.B.  $\hat{d}_{\alpha/2} = 1,18\%$ .

Für eine konkrete Partei, z.B. die *Grünen*, ist primär eine *untere* Grenze von Interesse, weil diese ihren Stimmanteil mit der entsprechenden Sicherheit nach unten abschätzt. Die untere Grenze des obigen *zweiseitigen* Intervalls hat die Sicherheit  $1 - \frac{\alpha}{2} = 97,5\%$ . Da diese *untere Grenze unter* der 5%-Hürde liegt, ist der Einzug ins Parlament nicht zu 97,5% abgesichert. Wenn man *nur* an der unteren Grenze interessiert ist, wird man die einseitige Grenze zur Sicherheit von 95% bestimmen. Diese lautet für die Grünen  $\tilde{p}_{un,\alpha} = 4,92\%$  und liegt (trotz der auf 95% verringerten Sicherheit) immer noch unter der 5%-Hürde. Kurioserweise liegt die zugehörige *normale* Grenze  $\tilde{p}_{un,\alpha} = 5,01\%$  knapp *oberhalb* von 5% und die *exakte* Grenze  $\hat{p}_{un,\alpha} = 4,95\%$  wieder *unterhalb* von 5%. Hieran erkennt man, daß selbst bei großem Umfang  $n$  die sich dann nur gering unterscheidenden Grenzen (exakt, normal oder grob) im Einzelfall zu unterschiedlichen Schlußfolgerungen führen können. □

## Tipps für *Microsoft Excel*

**Funktionen:** Das  $\alpha$ -Quantil der Standard-Normalverteilung bzw. der  $F$ -Verteilung (vgl. Anhang T) läßt sich mit folgenden Statistik-Funktionen in Excel berechnen:

$$\begin{aligned} z_\alpha &= \text{STANDNORMINV}[1 - \alpha] = -\text{STANDNORMINV}[\alpha] \\ F_{m,n;\alpha} &= \text{FINV}[\alpha; m; n] \end{aligned}$$

**3.1.4 Konfidenz-Box 1:** Konfidenzgrenzen für die Wahrscheinlichkeit  $p$  einer  $B(n, p)$ -verteilten Zufallsvariablen  $X$ .

Irrtums-Wahrscheinlichkeit:	$\alpha$	Sicherheit:	$1 - \alpha$
Stichprobenumfang:	$n$	(Anzahl der Wiederholungen)	
Beobachtete Anzahl:	$x$	(beobachtete absolute Häufigkeit)	
Schätzung von $p$ :	$\hat{p} = \frac{x}{n}$	(beobachtete relative Häufigkeit)	

**Exakte einseitige Konfidenzgrenzen:** exakte Sicherheit von mindestens  $1 - \alpha$

Anwendung: immer, zwingend bei kleinem  $n$

Erforderlich: Quantile der  $F$ -Verteilung (Anhang **T** oder per Computer)

einseitige untere Grenze	einseitige obere Grenze
$\hat{p}_{un,\alpha} = 0$	$\hat{p}_{ob,\alpha} = 1 - \alpha^{\frac{1}{n}}$
$\hat{p}_{un,\alpha} = \frac{1}{1+a}$	$\hat{p}_{ob,\alpha} = \frac{a}{1+a}$
$a = \frac{k}{l} \cdot F_{k,l;\alpha}$	$a = \frac{k}{l} \cdot F_{k,l;\alpha}$
$k = 2(n - x + 1), \quad l = 2x$	$k = 2(x + 1), \quad l = 2(n - x)$
$\hat{p}_{un,\alpha} = \alpha^{\frac{1}{n}}$	$\hat{p}_{ob,\alpha} = 1$
für $x = 0$	für $x = 0$
für $x > 0$	für $x < n$

**Normale einseitige Konfidenzgrenzen:** approximative Sicherheit  $1 - \alpha$

Anwendung: bei nicht zu kleinem  $n$ , Faustregel: ab  $x \geq 10$  und  $n - x \geq 10$

Erforderlich: Quantil  $z_\alpha$  der  $N(0, 1)$ -Verteilung (Anhang **T** oder per Computer)

einseitige untere Grenze	einseitige obere Grenze
$\tilde{p}_{un,\alpha} = p_m - \sqrt{D}$	$\tilde{p}_{ob,\alpha} = p_m + \sqrt{D}$
$p_m = \frac{b}{2a}$	$D = p_m^2 - \frac{c}{a}$
$a = n + z_\alpha^2$	$b = 2n\hat{p} + z_\alpha^2$
	$c = np\hat{p}^2$

**Große einseitige Konfidenzgrenzen:** approximative Sicherheit  $1 - \alpha$

Anwendung: bei großem  $n$ , Faustregel: ab  $x \geq 50$  und  $n - x \geq 50$

Erforderlich: Quantil  $z_\alpha$  der  $N(0, 1)$ -Verteilung (Anhang **T** oder per Computer)

einseitige untere Grenze	einseitige obere Grenze
$\tilde{\tilde{p}}_{un,\alpha} = \hat{p} - \hat{d}_\alpha$	$\tilde{\tilde{p}}_{ob,\alpha} = \hat{p} + \hat{d}_\alpha$

$$\hat{d}_\alpha = z_\alpha \cdot \sqrt{\frac{1}{n}\hat{p}(1 - \hat{p})} \quad (\text{Bandbreite})$$

**Zweiseitiges Konfidenzintervall:** beide einseitigen Grenzen mit  $\frac{\alpha}{2}$  statt  $\alpha$ .

### 3.2 Testen einer Wahrscheinlichkeit

Neben dem *Schätzen* unbekannter Parameter (mit *Konfidenzgrenzen*) über Verteilungen gehört das *Testen von Hypothesen* zu den wichtigsten Methoden der deduktiven Statistik. Bei einem *statistischen Test* geht es darum, daß man aufgrund von *beobachteten Daten* eine *Entscheidung* herbeiführen will zwischen zwei verschiedenen *Hypothesen* (über unbekannte Parameter). Zur Einführung in die Idee des *statistischen Tests* wollen wir zunächst Tests über die unbekannte Wahrscheinlichkeit  $p$  eines interessierenden Ziel-Ereignisses  $A$  betrachten. Wir betrachten zunächst einige Beispiele.

**Wahlumfrage:** Angenommen eine Partei steht vor der Entscheidung, ob sie vorzeitigen Neuwahlen zustimmen soll oder nicht. Bevor sie zustimmt, will sie wissen, ob sie bei den Neuwahlen ihr Wahl-Ziel erreichen würde, d.h ob ihr Stimmanteil  $p$  (in der betroffenen Region) größer ist als ein *Minimalwert*  $p_0$ , z.B.  $p_0 = 5\%$  (Überwindung der 5%-Hürde) bei einer *kleinen* Partei oder  $p_0 = 50\%$  (absolute Mehrheit) bei einer *großen* Partei. Die zugehörigen Hypothesen lassen sich wie folgt formalisieren:

$$\begin{aligned} H_{\leq}: p \leq p_0 & \quad (\text{Wahlziel wird nicht übertroffen}) & \quad \text{vs.} \\ H_{>}: p > p_0 & \quad (\text{Wahlziel wird übertroffen}). \end{aligned} \tag{1}$$

Die Entscheidung zwischen beiden Hypothesen will die Partei vom Ergebnis einer Wahlumfrage mit  $n$  (voneinander unabhängigen) Befragungen abhängig machen. Die Anzahl  $X$  der Für-Stimmen in der Umfrage ist dann  $B(n, p)$ -verteilt. Als Konsequenz will sie vorzeitigen Neuwahlen nur dann zustimmen, wenn sie sich aufgrund der Umfrage für die Hypothese  $H_{>}: p > p_0$  entschieden hat.  $\square$

**Klinische Studie:** Ein Klinikum steht vor der Entscheidung, eine Standardtherapie (z.B. eine Chemotherapie) durch eine neue und teurere Therapie zu ersetzen. Aus Kostengründen will man sich für die neue Therapie (als neuen Standard) nur dann entscheiden, wenn die Wahrscheinlichkeit  $p$  für das *Therapie-Versagen* geringer ist als die Versagens-Wahrscheinlichkeit  $p_0$  der bisherigen Standardtherapie. Die formalisierten Hypothesen lauten

$$\begin{aligned} H_{\geq}: p \geq p_0 & \quad (\text{neue Therapie nicht besser als Standard}) & \quad \text{vs.} \\ H_{<}: p < p_0 & \quad (\text{neue Therapie besser als Standard}). \end{aligned} \tag{2}$$

Die Entscheidung zwischen beiden Hypothesen – und die damit verbundene Entscheidung über das Ersetzen der Standardtherapie durch die neue Therapie – will das Klinikum vom Ergebnis einer klinischen Studie abhängig machen, bei der die neue Therapie in insgesamt  $n$  (voneinander unabhängigen) Fällen angewandt wird. Die Anzahl  $X$  der Therapie-Versager ist dann  $B(n, p)$ -verteilt.

Betrachtet man statt der Wahrscheinlichkeit  $p$  für das *Versagen* die komplementäre Wahrscheinlichkeit  $q = 1 - p$  für den *Erfolg* der Therapie, so lassen sich die Hypothesen äquivalent formulieren durch

$$\begin{aligned} H_{\leq}: q \leq q_0 & \quad (\text{neue Therapie nicht besser als Standard}) & \quad \text{vs.} & \quad (3) \\ H_{>}: q > q_0 & \quad (\text{neue Therapie besser als Standard}). \end{aligned}$$

Die Anzahl  $Y = n - X$  der Therapie-Erfolge ist  $B(n, q)$ -verteilt.  $\square$

**Genetik – Dominanter Erbgang:** Zur Klärung des Erbgangs eines interessierenden Merkmals (z.B. die Blütenfarbe des Löwenmäulchens) werden in der Genetik Kreuzungsversuche durchgeführt. Bei einem *dominanteren* Erbgang wird die interessierende Merkmalausprägung (z.B. rote Blüte) durch *ein* Gen mit *zwei* Allelen *A* (*dominant*, z.B. *rot*) und *a* (*rezessiv*, z.B. *weiss*) gesteuert, wobei das Merkmal (z.B. rote Blüte) genau dann auftritt, wenn *mindestens ein dominantes* Allel *A* vorhanden ist. Bei einer dihybriden Kreuzung  $AA \times aa$  hat die erste Tochtergeneration  $F_1$  immer den Genotyp  $Aa$  und das Merkmal tritt stets auf. Eine weitere Kreuzung  $Aa \times Aa$  ergibt in der Tochtergeneration  $F_2$  die drei Genotypen  $AA$ ,  $Aa$  und  $aa$  mit den zugehörigen Wahrscheinlichkeiten (bei *zufälliger* Weitergabe der jeweiligen Allele)

$$P\{AA\} = \frac{1}{4}, \quad P\{Aa\} = \frac{1}{2}, \quad P\{aa\} = \frac{1}{4}.$$

Die Wahrscheinlichkeit für das *Auftreten* des Merkmals (z.B. rote Blüte) in der  $F_2$ -Generation ist beim *dominanteren* Erbgang daher  $p_0 = \frac{3}{4}$ .

Will man überprüfen, ob ein bisher nicht untersuchter Erbgang dominant ist oder nicht, so lässt sich dies aus Hypothesen über die Wahrscheinlichkeit  $p$  für das Auftreten des Merkmals in der  $F_2$ -Generation ableiten:

$$\begin{aligned} H_=: p = p_0 & \quad (\text{Übereinstimmung mit dominantem Erbgang}) & \quad \text{vs.} & \quad (4) \\ H_{\neq}: p \neq p_0 & \quad (\text{keine Übereinstimmung mit dominantem Erbgang}). \end{aligned}$$

Bei einem Umfang  $n$  der  $F_2$ -Generation hat die Anzahl  $X$  der Nachkommen mit diesem Merkmal eine  $B(n, p)$ -Verteilung. Aufgrund der beobachteten Anzahl  $x$  will man sich dann für eine der beiden Hypothesen entscheiden.  $\square$

### 3.2.1 Einführung in statistische Tests am Beispiel

Wir wollen die grundlegenden Ideen eines statistischen Tests zunächst an einem Anwendungsbeispiel erläutern (vgl. *Kinder-Osius-Timm* 1982, Beispiel 7.1).

#### Genetik: Dominanz vs. komplementäre Polygenie

Hierzu betrachten wir für den Erbgang eines interessierenden phänotypischen Merkmals zweier konkurrierende Modelle: den *dominanteren Erbgang* und *komplementäre Polygenie*. Beim *dominanteren* Erbgang wird die Merkmalausprägung durch ein Gen mit *zwei* Allelen *A* (*dominant*) und *a* (*rezessiv*) gesteuert, wobei das Merkmal auftritt, wenn *mindestens*

ein dominantes Allel  $A$  vorhanden ist. Ist dagegen noch ein weiteres Gen mit den Allelen  $B$  und  $b$  am Erbgang beteiligt, so tritt das Merkmal z.B. bei *komplementärer Polygenie* nur dann auf, wenn *beide* dominanten Allele  $A$  und  $B$  vorhanden sind.

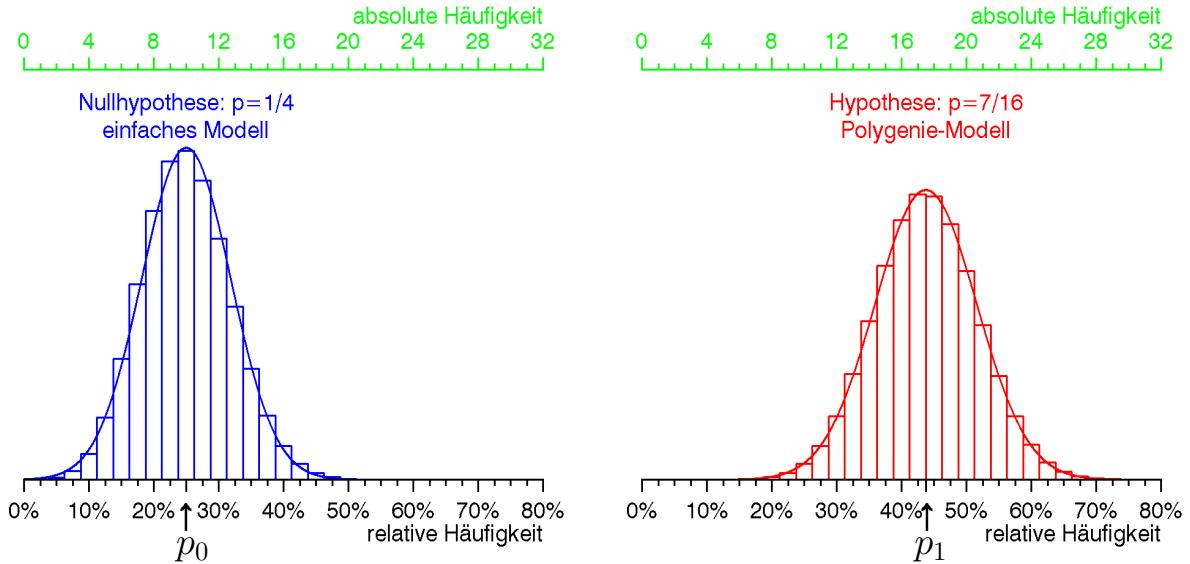
Bei einer dihybriden Kreuzung  $AABB \times aabb$  hat die erste Tochtergeneration ( $F_1$ ) stets den Genotyp  $AaBb$ , und das Merkmal tritt beim einfachen und beim Polygenie-Modell stets auf. In der zweiten Tochtergeneration ( $F_2$ ) treten bei der Kreuzung  $AaBb \times AaBb$  dann 16 verschiedenen Allel-Kombinationen mit jeweils gleicher Wahrscheinlichkeit von  $\frac{1}{16}$  auf. Die Wahrscheinlichkeit  $p$  für das *Fehlen* des Merkmals in der  $F_2$ -Generation ist für beide Vererbungsmodelle unterschiedlich und beträgt  $p_0 = \frac{1}{4}$  beim *dominanten* Erbgang und  $p_1 = \frac{7}{16}$  bei *komplementärer Polygenie*. Die beiden Modelle lassen sich formal durch zwei Hypothesen über die unbekannte Wahrscheinlichkeit  $p$  beschreiben:

$$\begin{aligned} \text{Nullhypothese} \quad H_0: p = p_0 & \quad \text{mit} \quad p_0 = \frac{1}{4} = 25\% \quad (\text{einfaches Modell}), \\ \text{Alternative (Hypothese)} \quad H_1: p = p_1 & \quad \text{mit} \quad p_1 = \frac{7}{16} = 43,75\% \quad (\text{Polygenie-Modell}). \end{aligned}$$

Hierbei haben wir den beiden Hypothesen schon Namen gegeben: *Nullhypothese* und *Alternative* oder *Hypothese*. Wie sich später noch zeigen wird, ist es von fundamentaler Bedeutung, welche der beiden Hypothesen als *Nullhypothese* angesehen wird, weil beide Hypothesen beim statistischen Test *unterschiedlich* behandelt werden. Zunächst wollen wir uns mit der Erklärung begnügen, daß die *Nullhypothese* (wie hier) oft eine bereits etablierte *Standard-Meinung* ist, während die *Alternative* typischerweise eine *neue Arbeitshypothese* ist, die man absichern möchte.

Wir betrachten nun ein Kreuzungsexperiment mit  $n = 40$  Nachkommen in der  $F_2$ -Generation. Die *absolute* Anzahl  $X$  der Nachkommen *ohne* das Merkmal ist dann  $B(n, p)$ -verteilt, wobei sich unter den jeweiligen Hypothesen unterschiedliche Verteilungen ergeben (vgl. Abb. 1). Die Entscheidung zwischen beiden Hypothesen soll nun aufgrund des konkreten Ausgangs des Experiments erfolgen, d.h. auf der Basis der beobachteten absoluten Häufigkeit  $x$  als der Realisierung der Zufallsvariablen  $X$ .

Bezeichnet  $\hat{p}(x) = \frac{1}{n}x$  die beobachtete relative Häufigkeit, so werden wir uns im Fall  $\hat{p}(x) \leq p_0 < p_1$  für die Nullhypothese  $H_0: p = p_0$  entscheiden, weil solche Realisierungen  $\hat{p}(x)$  unter der Alternative  $H_1: p = p_1$  unwahrscheinlich sind (vgl. Abb. 1). Und der Fall  $p_0 < p_1 \leq \hat{p}(x)$  spricht eher für die Alternative. Unklar ist aber, wie wir uns entscheiden sollen, wenn  $\hat{p}(x)$  zwischen  $p_0$  und  $p_1$  liegt. Hierzu werden wir eine „kritische Grenze“  $p_{krit}$  zwischen  $p_0$  und  $p_1$  festlegen, und uns für die Alternative entscheiden, wenn  $\hat{p}(x) \geq p_{krit}$  gilt. Bevor wir eine solche Grenze  $p_{krit}$  festlegen, wollen wir uns über die Konsequenzen klar werden. Dazu betrachten wir die möglichen *Fehlentscheidungen* des Tests.



**Abb. 1:** Die  $B(n, p)$ -Verteilung der *absoluten* Häufigkeiten  $X$  (*obere x-Achsen*) für  $n = 40$  (Umfang der  $F_2$ -Generation) unter beiden Hypothesen (*links bzw. rechts*). Die *unteren x-Achsen* gelten für die Verteilung der *relativen* Häufigkeit  $\hat{p}(X) = \frac{1}{n}X$ . Die glatte Kurve ist die Dichte der zugehörigen Normal-Approximation.

### Testentscheidungen und Fehlerrisiken

Ein *Fehler 1. Art* bzw. eine *falsch-positive Entscheidung* liegt vor, wenn der Test sich für die *Alternative* entscheidet, obwohl die Nullhypothese zutrifft (vgl. Tab. 1). Und ein *Fehler 2. Art* bzw. eine *falsch-negative Entscheidung* besteht darin, daß der Test sich für die *Nullhypothese* entscheidet, obwohl die Alternative zutrifft. Da die Testentscheidung von der beobachteten Realisierung  $\hat{p}(x)$  der Zufallsvariablen  $\hat{p}(X)$  abhängt, sind die Fehlentscheidungen ebenfalls *zufällig* und man kann deren *Wahrscheinlichkeiten* betrachten. Das *Fehlerrisiko 1. Art*  $\alpha$  ist die Wahrscheinlichkeit für einen Fehler 1. Art, d.h. die Wahrscheinlichkeit, sich für die *Alternative* zu entscheiden, obwohl diese *nicht* zutrifft. Und das *Fehlerrisiko 2. Art*  $\beta$  ist die Wahrscheinlichkeit für einen Fehler 2. Art, d.h. die Wahrscheinlichkeit, sich für die *Nullhypothese* zu entscheiden, obwohl diese *nicht* zutrifft.

Tabelle 1: Testentscheidungen und Fehlerrisiken

<b>Testentscheidung</b>	in Wirklichkeit trifft zu	
	<b>Nullhypothese</b>	<b>Alternative (Hypothese)</b>
<b>Nullhypothese</b> (keine Ablehnung der Nullhypothese) negative Entscheidung	<b>richtige (negative) Entscheidung</b> <b>Sicherheit 1. Art</b> $1 - \alpha$	<b>falsche (negative) Entscheidung</b> <b>Fehler 2. Art</b> <b>Fehlerrisiko 2. Art</b> $\beta$
<b>Alternative (Hypothese)</b> (Ablehnung der Nullhypothese) positive Entscheidung	<b>falsche (positive) Entscheidung</b> <b>Fehler 1. Art</b> <b>Fehlerrisiko 1. Art</b> $\alpha$	<b>richtige (positive) Entscheidung</b> <b>Sicherheit 2. Art,</b> <b>Schärfe, Power</b> $1 - \beta$

Zur Illustration gehen wir bei dem hier betrachteten Beispiel davon aus, daß das einfache Vererbungsmodell (Nullhypothese) die bisherige *Lehrbuchmeinung* ist, während die komplementäre Polygenie (Alternative) eine neue „revolutionäre“ Theorie darstellt. Falls aufgrund der experimentellen Ergebnisse die Nullhypothese abgelehnt wird, so soll die neue Theorie publiziert werden, während andernfalls noch weiter experimentiert werden soll. Der Fehler 1. Art besteht hier darin, die vermeintlich neue Theorie zu publizieren, obwohl sie nicht korrekt ist, d.h. eine wissenschaftliche *Blamage*. Und der Fehler 2. Art bedeutet hier, daß man die neue Theorie nicht publiziert (und weiter experimentiert), obwohl sie richtig ist, also eine *Verzögerung* von wissenschaftlicher Erkenntnis.

Die beiden Fehlerrisiken hängen von der Wahl der kritischen Grenze  $p_{krit}$  ab, und sie lassen sich als Binomial-Wahrscheinlichkeiten berechnen (vgl. Abb. 2)

$$\alpha = P\{\hat{p}(X) \geq p_{krit} \mid p = p_0\} = P\{X \geq np_{krit} \mid p = p_0\} \quad (1)$$

$$\beta = P\{\hat{p}(X) < p_{krit} \mid p = p_1\} = P\{X < np_{krit} \mid p = p_1\} \quad (2)$$

*Vergrößert* man den kritischen Wert  $p_{krit}$ , so wird zwar das Fehlerrisiko  $\alpha$  *kleiner*, aber gleichzeitig wird das andere Risiko  $\beta$  *größer*. Und umgekehrt wird bei *Verringerung* von  $p_{krit}$  zwar  $\beta$  *geringer*, dafür aber  $\alpha$  wieder *größer*. Folglich lassen sich durch die Wahl der kritischen Grenze  $p_{krit}$  nicht *beide* Fehlerrisiken *gleichzeitig* gering halten. Der Ausweg aus diesem Dilemma besteht darin, daß man Prioritäten setzt, indem man den beiden Fehlerrisiken unterschiedliche Bedeutung beimißt. Beim statistischen Test hat man sich darauf geeinigt, das Fehlerrisko 1. Art durch die *Vorgabe* des Wertes  $\alpha$  zu kontrollieren, wobei typischerweise (wie bei den Konfidenzgrenzen)  $\alpha = 5\%$  als *Standardwert* verwendet wird. Die Wahl von  $\alpha$  in einer konkreten Anwendungssituation sollte der inhaltlichen Bedeutung des Fehlers 1. Art Rechnung tragen, d.h. bei gravierenden Konsequenzen eines Fehlers 1. Art sollte  $\alpha$  auch entsprechend gering sein, z.B. 1% oder 0,1%. Für das vorgegebene Fehlerrisiko  $\alpha$  läßt sich dann der zugehörige kritische Wert  $p_\alpha = p_{krit}$  aus (1) bestimmen. Das Fehlerrisiko 2. Art  $\beta$  wird dadurch *nicht* kontrolliert und ist typischerweise *wesentlich größer* als  $\alpha$ .

Wegen dieser ungleichen Behandlung beider Fehlerrisiken spielt es eine entscheidende Rolle, welche der beiden Hypothesen man als *Nullhypothese* und welche als *Alternative* bezeichnet. Man beachte, daß sich beim Vertauschen beider Hypothesen auch die zugehörigen Fehlerrisiken vertauschen. Deshalb kann man die Hypothesen immer so wählen, daß der inhaltlich *wichtigere* Fehler auch zum Fehler 1. Art wird, der dann unmittelbar durch den Test über die Vorgabe von  $\alpha$  *kontrolliert* wird. – Im konkreten Beispiel aus der Genetik haben beide Fehler sehr unterschiedliche Bedeutung, und die von uns gewählte Formulierung des Testproblems kontrolliert das Risiko einer *wissenschaftlichen Blamage*, aber *nicht* das Risiko für die *Verzögerung von Erkenntnis*.

### Der normale (asymptotische) Test

Wir wollen jetzt für vorgegebenes  $\alpha$  (z.B. 5%) den zugehörigen kritischen Wert  $p_{krit}$  aus (1) zunächst *approximativ* bestimmen. Hierzu betrachten wir als sogenannten *Testwert* die standardisierte Abweichung der Schätzung  $\hat{p}(x)$  vom Wert  $p_0$  aus der Nullhypothese

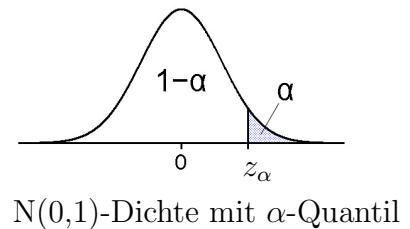
$$t(x) = \frac{(\hat{p}(x) - p_0) \cdot \sqrt{n}}{\sigma_0} \quad (\text{Testwert}) \quad \text{mit} \quad \sigma_0 = \sqrt{p_0(1 - p_0)}. \quad (3)$$

Die zugehörige Zufallsvariable  $t(X)$  ist unter  $H_0: p = p_0$  bei nicht zu kleinem Umfang  $n$  approximativ  $N(0, 1)$ -verteilt. Mit dem *oberen  $\alpha$ -Quantil*  $z_\alpha$  von  $N(0, 1)$ , gegeben durch (vgl. auch 3.1.3 und Anhang T)

$$\Phi(z_\alpha) = 1 - \alpha \quad (4)$$

ergibt sich  $p_{krit}$  aus (1) *approximativ* zu

$$p_{krit} = p_{krit}(\alpha) = p_0 + \frac{1}{\sqrt{n}} \sigma_0 z_\alpha. \quad (5)$$



Für vorgegebenes Fehlerrisiko  $\alpha$  erhalten wir damit als endgültiges Entscheidungskriterium: die *Nullhypothese* wird genau dann *abgelehnt* (d.h. man entscheidet sich für die *Alternative*), wenn  $\hat{p}(x) \geq p_{krit}(\alpha)$  bzw. wenn  $t(x) \geq z_\alpha$  ist:

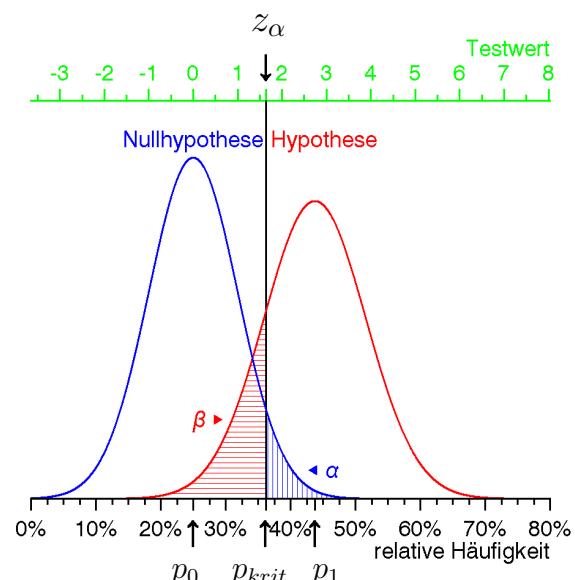
$$\begin{aligned} \text{Normaler Test: Ablehnung von } H_0: p = p_0 &\Leftrightarrow \hat{p}(x) \geq p_{krit}(\alpha) \\ &\Leftrightarrow t(x) \geq z_\alpha. \end{aligned} \quad (6)$$

Konkret für  $\alpha = 5\%$  ist  $z_\alpha = 1,645$  und mit  $n = 40$  und  $p_0 = 25\%$  erhält man den kritischen Wert  $p_{krit} = 36,26\%$ . Das zugehörige Fehlerrisiko 2. Art ergibt sich aus (2) zu  $\beta \approx 17\%$  (vgl. Abb. 2).

Abb. 2: Testentscheidung und Fehlerrisiken

Für relative Häufigkeiten  $\hat{p}(x)$  (untere Skala) bzw. Testwerte  $t(x)$  (obere Skala) **rechts** vom senkrechten Strich wird die Nullhypothese abgelehnt und die zugehörige Fläche unter der Dichte ist das Fehlerrisiko 1. Art  $\alpha$  (hier: 5%).

Für relative Häufigkeiten  $\hat{p}(x)$  bzw. Testwerte  $t(x)$  **links** vom senkrechten Strich wird die Nullhypothese nicht abgelehnt und die zugehörige Fläche unter der Dichte ist das Fehlerrisiko 2. Art  $\beta$  (hier: 17%).



**Experiment 1:** Bei einem konkreten Kreuzungsexperiment fehlte das Merkmal bei  $x = 14$  der  $n = 40$  Nachkommen, d.h. bei  $\hat{p} = 35\%$  aller Nachkommen. Da der kritische Wert  $p_{krit} = 36,26\%$  nicht erreicht wurde, entscheidet sich der Test für die Nullhypothese, d.h. das *einfache Modell* („Lehrbuchmeinung“). Da die Nullhypothese nicht abgelehnt wurde, kann ein Fehler 2. Art vorliegen, dessen Fehlerrisiko oben zu  $\beta \approx 17\%$  berechnet wurde.  $\square$

Das Entscheidungskriterium (6) lässt sich noch auf andere Weise äquivalent umformulieren (vgl. Abb. 4 oben):

$$t(x) \geq z_\alpha \quad \Leftrightarrow \quad \Phi(-t(x)) \leq \alpha. \quad (7)$$

Hierbei ist  $\Phi(-t(x)) = 1 - \Phi(t(x))$  die Wahrscheinlichkeit, daß die unter der Nullhypothese (approximativ)  $N(0, 1)$ -verteilte Testvariable  $t(X)$  mindestens so groß ist, wie ihre beobachtete Realisierung  $t(x)$  bzw., daß die zufällige relative Häufigkeit  $\hat{p}(X)$  mindestens so groß ist wie ihre beobachtete Realisierung  $\hat{p}(x)$ :

$$\Phi(-t(x)) \approx P\{t(X) \geq t(x) \mid p = p_0\} = P\{\hat{p}(X) \geq \hat{p}(x) \mid p = p_0\}. \quad (8)$$

Die Wahrscheinlichkeit  $\Phi(-t(x))$  wird als *Signifikanz der Beobachtung* oder als *P-Wert* bezeichnet. Da hohe Werte von  $t(x)$  bzw.  $\hat{p}(x)$  gegen die Nullhypothese sprechen, ist der *P-Wert* ein Maß für die „Glaubwürdigkeit“ der Nullhypothese im Lichte der Beobachtung  $x$ . Man lehnt die Nullhypothese deshalb genau dann ab, wenn dieser *P-Wert* zu gering ist, d.h. das Fehlerrisiko  $\alpha$  unterschreitet.

Eine weitere äquivalente Formulierung der Testentscheidung verwendet die normale untere Konfidenzgrenze  $\tilde{p}_{un,\alpha}(x)$  zur (approximativen) Sicherheit  $1 - \alpha$  bzw. Irrtums-Wahrscheinlichkeit  $\alpha$  (vgl. 3.1.3 a)). Und zwar wird die Nullhypothese  $H_0: p = p_0$  genau dann abgelehnt, wenn der Wert  $p_0$  die untere Konfidenzgrenze  $\tilde{p}_{un,\alpha}(x)$  unterschreitet (und somit zu „unwahrscheinlich“ ist), d.h.

$$\text{Normaler Test: Ablehnung von } H_0: p = p_0 \Leftrightarrow p_0 \leq \tilde{p}_{un,\alpha}(x). \quad (9)$$

Damit können wir die Testentscheidung durch unterschiedliche – aber im Ergebnis äquivalente – Methoden herbeiführen, und zwar durch den Vergleich

- des *Testwerts*  $t(x)$  mit dem *Quantil*  $z_\alpha$ ,
- der *Schätzung*  $\hat{p}(x)$  mit dem *kritischen Wert*  $p_{krit}(\alpha)$ ,
- des *P-Werts*  $\Phi(-t(x))$  mit dem *Fehlerrisiko*  $\alpha$ ,
- der *unteren Konfidenzgrenze*  $\tilde{p}_{un,\alpha}(x)$  mit dem *Wert*  $p_0$  aus der Nullhypothese.

Obwohl es für die Testentscheidung ausreicht, nur *einen* dieser vier Vergleiche durchzuführen, so ist es gute statistische Praxis, *sowohl* den *P-Wert* als auch die untere Konfidenzgrenze anzugeben, weil beide Größen eine zusätzliche Information – über die reine Testentscheidung hinaus – liefern. Rein rechentechnisch unterscheiden sich die vier Vergleiche insofern, als man für den *P-Wert* die *Funktion*  $\Phi$  an der Stelle  $-t(x)$  bestimmen muß, während man bei den restlichen 3 Vergleichen jeweils das  $\alpha$ -Quantil  $z_\alpha$  benötigt.

**Experiment 1 (Fortsetzung 1):** Für das bereits oben behandelte Experiment mit  $n = 40$  Nachkommen, unter denen das Merkmal in  $x = 14$  Fällen fehlte, sind zur Illustration alle vier Vergleiche zusammengestellt:

$$\begin{aligned} t(x) &= 1,461 &< & 1,645 = z_{5\%} \\ \hat{p} &= 35,0\% &< & 36,26\% = p_{krit}(5\%) \\ \varPhi(-t(x)) &= 7,21\% &> & 5,0\% = \alpha \\ \tilde{p}_{un,5\%} &= 23,91\% &< & 25,0\% = p_0. \end{aligned}$$

Der  $P$ -Wert von 7,21% bedeutet, daß das eingetretene Ereignis  $\{t(X) \geq t(x)\}$  bzw.  $\{\hat{p}(X) \geq \hat{p}(x)\}$  unter der Nullhypothese  $H_0: p = p_0$  eine Wahrscheinlichkeit von 7,21% hat und somit nicht zu „unwahrscheinlich“ ist. Und die *untere Konfidenzgrenze*  $\tilde{p}_{un,5\%}(x) = 23,91\%$  besagt, daß alle Werte oberhalb von 23,91% (und insbesondere  $p_0 = 25\%$ ) noch mit der Beobachtung  $x$  „verträglich“ (d.h. nicht zu „unwahrscheinlich“) sind.  $\square$

### Erweiterung der Hypothesen

Betrachten wir noch einmal die *Entscheidungskriterien* für die Ablehnung der Nullhypothese, so fällt auf, daß diese *nicht* von dem konkreten Wert  $p_1 = \frac{7}{16}$  aus der Alternative abhängen. Lediglich das *Fehlerrisiko 2. Art* hängt vom Wert  $p_1$  ab. Folglich ergibt sich für jeden Wert  $p_1 > p_0$  *dasselbe* Entscheidungskriterium. Deshalb können wir die Alternative jetzt auch in der *erweiterten* Form  $H: p > p_0$  fassen, die wir bereits bei anderen Anwendungen (Wahlumfrage, Therapie-Erfolg in 3.2) kennengelernt haben. Zu beachten ist jedoch, daß das Fehlerrisiko 2. Art  $\beta(p)$  jetzt vom Wert  $p > p_0$  abhängt (was wir in der Notation berücksichtigen), wobei das Risiko  $\beta(p)$  *geringer* wird, je *weiter*  $p$  von  $p_0$  entfernt ist. Oder umgekehrt:  $\beta(p_1)$  wird *größer*, je *näher*  $p$  oberhalb von  $p_0$  liegt (vgl. Abb. 3 rechts). Wenn  $p$  gegen  $p_0$  fällt, so wächst  $\beta(p)$  sogar bis  $1 - \alpha$  an, d.h. bis zum komplementären Fehlerrisiko 1. Art.

Gehen wir nun auch von der bisherigen Nullhypothese  $H_0: p = p_0$  zu der *erweiterten* Nullhypothese  $H_0: p \leq p_0$  über, so hängt das Fehlerrisiko 1. Art  $\alpha(p)$  jetzt vom Wert  $p \leq p_0$  ab. Dieses Risiko  $\alpha(p)$  *wächst*, je dichter der Wert  $p$  unterhalb von  $p_0$  liegt, und wird *maximal* für  $p = p_0$  (vgl. Abb. 3 links). Folglich ist beim bisherigen (normalen) Test das Fehlerrisiko 1. Art  $\alpha(p)$  für *jeden* Wert  $p$  unter der erweiterten Nullhypothese  $H_0: p \leq p_0$  noch höchstens gleich der Vorgabe  $\alpha$  (z.B. 5%).

Dies führt zu folgendem Test der erweiterten Hypothesen  $H_0: p \leq p_0$  vs.  $H: p > p_0$ , dessen (maximales) Fehlerrisiko 1. Art (approximativ) der Vorgabe  $\alpha$  entspricht

$$\text{Normaler Test: Ablehnung von } H_0: p \leq p_0 \Leftrightarrow t(x) \geq z_\alpha. \quad (10)$$

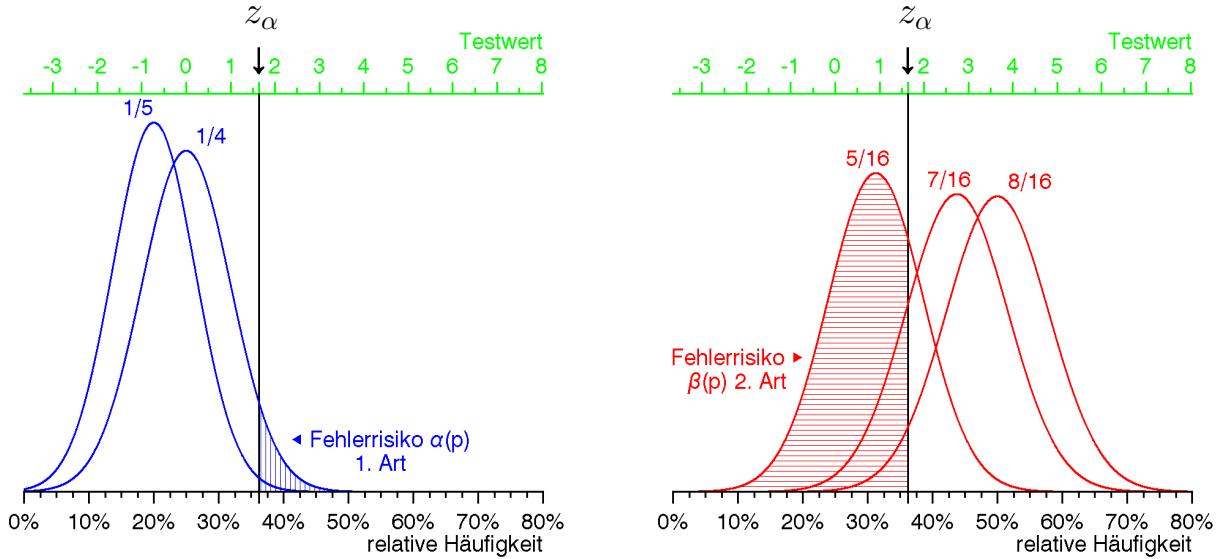


Abb. 3:

links: Unter der erweiterten Nullhypothese  $H_0: p \leq p_0$  hängt das Fehlerrisiko 1. Art  $\alpha(p)$  von  $p \leq p_0$  ab und wird für  $p = p_0$  maximal.

rechts: Unter der erweiterten Alternative  $H: p > p_0$  hängt das Fehlerrisiko 2. Art  $\beta(p)$  von  $p > p_0$  ab und wird größer, je geringer  $p$  von  $p_0$  abweicht.

### Exakter Test

Bisher haben wir bei allen Betrachtungen aus didaktischen Gründen die einfachere *Normal-Approximation* anstelle der *exakten* Verteilung von  $X$  bzw.  $\hat{p}(X)$  oder  $t(X)$  verwendet (vgl. Abb. 1). Dies liefert bei *nicht zu kleinem* Stichprobenumfang  $n$  auch hinreichend genaue Approximationen, d.h. das Fehlerrisiko 1. Art ist approximativ höchstens gleich der Vorgabe  $\alpha$  (von z.B. 5%). Will man das Fehlerrisiko 1. Art nicht nur *approximativ*, sondern *exakt* durch die Vorgabe  $\alpha$  beschränkt wissen, so sollte man einen entsprechenden *exakten Test* durchführen. Dieser unterscheidet sich von bisherigen *normalen* (oder *asymptotischen*) Tests lediglich dadurch, daß man statt der *normalen* jetzt die *exakte untere Konfidenzgrenze*  $\hat{p}_{un,\alpha}(x)$  aus 3.1.2 a verwendet, d.h. die Testentscheidung ist statt (9) beim *exakten Test* gegeben durch

$$\text{Exakter Test: } \text{Ablehnung von } H_0: p = p_0 \Leftrightarrow p_0 \leq \hat{p}_{un,\alpha}(x). \quad (11)$$

Als äquivalentes Kriterium für die Ablehnung der Nullhypothese ergibt sich

$$p_0 \leq \hat{p}_{un,\alpha}(x) \Leftrightarrow P\{\hat{p}(X) \geq \hat{p}(x) \mid p = p_0\} \leq \alpha \quad (12)$$

wobei die Wahrscheinlichkeit  $P\{\hat{p}(X) \geq \hat{p}(x) \mid p = p_0\}$  als *exakter P-Wert* oder *exaktes Signifikanzniveau der Beobachtung* bezeichnet wird. Dieser exakte P-Wert kann wieder als „Glaubwürdigkeit“ der Nullhypothese  $H_0: p = p_0$  im Lichte der Beobachtung  $x$  interpretiert werden. – Auf die restlichen beiden äquivalenten Entscheidungskriterien unter Verwendung eines *exakten* kritischen Werts für  $\hat{p}(x)$  bzw.  $t(x)$  wollen wir hier nicht näher eingehen, da diese einerseits etwas aufwendiger zu bestimmen sind und andererseits die Testentscheidung auch ohne sie herbeigeführt werden kann.

Der exakte Test sollte in jedem Fall bei *kleinem* Umfang  $n$  angewandt werden, weil dann die dem normalen Test zugrunde liegenden Approximationen unzuverlässig sein können.

**Experiment 1 (Fortsetzung 2):** Für das obige Experiment mit  $n = 40$  und  $x = 14$  lehnt auch der *exakte* Test die Nullhypothese *nicht* ab, weil:

$$\begin{aligned} P\{\hat{p}(X) \geq \hat{p}(x) \mid p = p_0\} &= 10,32\% > 5,0\% = \alpha. \\ \hat{p}_{un,\alpha}(x) &= 22,55\% < 25,0\% = p_0. \end{aligned}$$

□

### 3.2.2 Ein- und zweiseitige Tests für eine Wahrscheinlichkeit

Nachdem wir die grundlegenden Ideen eines Tests am Beispiel diskutiert haben, wollen wir jetzt die verschiedenen Tests über die unbekannte Wahrscheinlichkeit  $p$  systematisch zusammenstellen (vgl. *Test-Box 1*) und noch einmal kurz erläutern.

#### Einseitiger Test mit oberer Alternative

Wir beginnen mit den *einseitigen* Hypothesen mit *oberer* Alternative (der Name kommt daher, daß die Alternative aus allen Werten von  $p$  besteht, die auf *einer Seite* – und zwar *oberhalb* – des Referenzwertes  $p_0$  liegen):

$$\begin{aligned} \text{Nullhypothese } H_0: p \leq p_0 \quad (\text{Referenzwert nicht überschritten}) &\quad \text{vs.} \quad (1) \\ \text{Alternative } H_>: p > p_0 \quad (\text{Referenzwert überschritten}). \end{aligned}$$

Als Beispiel hierfür haben wir bereits die Wahlumfragen (in **3.2**) erwähnt. Das in **3.2.1** diskutierte Anwendungsbeispiel aus der Genetik läßt sich ebenfalls in diese Form bringen, wenn man die ursprünglich nur aus zwei Werten  $p_0 < p_1$  bestehenden Hypothesen  $H_0: p = p_0$  und  $H_1: p = p_1$  wie dort besprochen zu (1) *erweitert*. Da wir dieses Testproblem ausführlich in **3.2.1** behandelt haben, wollen wir die erforderlichen Schritte für den *normalen (asymptotischen)* bzw. *exakten* Test hier nicht mehr wiederholen und verweisen auf die Zusammenfassung in *Test-Box 1*.

#### Einseitiger Test mit unterer Alternative

Am Beispiel der klinischen Studie in **3.2** sind wir bereits auf das *einseitige* Testproblem mit *unterer* Alternative gestoßen:

$$\begin{aligned} \text{Nullhypothese } H_0: p \geq p_0 \quad (\text{Referenzwert nicht unterschritten}) &\quad \text{vs.} \quad (2) \\ \text{Alternative } H_<: p < p_0 \quad (\text{Referenzwert unterschritten}). \end{aligned}$$

Obwohl man diese Hypothesen durch Übergang auf die komplementäre Wahrscheinlichkeit  $q = 1 - p$  auch auf die Form (1) zurückführen kann, wollen wir die zugehörigen Tests zumindest kurz erläutern (eine Zusammenfassung findet sich in der *Test-Box 1*).

Hier sprechen *kleine* Werte (unterhalb von  $p_0$ ) der relativen Häufigkeit  $\hat{p}(x)$  bzw. *negative* Testwerte  $t(x)$  eher für die Alternative. Dies führt zur Testentscheidung

$$\begin{aligned} \text{Normaler Test: } & \text{ Ablehnung von } H_0: p \geq p_0 \Leftrightarrow \\ & t(x) \leq -z_\alpha \Leftrightarrow \\ & \hat{p}(x) \leq p'_{krit}(\alpha) := p_0 - \frac{1}{\sqrt{n}}\sigma_0 z_\alpha. \end{aligned} \quad (3)$$

Die Testentscheidung lässt sich hier auch unter Verwendung des *P-Werts* (vgl. Abb. 4 *Mitte*)

$$\Phi(t(x)) \approx P\{t(X) \leq t(x) \mid p = p_0\} = P\{\hat{p}(X) \leq \hat{p}(x) \mid p = p_0\} \quad (4)$$

oder der normalen *oberen Konfidenzgrenze*  $\tilde{p}_{ob,\alpha}(x)$  formulieren:

$$t(x) \leq -z_\alpha \Leftrightarrow \Phi(t(x)) \leq \alpha \Leftrightarrow \tilde{p}_{ob,\alpha}(x) \leq p_0. \quad (5)$$

Und der zugehörige *exakte* Test verwendet die *exakte obere Konfidenzgrenze*  $\hat{p}_{ob,\alpha}(x)$  bzw. den *exakten P-Wert*

$$\begin{aligned} \text{Exakter Test: } & \text{ Ablehnung von } H_0: p \geq p_0 \Leftrightarrow \hat{p}_{ob,\alpha}(x) \leq p_0 \quad (6) \\ & \Leftrightarrow P\{\hat{p}(X) \leq \hat{p}(x) \mid p = p_0\} \leq \alpha. \end{aligned}$$

### Zweiseitiger Test

Es gibt Anwendungen, bei denen man sich dafür interessiert, ob der Referenzwert  $p_0$  *genau* eingehalten wird (Nullhypothese) oder ob eine Abweichung *nach oben oder unten* (also *zweiseitig*) vorliegt. Dies ist beim Genetik-Beispiel aus 3.2 der Fall, wenn kein anderweitiges Vererbungsmodell (wie z.B. die komplementäre Polygenie in 3.2.1) als Alternative vorliegt. Das zugehörige *zweiseitige* Testproblem lautet

$$\begin{aligned} \text{Nullhypothese } & H_0: p = p_0 \quad (\text{Referenzwert } \text{eingehalten}) \quad \text{vs.} \quad (7) \\ \text{Alternative } & H_\neq: p \neq p_0 \quad (\text{Referenzwert } \text{nicht eingehalten}). \end{aligned}$$

Wegen

$$H_\neq: p \neq p_0 \Leftrightarrow H_>: p > p_0 \quad \text{oder} \quad H_<: p < p_0 \quad (8)$$

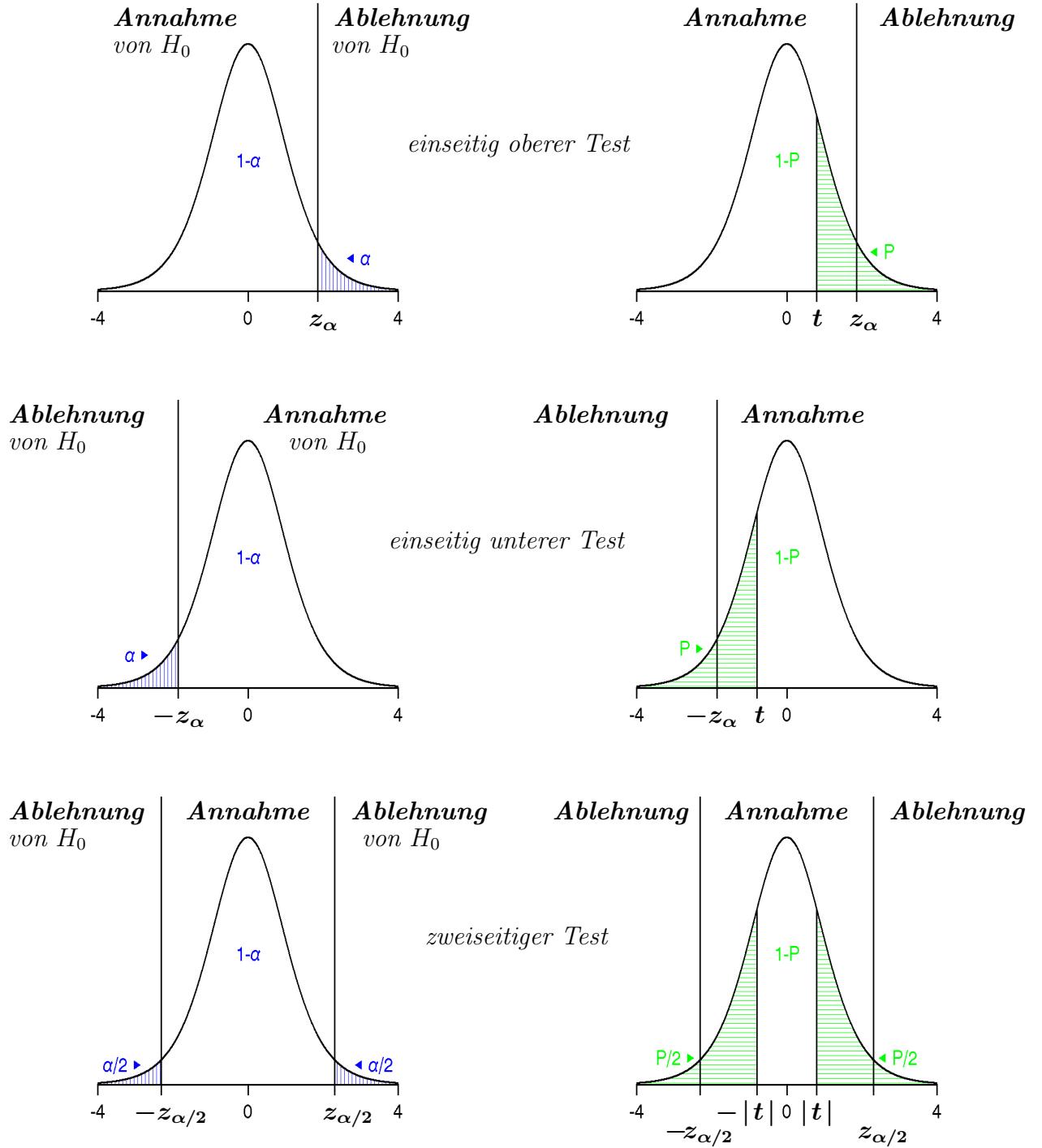
ist folgende Testprozedur naheliegend. Man führt *beide einseitigen* Tests

$$H_0: p \leq p_0 \quad \text{vs.} \quad H_>: p > p_0 \quad \text{und} \quad H_0: p \geq p_0 \quad \text{vs.} \quad H_<: p < p_0$$

durch. Dann lehnt man die *zweiseitige Nullhypothese*  $H_0: p = p_0$  ab, wenn *einer* der beiden einseitigen Tests die *einseitige Nullhypothese abgelehnt hat*. Der so zusammengesetzte *zweiseitige Test* hat als Fehlerrisiko 1. Art allerdings die *Summe* der beiden Fehlerrisiken 1. Art *beider* einseitiger Tests, also  $2\alpha$ . Damit das Fehlerrisiko 1. Art beim zweiseitigen Test der Vorgabe von  $\alpha$  (z.B. 5%) entspricht, müssen *beide* einseitige Tests mit  $\frac{\alpha}{2}$  statt  $\alpha$  (z.B. 2,5% statt 5%) durchgeführt werden (vgl. Abb. 4 *unten*). Die hierfür erforderlichen Rechenschritte sind in der *Test-Box 1* zusammengestellt.

<b>Test-Box 1:</b> Tests über die Wahrscheinlichkeit $p$ einer $B(n, p)$ -verteilten Zufallsvariablen $X$ .		
einseitig oben	einseitig unten	zweiseitig
$H_0: p \leq p_0$ vs. $H: p > p_0$	$H_0: p \geq p_0$ vs. $H: p < p_0$	$H_0: p = p_0$ vs. $H: p \neq p_0$
Fehlerrisiko 1. Art Testniveau: $\alpha$		
Stichprobenumfang: $n$		
beobachtete absolute Häufigkeit: $x$ (Realisierung von $X$ )		
beobachtete relative Häufigkeit: $\hat{p}(x) = \frac{1}{n} x$ (Schätzung von $p$ )		
<b>Asymptotischer (normaler) Test:</b> Fehlerrisiko 1. Art $\approx \alpha$		
Anwendung: bei <i>nicht zu kleinem</i> $n$ ,		Faustregel: $n p_0(1 - p_0) \geq 5$
Testwert: $t = \frac{(\hat{p}(x) - p_0)\sqrt{n}}{\sigma_0}$ mit $\sigma_0 = \sqrt{p_0(1 - p_0)}$		
Entscheidung: Ablehnung von $H_0$ (Annahme von $H$ ), falls gilt		
einseitig oben	einseitig unten	zweiseitig
$Testwert t \geq z_\alpha$ bzw. $P\text{-Wert } \Phi(-t) \leq \alpha$ bzw. $p_0 \leq \tilde{p}_{un,\alpha}(x)$	$Testwert t \leq -z_\alpha$ bzw. $P\text{-Wert } \Phi(t) \leq \alpha$ bzw. $\tilde{p}_{ob,\alpha}(x) \leq p_0$	$Testwert  t  \geq z_{\frac{\alpha}{2}}$ bzw. $P\text{-Wert } 2\Phi(- t ) \leq \alpha$ bzw. $p_0 \leq \tilde{p}_{un,\frac{\alpha}{2}}(x) \quad oder \quad \tilde{p}_{ob,\frac{\alpha}{2}}(x) \leq p_0$
<b>Exakter Test:</b> Fehlerrisiko 1. Art $\leq \alpha$		
Anwendung: immer, zwingend bei <i>kleinem</i> $n$		
Entscheidung: Ablehnung von $H_0$ (Annahme von $H$ ), falls gilt		
einseitig oben	einseitig unten	zweiseitig
$P\{X \geq x   p = p_0\} \leq \alpha$ (exakter P-Wert) bzw. $p_0 \leq \hat{p}_{un,\alpha}(x)$	$P\{X \leq x   p = p_0\} \leq \alpha$ (exakter P-Wert) bzw. $\hat{p}_{ob,\alpha}(x) \leq p_0$	$P\{X \geq x   p = p_0\} \leq \frac{\alpha}{2} \quad oder \quad P\{X \leq x   p = p_0\} \leq \frac{\alpha}{2}$ bzw. $p_0 \leq \hat{p}_{un,\frac{\alpha}{2}}(x) \quad oder \quad \hat{p}_{ob,\frac{\alpha}{2}}(x) \leq p_0$

**Erläuterung zur Test-Box 1:** Das  $\alpha$ -Quantil  $z_\alpha$  bzw. die Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung ist im Anhang **T** tabelliert oder kann per Computer berechnet werden. Die Berechnung der *normalen* bzw. *exakten* Konfidenzgrenzen  $\tilde{p}_{un}, \tilde{p}_{ob}$  bzw.  $\hat{p}_{un}, \hat{p}_{ob}$  ist in der *Konfidenz-Box 1* (in **3.1.4**) erläutert.



**Abb. 4:** Ablehnungsbereiche und P-Wert beim normalen (asymptotischen) Test

**links:** Dichte der  $N(0, 1)$ -verteilten Teststatistik  $t(X)$  für  $p = p_0$  mit Ablehnungsbereichen der Nullhypothese beim ein- und zweiseitigen Test zum Fehlerrisiko 1. Art  $\alpha = 5\%$ .

**rechts:** Der zum Testwert  $t = t(x)$  gehörige P-Wert  $P$  als markierte Fläche.

### 3.2.3 Fehlerrisiko 2. Art und Testschärfe (Power)

Im Gegensatz zum Fehlerrisiko 1. Art, das beim Test der Vorgabe von  $\alpha$  entspricht, wird das Fehlerrisiko 2. Art *nicht* durch den Test kontrolliert. Wenn man die Nullhypothese nicht abgelehnt hat, so ist ein Fehler 2. Art möglich, dessen Risiko man (zumindest für

relevante Werte aus der Alternative) berechnen sollte und zwar eigentlich schon bei der Planung des Versuchs (vgl. auch **3.2.4**).

Zur Illustration betrachten wir zuerst das *einseitig obere* Testproblem

$$H_0: p \leq p_0 \quad vs. \quad H: p > p_0 \quad (1)$$

Wenn die Alternative zutrifft, also  $p > p_0$  ist, so hängt das Fehlerrisiko 2. Art  $\beta(p)$  vom unbekannten wahren Wert  $p$  ab. Man kann – und sollte – es deshalb für praktisch relevante (hypothetische) Werte  $p_1 > p_0$  berechnen. Wir geben hierfür zwei inhaltliche Beispiele.

**Dominanz vs. komplementäre Polygenie (Fortsetzung 1):** Hier war  $p_0 = \frac{1}{4}$  und aus der Alternative hat uns speziell der Wert  $p_1 = \frac{7}{16}$  interessiert, der dem *Polygenie-Modell* entspricht. Das zugehörige Fehlerrisiko  $\beta(p_1) \approx 17\%$  haben wir zwar schon angegeben, aber noch nicht ausgeführt, wie es *berechnet* wurde.  $\square$

**Wahlumfrage (Fortsetzung 1):** Wir betrachten eine kleine Partei, die wissen möchte, ob sie an der 5%-Hürde scheitert (Nullhypothese) oder nicht (Alternative). Dies entspricht dem Testproblem (1) mit  $p_0 = 5\%$  (vgl. auch **3.2**). Der *Fehler 1. Art* besteht darin, daß sie Neuwahlen zustimmt, obwohl sie an der 5%-Hürde scheitern würde („katastrophale Wahlniederlage“). Daher ist ein *extrem kleines*  $\alpha = 0,5\%$  angemessen. Der *Fehler 2. Art* besteht darin, daß sie Neuwahlen nicht zustimmt, obwohl der Einzug ins Parlament gesichert wäre, d.h. ein günstiger Wahltermin wird nicht genutzt. Hier ist z.B. von Interesse, wie groß das Fehlerrisiko  $\beta(p)$  ist, wenn der wahre Stimmanteil mit  $p = 7\%$  deutlich über 5% läge.  $\square$

Ausgangspunkt für die Berechnung des Fehlerrisiko 2. Art  $\beta(p)$  ist die approximative Normalverteilung der Teststatistik  $t(X)$ . Wie in Abb. 3 rechts erkennbar, entspricht das Fehlerrisiko  $\beta(p)$  einer Fläche unter der Dichte von  $t(X)$  und lässt sich daher unter Verwendung der Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung berechnen (zu Einzelheiten vgl. *Power-Box 1*).

Neben dem Fehlerrisiko 2. Art ist auch die *Schärfe* (engl.: *Power*) des Tests von Interesse. Hierunter versteht man ganz allgemein die Wahrscheinlichkeit, daß der Test die Nullhypothese *ablehnt*, d.h. sich für die Alternative entscheidet:

$$\text{Schärfe} = \text{Power} = P \{ \text{Test lehnt Nullhypothese ab} \}. \quad (2)$$

Wenn die Nullhypothese *falsch* ist, so ist die Schärfe also die Wahrscheinlichkeit, daß der Test dies auch *entdeckt*, d.h. die Nullhypothese auch *ablehnt*. In diesem Fall ist die Schärfe gerade das *komplementäre Fehlerrisiko 2. Art*  $\gamma(p)$ , d.h.

$$\gamma(p) = 1 - \beta(p) \quad (\text{Schärfe, Power}). \quad (3)$$

Je größer die Schärfe, desto kleiner das Fehlerrisiko 2. Art und umgekehrt.

Für das *einseitig untere* Testproblem

$$H_0: p \geq p_0 \quad \text{vs.} \quad H: p < p_0 \quad (4)$$

hängt das Fehlerrisiko 2. Art  $\beta(p)$  wieder vom Wert  $p < p_0$  ab. Es wird wieder *größer* je *dichter*  $p$  unterhalb von  $p_0$  liegt (zu Einzelheiten vgl. *Power-Box 1*).

Und beim *zweiseitigen* Testproblem

$$H_0: p = p_0 \quad \text{vs.} \quad H: p \neq p_0 \quad (5)$$

hängt das Fehlerrisiko 2. Art  $\beta(p)$  ebenfalls vom Wert  $p \neq p_0$  ab. Es wird *größer* je *weiter*  $p$  von  $p_0$  entfernt ist. Es kann (approximativ) berechnet werden wie das zugehörige Fehlerrisiko beim *einseitigen* Test zum *halben* Niveau  $\frac{\alpha}{2}$  (zu Einzelheiten vgl. *Power-Box 1*).

**Power-Box 1:** Für Tests über die Wahrscheinlichkeit  $p$  einer  $B(n, p)$  Verteilung

einseitig oben	einseitig unten	zweiseitig
$H_0: p \leq p_0$ vs. $H: p > p_0$	$H_0: p \geq p_0$ vs. $H: p < p_0$	$H_0: p = p_0$ vs. $H: p \neq p_0$
Fehlerrisiko 1. Art (Testniveau): $\alpha$		Stichprobenumfang: $n$
Wert aus der Alternative $H$ : $p_1$		
einseitig oben	einseitig unten	zweiseitig
$z = z_\alpha$	$z = z_\alpha$	$z = z_{\alpha/2}$
Hilfsgrößen:	$\Delta =  p_1 - p_0  > 0$ (Abweichung von $H_0$ ) $\sigma_0 = \sqrt{p_0(1-p_0)}$ , $\sigma_1 = \sqrt{p_1(1-p_1)}$ $u = \frac{z\sigma_0 - \Delta\sqrt{n}}{\sigma_1}$	
Fehlerrisiko 2. Art: Testschärfe (Power):	$\beta(p_1) \approx \Phi(u)$ mit $\Phi$ = Verteilungsfunktion $\gamma(p_1) \approx \Phi(-u)$ von $N(0, 1)$	

**Dominanz vs. komplementäre Polygenie (Fortsetzung 2):**

Für  $p_0 = \frac{1}{4} = 0,25$  und  $p_1 = \frac{7}{16} = 0,4375$  ergeben sich mit  $\alpha = 5\%$  und  $n = 40$ :

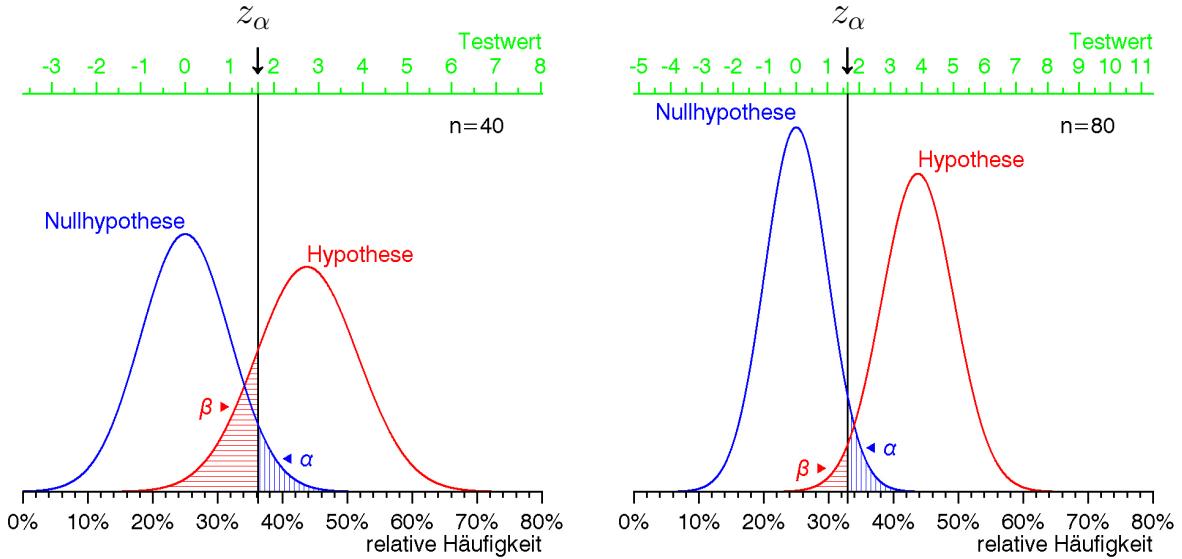
$$\begin{array}{lll} \Delta = 0,18754 & \sigma_0 = 0,4330 & \sigma_1 = 0,4961 \\ u = -0,9547 & \beta(p_1) = 16,99\% & \gamma(p_1) = 83,01\% \end{array} \quad \text{für } n = 40.$$

Das Fehlerrisiko  $\beta(p_1) \approx 17\%$  ist noch relativ hoch (es entspricht dem Werfen einer 6 beim homogenen Würfel) und folglich ist auch die Schärfe  $\gamma(p_1) \approx 83\%$  noch nicht ausreichend

groß. – Verdoppelt man jedoch den Umfang der  $F_2$ -Generation auf  $n = 80$ , so ändern sich die Werte wie folgt

$$u = -1,9449 \quad \beta(p_1) = 2,59\% \quad \gamma(p_1) = 97,41\% \quad \text{für } n = 80,$$

und jetzt ist  $\beta(p_1)$  sogar schon kleiner als  $\alpha = 5\%$  (vgl. Abb. 5).  $\square$



**Abb. 5:** Das Fehlerrisiko 2. Art  $\beta$  im Genetik-Beispiel (Dominanz vs. komplementäre Polygenie) für  $n = 40$  (links:  $\beta \approx 17\%$ ) und  $n = 80$  (rechts:  $\beta \approx 2,6\%$ ) bei  $\alpha = 5\%$ . Die untere  $x$ -Achse der relativen Häufigkeit ist in beiden Grafiken gleich und die Flächen unter den Kurven sind jeweils gleich groß (100%). Die oberen  $x$ -Achsen für den Testwert sind jedoch unterschiedlich skaliert, weil der Testwert sich mit  $\sqrt{n}$  ändert.

### 3.2.4 Versuchsplanung: erforderlicher Stichprobenumfang

Es sei wieder  $p_1$  ein für die konkrete Anwendung relevanter Wert aus der Alternative und  $\Delta = |p_1 - p_0| > 0$  die zugehörige Abweichung von der Nullhypothese. Dann fällt das Fehlerrisiko  $\beta(p_1)$  bzw. die Schärfe  $\gamma(p_1)$  wächst, wenn man den Stichprobenumfang  $n$  erhöht (vgl. Abb. 5).

Für  $n \rightarrow \infty$  strebt  $\beta(p_1)$  sogar gegen Null bzw.  $\gamma(p_1)$  strebt gegen 1. Folglich lässt sich durch einen geeignet großen Stichprobenumfang auch das Fehlerrisiko  $\beta(p_1)$  kontrollieren. Im Rahmen einer *Versuchsplanung* gibt man sich daher einen Wert  $\beta$  vor und bestimmt den *Mindestumfang*  $n$ , bei dem  $\beta(p_1)$  höchstens gleich der Vorgabe  $\beta$  ist. Die erforderlichen Rechenschritte sind in der *Planungs-Box 1* zusammengestellt. Typischerweise verwendet man für die Vorgabe  $\beta$  das *doppelte* Fehlerrisiko 1. Art, also  $\beta = 2\alpha$  (z.B.  $\beta = 10\%$  bei  $\alpha = 5\%$ ).

<b>Planungs-Box 1:</b> Für Tests über die Wahrscheinlichkeit $p$ einer $B(n, p)$ Verteilung		
einseitig oben	einseitig unten	zweiseitig
$H_0: p \leq p_0$ vs. $H: p > p_0$	$H_0: p \geq p_0$ vs. $H: p < p_0$	$H_0: p = p_0$ vs. $H: p \neq p_0$
<b>Vorgaben:</b>	Fehlerrisiko 1. Art Wert aus der Alternative $H$ : Fehlerrisiko 2. Art für $p = p_1$ : Schärfe	$\alpha$ $p_1$ $\beta$ bzw. $\gamma = 1 - \beta$
<i>Hilfsgrößen:</i>	$\Delta =  p_1 - p_0  > 0$ $\sigma_0 = \sqrt{p_0(1 - p_0)}$ , $\sigma_1 = \sqrt{p_1(1 - p_1)}$ $z_\alpha, z_\beta$ als oberes $\alpha$ -Quantil bzw. $\beta$ -Quantil von $N(0, 1)$	(Abweichung von $H_0$ )
einseitig oben	einseitig unten	zweiseitig
$z = z_\alpha$	$z = z_\alpha$	$z = z_{\frac{\alpha}{2}}$
<i>Erforderlicher Mindestumfang</i>	$n \approx \left[ \frac{z\sigma_0 + z_\beta\sigma_1}{\Delta} \right]^2$ (aufzurunden auf eine ganze Zahl)	

Der erforderliche Mindestumfang  $n$  in der Planungs-Box 1 hängt ab von:

- dem Fehlerrisiko 1. Art  $\alpha$ : je kleiner  $\alpha$ , desto größer ist  $z$  und somit  $n$ .
- dem Fehlerrisiko 2. Art  $\beta$ : je kleiner  $\beta$ , desto größer ist  $z_\beta$  und somit  $n$ .
- dem Unterschied  $\Delta$ : je kleiner  $\Delta$ , desto größer ist  $n$ ,
- den Werten  $p_0$  und  $p_1$ : je dichter  $p_0$  bzw.  $p_1$  bei  $\frac{1}{2}$  liegen, desto größer wird  $\sigma_0$  bzw.  $\sigma_1$  und somit  $n$ .

**Wahlumfrage (Fortsetzung 2):** Die Partei will das Fehlerrisiko 2. Art  $\beta(p_1)$  (für einen nicht genutzten günstigen Wahltermin) auf  $\beta = 10\%$  begrenzen, für den Fall, daß ihr wahrer Stimmanteil mit  $p_1 = 7\%$  deutlich über der 5%-Hürde liegt. Aus

$$\begin{array}{lllll} \alpha & = & 0,5\% & z_\alpha & = 2,576 \\ \Delta & = & 0,02 & \sigma_0 & = 0,2179 \\ & & & \sigma_1 & = 0,2551 \end{array} \quad \begin{array}{lllll} \beta & = & 10\% & z_\beta & = 1,282 \\ n & \approx & 1973,02 & & \end{array}$$

ergibt sich der aufgerundete Mindestumfang zu  $n = 1974$  Befragungen.  $\square$

### 3.3 Vergleich zweier Wahrscheinlichkeiten

Bisher haben wir die *unbekannte Wahrscheinlichkeit*  $p = P(A)$  eines interessierenden Ziel-Ereignisses  $A$  geschätzt und Tests für Hypothesen über  $p$  behandelt. Wir betrachten jetzt *zwei* verschiedene *Gruppen* (z.B. zwei Populationen oder zwei Behandlungsgruppen) und wollen die Wahrscheinlichkeit des Zielereignisses in beiden Gruppen miteinander *vergleichen*. Wir unterscheiden die beiden Gruppen durch die Nummern 1 bzw. 2 und bezeichnen die zugehörigen Ereignisse in der Gruppe mit  $A_1$  bzw.  $A_2$  sowie die unbekannten Wahrscheinlichkeiten mit  $p_1 = P(A_1)$  bzw.  $p_2 = P(A_2)$ .

**Klinischer Vergleich:** Bei einer spezifischen Erkrankung stehen zwei konkurrierende Therapien 1 bzw. 2 mit Wahrscheinlichkeiten  $p_1$  bzw.  $p_2$  für den Behandlungserfolg zur Verfügung. Von Interesse ist, ob eine der beiden Therapien wirksamer ist als die andere.

In einer konkreten Studie sollte überprüft werden, ob eine neue *Kombinationstherapie* zur Behandlung eines metastasierenden Brustkrebses gegenüber der bisherigen *Standard-Chemotherapie* überlegen ist oder nicht, wobei eine partielle Rückbildung des Tumors als *Behandlungserfolg* angesehen wird. Die Ergebnisse bei insgesamt 240 behandelten Patientinnen sind in Tabelle 1 zusammengestellt. Obwohl die beobachtete Erfolgsquote der neuen Therapie hier mit rund 42% deutlich besser ist als beim Standard (32,5%), ist noch zu klären, inwieweit dieser Unterschied auch *zufallsbedingt* sein können, falls beide Therapien gleichwertig sind.  $\square$

Therapie	Erfolg	kein Erfolg	Behandlungen	Erfolgsquote
1: neu	67	93	160	$\hat{p}_1 = 41,88\%$
2: Standard	26	54	80	$\hat{p}_2 = 32,50\%$
Total	93	147	240	$\hat{p} = 38,75\%$

Tab. 1: Ergebnisse einer Studie zum Vergleich zweier Chemotherapien beim metastasierenden Brustkrebs.

**Geschlechtsspezifischer Vergleich:** Gibt es einen geschlechtspezifischen Unterschied bei der Säuglingssterblichkeit?  $\square$

**Regionaler Vergleich:** Ist ein spezifisches Gesundheitsrisiko (z.B. für Lungenkrebs) in einer Industrieregion höher als in einer landwirtschaftlichen Region?  $\square$

### 3.3.1 Schätzung des Unterschieds mit Konfidenzgrenzen

Ausgangspunkt der statistischen Analyse der Wahrscheinlichkeit  $p_1$  bzw.  $p_2$  in den beiden Gruppen sind  $n_1$  unabhängige Wiederholungen in Gruppe 1 und *davon unabhängige* weitere  $n_2$  *unabhängige* Wiederholungen in Gruppe 2. Die resultierenden zufälligen Anzahlen  $X_1$  und  $X_2$  in den jeweiligen Gruppen sind dann *voneinander unabhängig* und binomial-verteilt:

$X_1$ ist $B(n_1, p_1)$ -verteilt,	$X_2$ ist $B(n_2, p_2)$ -verteilt, (1)
$X_1$ und $X_2$ sind <i>voneinander unabhängig</i> .	

Genau genommen handelt es sich bei (1) um eine *Voraussetzung*, die erfüllt sein muß, damit die folgenden Analysen korrekt sind. Diese Voraussetzung ist bei einem klinischen Vergleich z.B. beim folgenden Behandlungsschema *nicht* gegeben. Man behandelt  $n_1$  Patienten mit Therapie 1 und von *denjenigen Patienten, bei denen Therapie 1 versagt hat*, behandelt man *insgesamt*  $n_2$  Patienten mit der Therapie 2. Ein weiteres Gegen-Beispiel für (1) liegt vor, wenn  $X_1$  und  $X_2$  die Anzahl der Für-Stimmen für *zwei* Parteien (z.B. *SPD* und *CDU*) aus *derselben* Wahlumfrage sind.

Die beobachteten absoluten Häufigkeiten  $x_1$  und  $x_2$ , mit der das Ereignis eingetreten ist (also die Realisierungen von  $X_1$  und  $X_2$ ), stellt man mit den Häufigkeiten  $y_1 = n_1 - x_1$  und  $y_2 = n_2 - x_2$  der Nicht-Eintritte in einer sogenannten  $2 \times 2$ -Tafel übersichtlich zusammen (vgl. Tab. 1 und 2).

Gruppe	Eintritte	Nicht-Eintritte	Total
1	$x_1$	$y_1$	$n_1 = x_1 + y_1$
2	$x_2$	$y_2$	$n_2 = x_2 + y_2$
Total	$x = x_1 + x_2$	$y = y_1 + y_2$	$n = n_1 + n_2$

Tab. 2:  $2 \times 2$  Tafel für die beobachteten Häufigkeiten  
beim Vergleich zweier Gruppen.

Man berechnet nun zuerst die Schätzungen beider Wahrscheinlichkeiten  $p_1$  und  $p_2$

$$\hat{p}_1 = \hat{p}_1(x_1) = \frac{x_1}{n_1}, \quad \hat{p}_2 = \hat{p}_2(x_2) = \frac{x_2}{n_2} \quad (\text{Schätzungen}). \quad (2)$$

Für den interessierenden Unterschied beider Wahrscheinlichkeiten

$\Delta p = p_1 - p_2$	<i>(Differenz, Unterschied)</i> (3)
------------------------	-------------------------------------

ergibt sich dann als zugehörige Schätzung

$$\Delta\hat{p} = \hat{p}_1 - \hat{p}_2 \quad (\text{geschätzte Differenz, geschätzter Unterschied}). \quad (4)$$

Neben dieser *Schätzung* sind auch *Konfidenzgrenzen* für den Unterschied von Interesse. Hierzu brauchen wir Information über die *Streuung* der Schätzung  $\Delta\hat{p}$ . Die Varianz der zugehörigen *zufälligen* Differenz  $\hat{p}_1(X_1) - \hat{p}_2(X_2)$  ist (nach 2.6.4(5)) gleich der Summe beider Varianzen von  $\hat{p}_1(X_1)$  und  $\hat{p}_2(X_2)$ , also (nach 3.1.1(5))

$$\sigma_{\Delta}^2 := \text{Var}(\hat{p}_1(X_1) - \hat{p}_2(X_2)) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}. \quad (5)$$

Die Varianz  $\sigma_{\Delta}^2$  bzw. die Standardabweichung  $\sigma_{\Delta}$  lässt sich nicht berechnen, da die Wahrscheinlichkeiten  $p_1$  und  $p_2$  unbekannt sind. Aber man kann sie *schätzen*, indem man die Wahrscheinlichkeiten durch ihre Schätzungen ersetzt, und erhält dann:

$$\hat{\sigma}_{\Delta}^2 = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \quad (\text{geschätzte Varianz}). \quad (6)$$

Da die Differenz  $\hat{p}_1(X_1) - \hat{p}_2(X_2)$  approximativ  $N(\Delta\hat{p}, \sigma_{\Delta}^2)$ -verteilt ist, ergeben sich die *einseitigen* (asymptotischen) Konfidenzgrenzen für  $\Delta\hat{p}$  zur approximativen Sicherheit  $1-\alpha$  wie folgt:

$$\Delta\hat{p}_{un,\alpha} = \Delta\hat{p} - \hat{d}_{\alpha} \quad (\text{untere Konfidenzgrenze für } \Delta\hat{p}), \quad (7)$$

$$\Delta\hat{p}_{ob,\alpha} = \Delta\hat{p} + \hat{d}_{\alpha} \quad (\text{obere Konfidenzgrenze für } \Delta\hat{p}), \quad \text{mit} \quad (8)$$

$$\hat{d}_{\alpha} = z_{\alpha} \cdot \hat{\sigma}_{\Delta} \quad (\text{Bandbreite}), \quad (9)$$

wobei  $z_{\alpha}$  wieder das  $\alpha$ -Quantil von  $N(0,1)$  ist (vgl. Anhang T). Die beiden Konfidenzgrenzen liegen symmetrisch um die geschätzte Differenz  $\Delta\hat{p}$  im Abstand  $\hat{d}_{\alpha}$  und sind analog zu den *groben* Grenzen in 3.1.3 b konstruiert.

Die Sicherheit dieser Grenzen ist nur *approximativ* gleich der Vorgabe  $1-\alpha$ , wobei die Approximation beliebig genau wird, wenn *beide* Umfänge anwachsen:  $n_1 \rightarrow \infty$  und  $n_2 \rightarrow \infty$ . Deshalb sollte man diesen Grenzen nur vertrauen, wenn *alle vier* beobachteten Anzahlen  $x_1, y_1, x_2, y_2$  in Tab. 2 *nicht zu klein* sind, d.h. *mindestens* 5 (besser: 10) betragen.

Das zugehörige *zweiseitige* Konfidenz-Intervall zur approximativen Sicherheit  $1-\alpha$  ergibt sich wieder aus den beiden einseitigen groben Grenzen mit  $\frac{\alpha}{2}$  statt  $\alpha$ :

$$(\Delta\hat{p}_{un,\frac{\alpha}{2}}, \Delta\hat{p}_{ob,\frac{\alpha}{2}}) = \{\Delta\hat{p} \mid \Delta\hat{p}_{un,\frac{\alpha}{2}} < \Delta\hat{p} < \Delta\hat{p}_{ob,\frac{\alpha}{2}}\} \quad (10)$$

(zweiseitiges Konfidenzintervall für  $\Delta\hat{p}$ ).

**Klinischer Vergleich (Fortsetzung 1):** Für die Studie aus Tab. 1 ist der beobachtete Unterschied  $\hat{\Delta p} \approx 9,38\%$  relativ hoch. Die untere Konfidenzgrenze zur (einseitigen) Sicherheit 95% ergibt sich aus

$$\alpha = 5\% \quad z_{5\%} = 1,645 \quad \hat{\sigma}_\Delta = 0,0653 \quad \hat{d}_{5\%} = 10,74\%$$

zu  $\hat{\Delta p}_{un,5\%} = -1,4\%$ . Da die untere Grenze *negativ* ist (d.h. unterhalb von 0 liegt, was ja der Differenz  $\Delta p$  bei *Gleichheit*  $p_1 = p_2$  entspricht), ist die Überlegenheit der neuen Therapie durch diese Studie *nicht* (zu 95%) abgesichert.

Andererseits liefert die *obere* Grenze von  $\hat{\Delta p}_{ob,5\%} = 20,1\%$  einen Hinweis, um wieviel die Erfolgs-Wahrscheinlichkeit  $p_1 = p_2 + \Delta p$  den Wert  $p_2$  der Standard-Therapie übersteigen kann. Man beachte, daß das *zweiseitige* Intervall von  $\hat{\Delta p}_{un,5\%}$  bis  $\hat{\Delta p}_{ob,5\%}$  nur die *zweiseitige* Sicherheit von 90% (also *nicht* von 95%) hat.  $\square$

### 3.3.2 Testen von Hypothesen

Neben dem *Schätzen* des Unterschieds  $\Delta p = p_1 - p_2$  (mit Konfidenzgrenzen) ist es in vielen Anwendungen erforderlich, sich aufgrund der beobachteten Daten zwischen konkurrierenden Hypothesen über beide Wahrscheinlichkeiten  $p_1$  und  $p_2$  bzw. über den Unterschied  $\Delta p$  zu *entscheiden*.

#### Einseitiger Test

Wir betrachten zuerst das *einseitige* Testproblem mit den Hypothesen

$$\begin{aligned} \text{Nullhypothese } H_0 : p_1 &\leq p_2 \quad \text{d.h.} \quad \Delta p \leq 0 \quad (\text{kein positiver Unterschied}) \text{ vs. } (1) \\ \text{Alternative } H : p_1 &> p_2 \quad \text{d.h.} \quad \Delta p > 0 \quad (\text{positiver Unterschied}). \end{aligned}$$

Man beachte, daß man das duale Testproblem mit  $p_2 \leq p_1$  als Nullhypothese und  $p_2 > p_1$  als Alternative hier nicht separat betrachten muß, weil es nach *Vertauschen* beider Gruppen auch von der Form (1) ist.

Wir wollen jetzt (wie in 3.2) einen Testwert definieren, mit dessen Hilfe die Entscheidung für eine der beiden Hypothesen herbeigeführt werden soll. Als Testwert wollen wir eine geeignete „Standardisierung“ des geschätzten Unterschieds  $\hat{\Delta p}$  verwenden. Hierzu betrachten wir die absolute Häufigkeit  $x = x_1 + x_2$  in beiden Gruppen *zusammen* und die zugehörige *gemeinsame* relative Häufigkeit (vgl. Tab. 2)

$$\hat{p} = \frac{x}{n} \quad \text{mit} \quad n = n_1 + n_2. \quad (2)$$

Ersetzt man in der Varianz  $\sigma_\Delta^2$  jetzt  $p_1$  und  $p_2$  durch die Schätzung  $\hat{p}$ , so ergibt sich als *geschätzte* Varianz von  $\hat{\Delta p}$  unter dem Spezialfall  $p_1 = p_2$  der Nullhypothese

$$\hat{\sigma}_0^2 := \left[ \frac{1}{n_1} + \frac{1}{n_2} \right] \hat{p}(1 - \hat{p}) = \frac{n}{n_1 n_2} \hat{p}(1 - \hat{p}) = \frac{xy}{n_1 n_2 n}. \quad (3)$$

Als Testwert verwenden wir nun die folgende *Standardisierung* von  $\Delta\hat{p} = \hat{p}_1 - \hat{p}_2$

$$t(x_1x_2) = \frac{\hat{p}_1(x_1) - \hat{p}_2(x_2)}{\hat{\sigma}_0} = (x_1y_2 - x_2y_1) \cdot \sqrt{\frac{n}{x_1y_1n_1n_2}} \quad (\text{Testwert}). \quad (4)$$

Die zugehörige *zufällige* Testvariable  $t(X_1, X_2)$  ist unter dem Spezialfall  $p_1 = p_2$  der Nullhypothese approximativ  $N(0, 1)$ -verteilt. Da *große* Testwerte für die Alternative sprechen, ergibt sich (analog **3.2.1**) folgender Test zum Niveau  $\alpha$  (vgl. **3.2 Abb. 4**):

$$\begin{aligned} \text{Ablehnung von } H_0: p_1 \leq p_2 &\Leftrightarrow t(x_1, x_2) \geq z_\alpha \\ &\Leftrightarrow \Phi(-t(x_1, x_2)) \leq \alpha. \end{aligned} \quad (5)$$

Hierbei ist die *Signifikanz der Beobachtung* oder der *P-Wert*

$$\Phi(-t(x_1, x_2)) \approx P\{t(X_1, X_2) \geq t(x_1, x_2) \mid p_1 = p_2\} \quad (6)$$

eine Approximation der Wahrscheinlichkeit unter dem Spezialfall  $p_1 = p_2$  der Nullhypothese, einen *mindestens* so großen Testwert wie den konkret beobachteten zu erhalten.

Man könnte auch – inspiriert durch **3.2.1** – auf die Idee kommen, sich für die Alternative  $H: \Delta p > 0$  zu entscheiden, wenn die *untere* Konfidenzgrenze für  $\Delta p$  nicht negativ ist, d.h.  $0 \leq \Delta\hat{p}_{un,\alpha}$  gilt, weil ja  $\Delta\hat{p}_{un,\alpha} < \Delta p$  mit der (approximativen) Sicherheit von  $1 - \alpha$  gilt. Dieses Entscheidungskriterium entspricht allerdings *nicht genau* dem Test (5), weil die Schätzungen  $\hat{\sigma}_0$  und  $\hat{\sigma}_\Delta$  im allgemeinen *nicht genau* übereinstimmen, obwohl die Abweichungen typischerweise nur gering sind.

**Klinischer Vergleich (Fortsetzung 2):** Für die Studie aus *Tab. 1* soll jetzt der einseitige Test mit den Hypothesen

$$\begin{array}{ll} \text{Nullhypothese} & H_0: p_1 \leq p_2 \quad (\text{neue Therapie nicht besser als Standard}) \quad \text{vs.} \\ \text{Alternative} & H: p_1 > p_2 \quad (\text{neue Therapie ist besser als Standard}) \end{array}$$

zum Niveau  $\alpha = 5\%$  überprüft werden. Aus

$$\Delta\hat{p} \approx 9,375\% \quad \hat{\sigma}_0 = 0,0667 \quad t = 1,405$$

ergibt sich, daß die Nullhypothese *nicht* abgelehnt wird, weil

$$\begin{array}{llll} \text{Testwert} & t & = & 1,405 & < 1,645\% = z_{5\%} \\ \text{P-Wert} & \Phi(-t) & = & 8,00\% & > 5\% = \alpha. \end{array}$$

Der P-Wert bedeutet hier, daß selbst bei *gleichen* Erfolgs-Wahrscheinlichkeiten der Therapien (also  $p_1 = p_2$ ) ein Testwert von *mindestens* 1,405 noch mit einer Wahrscheinlichkeit von 8% auftritt. Da hier die Nullhypothese *nicht* abgelehnt wird, kann ein Fehler 2. Art vorliegen, und wir werden das zugehörige Fehlerrisiko in **3.3.4** weiter analysieren.  $\square$

### Zweiseitiger Test

Zum Abschluß betrachten wir das *zweiseitige* Testproblem mit den Hypothesen

$$\begin{aligned} \text{Nullhypothese } H_0 : p_1 = p_2 &\quad \text{d.h.} \quad \Delta p = 0 \quad (\text{kein Unterschied}) \\ \text{Alternative } H : p_1 \neq p_2 &\quad \text{d.h.} \quad \Delta p \neq 0 \quad (\text{Unterschied vorhanden}). \end{aligned} \quad \text{vs.} \quad (7)$$

Da jetzt sowohl *hohe positive* als auch *hohe negative* Testwerte für die Alternative sprechen, ergibt sich (analog **3.2.2**) folgender *zweiseitiger* Test zum Niveau  $\alpha$  (vgl. **3.2 Abb. 4**):

$$\text{Ablehnung von } H_0: p_1 = p_2 \Leftrightarrow |t(x_1, x_2)| \geq z_{\alpha/2}. \quad (8)$$

Dieser zweiseitige Test wird auch als *Chiquaret-Test* bezeichnet, weil die *quadrierte* Teststatistik  $t^2(X_1, X_2)$  unter der Nullhypothese  $H_0 : p_1 = p_2$  approximativ eine sogenannte *Chiquaret-Verteilung*  $\chi_1^2$  mit Freiheitsgrad 1 besitzt. Das  $\alpha$ -Quantil  $\chi_{1,\alpha}^2$  dieser Verteilung (vgl. Anhang **T**) stimmt mit dem Quadrat  $z_{\alpha/2}^2$  überein, und der Test läßt sich äquivalent formulieren durch:

$$\text{Ablehnung von } H_0: p_1 = p_2 \Leftrightarrow t^2(x_1, x_2) \geq \chi_{1,\alpha}^2. \quad (9)$$

**Beispiel:** vgl. **3.3.6.**

### 3.3.3 Test- und Konfidenz-Box 2:

Vergleich der Eintritts-Wahrscheinlichkeiten  $p_1$  und  $p_2$  bei *zwei unabhängigen*  $B(n_1, p_1)$ - und  $B(n_2, p_2)$ -verteilten Zufallsvariablen  $X_1$  und  $X_2$ .

einseitiger Test	zweiseitiger Test
$H_0: p_1 \leq p_2$ vs. $H: p_1 > p_2$	$H_0: p_1 = p_2$ vs. $H: p_1 \neq p_2$

Fehlerrisiko 1. Art (Testniveau):  $\alpha$

Stichprobenumfänge:  $n_1$  (Gruppe 1),  $n_2$  (Gruppe 2)

Beobachtete Anzahl der Eintritte:  $x_1$  (Gruppe 1),  $x_2$  (Gruppe 2)

Anordnung in  $2 \times 2$ -Tafel:

Gruppe	Eintritte	Nicht-Eintritte	Total	Schätzung
1	$x_1$	$y_1$	$n_1 = x_1 + y_1$	$\hat{p}_1 = \frac{1}{n_1} x_1$
2	$x_2$	$y_2$	$n_2 = x_2 + y_2$	$\hat{p}_2 = \frac{1}{n_2} x_2$
Total	$x = x_1 + x_2$	$y = y_1 + y_2$	$n = n_1 + n_2$	$\hat{p} = \frac{1}{n} x$

**Asymptotischer Test:** Fehlerrisiko 1. Art  $\approx \alpha$

Anwendbar: wenn  $x_1, y_1, x_2, y_2$  nicht zu klein, Faustregel: mindestens 5 (besser: 10)

$$\text{Testwert } t = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_0} \quad \text{mit} \quad \hat{\sigma}_0 = \sqrt{\frac{n\hat{p}(1-\hat{p})}{n_1 n_2}}$$

Entscheidung: Ablehnung von  $H_0$  (Annahme von  $H$ ), falls gilt

einseitiger Test	zweiseitiger Test
Testwert $t \geq z_\alpha$ bzw.	Testwert $ t  \geq z_{\frac{\alpha}{2}}$ bzw.
$P$ -Wert $\Phi(-t) \leq \alpha$	$P$ -Wert $2\Phi(- t ) \leq \alpha$

**Einseitige Konfidenzgrenzen für  $\Delta p = p_1 - p_2$ :** Sicherheit  $\approx 1 - \alpha$

Anwendbar: wenn  $x_1, y_1, x_2, y_2$  nicht zu klein, Faustregel: mindestens 5 (besser: 10)

einseitige untere Grenze	einseitige obere Grenze
$\Delta \hat{p}_{un,\alpha} = \Delta \hat{p} - \hat{d}_\alpha$	$\Delta \hat{p}_{ob,\alpha} = \Delta \hat{p} + \hat{d}_\alpha$

$$\Delta \hat{p} = \hat{p}_1 - \hat{p}_2, \quad \hat{d}_\alpha = z_\alpha \cdot \hat{\sigma}_\Delta, \quad \hat{\sigma}_\Delta = \sqrt{\frac{1}{n_1} \hat{p}_1 (1 - \hat{p}_1) + \frac{1}{n_2} \hat{p}_2 (1 - \hat{p}_2)}$$

**Zweiseitiges Konfidenzintervall für  $\Delta p$ :**  $(\Delta \hat{p}_{un,\frac{\alpha}{2}}, \Delta \hat{p}_{ob,\frac{\alpha}{2}})$

**Erläuterung zur Test- und Konfidenz-Box 2:** Das  $\alpha$ -Quantil bzw. die Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung ist im Anhang T tabelliert oder kann per Computer berechnet werden. Bei *nicht zu großem* Unterschied  $\Delta\hat{p} = \hat{p}_1 - \hat{p}_2$  ist  $\hat{\sigma}_\Delta \approx \hat{\sigma}_0$ .

### 3.3.4 Fehlerrisiko 2. Art und Testschärfe (Power)

Wenn man die Nullhypothese *nicht* abgelehnt hat, so ist ein Fehler 2. Art möglich, dessen Risiko man (zumindest für relevante Werte aus der Alternative) berechnen sollte und zwar eigentlich schon bei der Planung des Versuchs (vgl. auch 3.3.5). Beim *einseitigen* Testproblem

$$H_0: p_1 \leq p_2 \quad \text{vs.} \quad H: p_1 > p_2 \quad (1)$$

hängen das Fehlerrisiko 2. Art  $\beta = \beta(p_1, p_2)$  und die dazu komplementäre Testschärfe  $\gamma = \gamma(p_1, p_2)$  von den konkreten Werten  $p_1 > p_2$  aus der Alternative ab. Je *kleiner* der Unterschied  $\Delta p = p_1 - p_2$  ist, desto *größer* ist das Fehlerrisiko  $\beta(p_1, p_2)$ . Die Berechnung von  $\beta(p_1, p_2)$  basiert auf der approximativen Normalverteilung der Test-Variablen  $t(X_1, X_2)$  und die erforderlichen Schritte sind in der *Power-Box 2* zusammengestellt.

Für das *zweiseitige* Testproblem

$$H_0: p_1 = p_2 \quad \text{vs.} \quad H: p_1 \neq p_2 \quad (2)$$

wird das Fehlerrisiko  $\beta = \beta(p_1, p_2)$  analog berechnet (vgl. *Power-Box 2*). Es wird wieder größer, wenn der *absolute* Unterschied  $\Delta = |p_1 - p_2|$  kleiner wird.

**Klinischer Vergleich (Fortsetzung 3):** Da für die Studie aus Tab. 1 die Nullhypothese *nicht* abgelehnt wird, kann ein Fehler 2. Art vorliegen, und wir wollen das zugehörige Fehlerrisiko für zwei relevante Situationen berechnen. Hierbei gehen wir davon aus, daß die Standard-Therapie in rund  $p_2 = 30\%$  erfolgreich ist (dies kann z.B. aus langjähriger Praxis bekannt sein und deckt sich auch ungefähr mit der beobachteten Quote  $\hat{p}_2$ ). Falls die neue Therapie mit  $p_1 = 40\%$  um *ein Drittel* erfolgreicher als der Standard ist, so ergibt sich das zugehörige Fehlerrisiko 2. Art wie folgt

$$\begin{array}{llll} \Delta=0,10 & p=0,3667 & \sigma_\Delta=0,0660 & \sigma_\Delta=0,0642, \\ u=0,1329 & \beta(30\%, 40\%)=55,3\% & & \gamma(30\%, 40\%)=44,7\%. \end{array}$$

Das Fehlerrisiko ist mit  $\beta = 55,3\%$  inakzeptabel hoch, d.h. der Test entdeckt diesen Unterschied nur mit einer Wahrscheinlichkeit von  $\gamma = 44,3\%$ . Selbst wenn die neue Therapie um *die Hälfte* erfolgreicher ist als der Standard, d.h.  $p_1 = 45\%$ , so ergeben sich mit

$$\beta(30\%, 45\%) = 27\% \quad \gamma(30\%, 45\%) = 73\%$$

noch keine zufriedenstellenden Werte. Die Ursache hierfür sind die zu geringen Stichprobenumfänge  $n_1 = 160$  und  $n_2 = 80$ .  $\square$

**Power-Box 2:** Für Tests über die Wahrscheinlichkeiten  $p_1$  und  $p_2$  zweier unabhängiger  $B(n_1, p_1)$ - und  $B(n_2, p_2)$ -Verteilungen.

einseitiger Test	zweiseitiger Test
$H_0: p_1 \leq p_2$ vs $H: p_1 > p_2$	$H_0: p_1 = p_2$ vs. $H: p_1 \neq p_2$
Fehlerrisiko 1. Art (Testniveau):	$\alpha$
Stichprobenumfänge:	$n_1, n_2$
Wertepaar aus der Alternative $H$ :	$p_1 > p_2$
einseitiger Test	zweiseitiger Test
$z = z_\alpha$	$z = z_{\frac{\alpha}{2}}$
<i>Hilfsgrößen:</i>	
$\Delta =  p_1 - p_2  > 0$	(Abweichung von $H_0$ )
$p = \frac{n_1 p_1 + n_2 p_2}{n}$	(Eintritts-Wahrscheinlichkeit in der Gesamt-Stichprobe)
$\sigma_0 = \sqrt{\frac{np(1-p)}{n_1 n_2}}$	(Standardabweichung von $\hat{p}_1 - \hat{p}_2$ unter $H_0: p_1 = p_2$ )
$\sigma_\Delta = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	(Standardabweichung von $\hat{p}_1 - \hat{p}_2$ )
$u = \frac{z\sigma_0 - \Delta}{\sigma_\Delta}$	
<i>Fehlerrisiko 2. Art:</i>	
$\beta(p_1, p_2) \approx \Phi(u)$	mit $\Phi = \text{Verteilungsfunktion}$
Testschärfe (Power):	$\gamma(p_1, p_2) \approx \Phi(-u)$ von $N(0, 1)$

**Eräuterung zur Power-Box 2:** Das  $\alpha$ -Quantil  $z_\alpha$  bzw. die Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung ist im Anhang T tabelliert oder kann per Computer berechnet werden. Mit wachsenden Umfängen  $n_1, n_2$  oder wachsender Abweichung  $\Delta$  fällt  $u$  und somit auch das Fehlerrisiko  $\beta$ . Bei nicht zu großem Unterschied  $\Delta$  ist  $\sigma_\Delta \approx \sigma_0$ .

### 3.3.5 Versuchsplanung: erforderlicher Stichprobenumfang

Es sei wieder  $p_1 > p_2$  ein für die konkrete Anwendung relevantes Wertepaar aus der Alternative und  $\Delta = |p_1 - p_2| > 0$  die zugehörige Abweichung von der Nullhypothese. Dann fällt das Fehlerrisiko  $\beta(p_1, p_2)$  bzw. die Schärfe  $\gamma(p_1, p_2)$  wächst, wenn man die Stich-

probenumfänge  $n_1$  und  $n_2$  erhöht. Im Rahmen einer *Versuchsplanung* kann man wieder für einen vorgegebenen Wert  $\beta$  die erforderlichen Mindestumfänge  $n_1$  und  $n_2$  bestimmen, bei denen  $\beta(p_1, p_2)$  höchstens gleich der Vorgabe  $\beta$  ist. Hierbei ist es (aus statistischer Sicht) optimal, *gleiche Stichprobenumfänge*  $n_1 = n_2$  zu verwenden. Man spricht dann von einem *balancierten Design*. Die erforderliche Rechenschritte sind in der *Planungs-Box 2* zusammengestellt. Typischerweise verwendet man für die Vorgabe  $\beta$  das *doppelte Fehlerrisiko 1. Art*, also  $\beta = 2\alpha$  (z.B.  $\beta = 10\%$  bei  $\alpha = 5\%$ ).

**Planungs-Box 2:**

Für Tests über die Wahrscheinlichkeiten  $p_1$  und  $p_2$  zweier unabhängiger  $B(n_1, p_1)$ - und  $B(n_2, p_2)$ -Verteilungen mit gleichen Umfängen  $n_1 = n_2$

einseitiger Test	zweiseitiger Test
$H_0: p_1 \leq p_2$ vs. $H: p_1 > p_2$	$H_0: p_1 = p_2$ vs. $H: p_1 \neq p_2$

**Vorgaben:** Fehlerrisiko 1. Art (Testniveau):  $\alpha$

Wertepaar aus der Alternative  $H$ :  $p_1 > p_2$

Fehlerrisiko 2. Art für  $p_1 > p_2$ :  $\beta$  bzw.

Schärfe für  $p_1 > p_2$ :  $\gamma = 1 - \beta$

*Hilfsgrößen:*  $\Delta = |p_1 - p_2| > 0$  (Abweichung von  $H_0$ )

$\bar{p} = \frac{1}{2}(p_1 + p_2)$  (Mittelwert der Wahrscheinlichkeiten)

einseitiger Test	zweiseitiger Test
$z = z_\alpha$	$z = z_{\frac{\alpha}{2}}$

**Mindestumfang**

pro Gruppe:  $n_1 = n_2 \approx 2\bar{p}(1 - \bar{p}) \left[ \frac{z + z_\beta}{\Delta} \right]^2$  (aufzurunden auf  
insgesamt:  $n = n_1 + n_2 = 2n_1$  eine ganze Zahl)

**Klinischer Vergleich (Fortsetzung 4):** Die Studie aus Tab. 1 hat für den interessierenden Fall mit  $p_1 = 45\% > p_2 = 30\%$  noch das relative hohe Fehlerrisiko 2. Art von  $\beta = 27\%$  (vgl. 3.3.4). Wir wollen jetzt die Mindestumfänge (beim balancierten Design) bestimmen, die ein Fehlerrisiko 2. Art von nur  $\beta = 10\%$  garantieren. Aus

$$\begin{array}{llll} \alpha = 5\% & z_\alpha = 1,645 & \beta = 10\% & z_\beta = 1,282 \\ \Delta = 0,15 & \bar{p} = 0,375 & n_1 = n_2 \approx 178,4 & \end{array}$$

ergibt sich der *aufgerundete* Mindestumfang pro Gruppe zu  $n_1 = n_2 = 179$  Anwendungen, also insgesamt  $n = 358$  Patientinnen.  $\square$

### 3.3.6 Anwendungen

#### *Säuglingssterblichkeit nach Geschlecht*

Bezeichnet  $p_1$  bzw.  $p_2$  die Rate der Säuglingssterblichkeit (worunter wir hier Totgeburtten und Todesfälle in den ersten 28 Lebenstagen verstehen wollen) bei den männlichen bzw. weiblichen Geborenen, so wollen wir einen *zweiseitigen* Test zum Niveau  $\alpha = 5\%$  durchführen mit den Hypothesen

$$\begin{aligned} H_0: p_1 &= p_2 && (\text{kein geschlechtspezifischer Unterschied der Sterblichkeit}), \\ H_H: p_1 &\neq p_2 && (\text{geschlechtspezifischer Unterschied der Sterblichkeit}). \end{aligned}$$

In der (bereits in **1.7.3** erwähnten) Arbeit von Karn und Penrose (*Birth weight and gestation time in relation to maternal age, parity and infant survival*, Annals of Eugenetics 16, 1951-52, 147-164) werden – basierend auf Krankenhausunterlagen der Jahre 1935-1946 in England – die Daten aus *Tab. 3* zur Säuglingssterblichkeit angegeben.

Geschlecht	Totgeborene und Gestorbene	Überlebende (28 Tage)	Geborene	beobachtete Sterberate
männlich: 1	340	6 697	7 037	$\hat{p}_1 = 4,83\%$
weiblich: 2	274	6 419	6 693	$\hat{p}_2 = 4,09\%$
insgesamt	614	13 116	13 730	$\hat{p} = 4,47\%$

**Tab. 3:** *Säuglingssterblichkeit (Totgeburtten und Todesfälle in den ersten 28 Lebenstagen) nach Geschlecht, England 1935-1946* (Quelle: Karn und Penrose, 1951-52).

Die Nullhypothese wird *abgelehnt*, weil:

$$\begin{aligned} \text{Testwert} \quad |t| &= 2,091 \quad \geq z_{2,5\%} = 1,96 && \text{bzw.} \\ \text{P-Wert} \quad \Phi(-t) &= 3,66\% \quad \leq \quad \alpha = 5\%. \end{aligned}$$

Die Hypothese  $H$  einer geschlechtsspezifischen Säuglingssterblichkeit ist also mit der Irrtumswahrscheinlichkeit  $\alpha = 5\%$  bzw. mit der Sicherheit von  $1 - \alpha = 95\%$  *abgesichert*. Und zwar ist die männliche Sterblichkeit *höher* als die weibliche. Daß hier trotz des relativ *geringen* Unterschieds  $\Delta\hat{p} = \hat{p}_1 - \hat{p}_2 = 0,74\%$  der beobachteten Sterberisiken der Test die Nullhypothese ablehnt, liegt an den *großen* Gruppenumfängen (Anzahl der Geborenen).

Das *zweiseitige 95%-Konfidenzintervall* für den Unterschied  $\Delta p$  ist nach Test- und Konflikt-Box 2 gegeben durch:

$$0,74\% \pm 0,69\% \quad \text{bzw. von } 0,05\% \text{ bis } 1,43\%$$

□

### **Gefährdung von Schwangerschaften nach Tschernobyl**

Thieme und Lack (*Der Frauenarzt* 6/1987) haben Schwangerschaften in verschiedenen Regionen betrachtet und Totgeburten sowie ausgewählte Mißbildungen als „Schadensfälle“ eingestuft. Die Belastung wurde durch die Bodenkontamination mit Cäsium-137 definiert. Im Vergleich der am geringsten belasteten Region in Niedersachsen („nicht exponiert“) mit der am höchsten belasteten Region in Bayern („exponiert“) ergab sich folgende Tabelle:

Exposition	Schaden	kein Schaden	Umfang	Schadensquote
Ja (1)	15	1 247	1 262	$\hat{p}_1 = 1,189\%$
Nein (2)	15	1 287	1 287	$\hat{p}_2 = 1,166\%$
Gesamt	30	2 519	2 549	$\hat{p} = 1,177\%$

Es wird ein *einseitiger* Test zum Niveau  $\alpha = 5\%$  durchgeführt mit den Hypothesen

$$H_0 : p_1 \leq p_2 \quad (\text{Schadensrisiko bei Exposition nicht höher als ohne Exposition}),$$

$$H: p_1 > p_2 \quad (\text{Schadensrisiko bei Exposition höher als ohne Exposition}).$$

Die Nullhypothese wird *nicht abgelehnt*, weil:

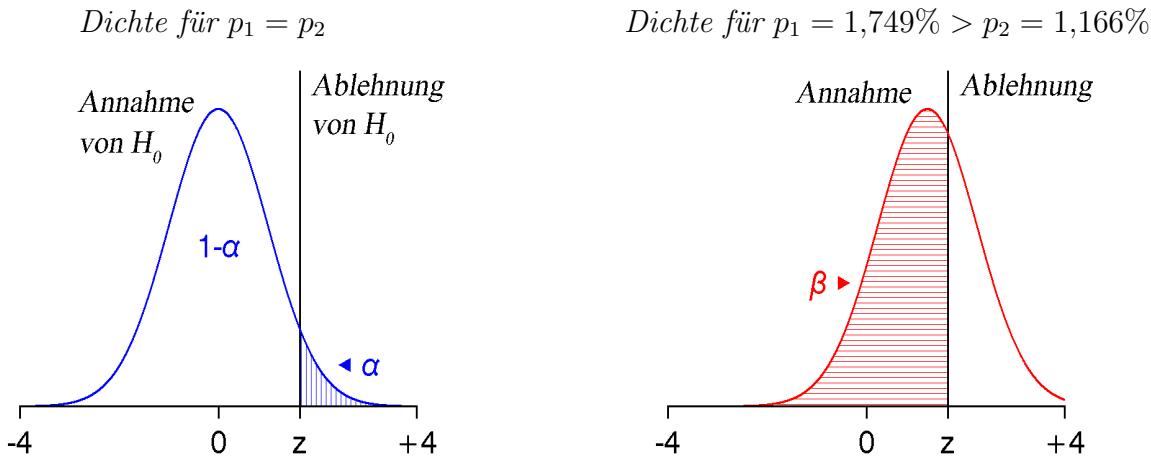
$$\begin{aligned} \text{Testwert:} \quad t &= 0,054 &< z_\alpha = 1,645, \\ \text{Signifikanzniveau:} \quad P &= 47,8\% &> \alpha = 5\%. \end{aligned}$$

Ein höheres Schadensrisiko unter Exposition (d.h.  $H$  gilt) ist mit diesen Daten nicht nachweisbar. Es kann aber ein Fehler 2. Art („trügerische Sicherheit“) vorliegen, und daher ist die Nullhypothese durch den Test *nicht quantifizierbar abgesichert*. Das *zweiseitige 90%-Konfidenzintervall* für den Unterschied  $\Delta p = p_1 - p_2$  reicht allerdings von  $-0,68\%$  bis  $0,73\%$  und deshalb sind auch positive Werte für  $\Delta p$ , z.B.  $\Delta p = 0,5\%$  bzw.  $p_1 = p_2 + 0,5\%$ , noch „kompatibel“ mit den beobachteten Daten.

*Analyse des Fehlerrisiko 2. Art*  $\beta$  und der Schärfe  $\gamma = 1 - \beta$  für hypothetisches  $p_2 = \hat{p}_2 = 1,166\%$  (beobachtete Schadensquote ohne Exposition) und *vorgegebene* Werte des *relativen Risikos*  $RR = p_1/p_2$  der Exposition gegenüber keiner Exposition:

$RR$	$p_1 = RR \cdot p_2$	Risiko $\beta$	Schärfe $\gamma = 1 - \beta$
1,5	1,749%	66,2%	33,8%
2,0	2,332%	27,5%	72,5%
2,5	2,915%	6,9%	93,1%
3,0	3,498%	1,2%	98,8%

Beim relativen Risiko von 1,5 erhält man das extrem große Risiko  $\beta$  von 66,2% (vgl. Abb. 1). Ein akzeptables Fehlerrisiko  $\beta$  von unter 10% ergibt sich hier erst ab einem relativen Risiko von rund 2,5 oder mehr.



**Abb. 1:** Dichte der (approximativ) normalverteilten Teststatisik  $t(X_1, X_2)$  mit Fehlerrisiken für den **einseitigen Test** (vgl. Beispiel: Gefährdung von Schwangerschaften nach Tschernobyl)

*Versuchsplanung:* Erforderliche Mindestumfänge  $n_1 = n_2$ , damit der Test zum Niveau  $\alpha = 5\%$  ein Fehlerrisiko 2. Art von (höchstens)  $\beta = 10\%$  bzw. eine Schärfe von (mindestens)  $\gamma = 90\%$  hat für hypothetisches  $p_2 = \hat{p}_2 = 1,166\%$  (beobachtete Schadensquote ohne Exposition) und *vorgegebene* Werte des relativen Risikos  $RR$ :

$RR$	$p_1 = RR \cdot p_2$	Umfänge	Gesamtumfang
		$n_1 = n_2$	$n = n_1 + n_2$
2,0	2,332%	2 164	4 332
1,5	1,749%	7 239	14 478
1,2	1,399%	39 895	79 916
1,1	1,283%	151 394	302 778

### 3.4 Konfidenzgrenzen und Tests für einen Erwartungswert

Nachdem wir (in **3.1-2**) Konfidenzgrenzen und Tests für die Wahrscheinlichkeit eines Ziel-Ereignisses behandelt haben, wollen wir diese Konzepte jetzt auf den Erwartungswert  $\mu$  einer interessierenden Zufallsvariablen  $X$  ausdehnen. Anwendungsbeispiele hierfür sind:

- Der Schadstoffgehalt  $X$  (z.B. in einem Nahrungsmittel oder in der Luft) kann als Zufallsvariable aufgefaßt werden mit  $\mu = E(X)$  als „mittlerem“ Wert.
- Die Lebensdauer  $X$  eines technischen Produkts mit  $\mu = E(X)$  als durchschnittlicher Lebensdauer.
- Der Wirkstoffgehalt  $X$  eines Medikaments mit  $\mu = E(X)$  als „mittlerem“ Gehalt.

#### 3.4.1 Konfidenzgrenzen für den Erwartungswert

Der Ausgangspunkt zur Konstruktion von Konfidenzgrenzen für  $\mu$  sind  $n$  unabhängige Wiederholungen  $X_1, X_2, \dots, X_n$  der Zufallsvariablen  $X$ . Für die zugehörigen Realisierungen (beobachtete Ergebnisse)  $x_1, x_2, \dots, x_n$  haben wir in **2.6.5** bereits die Schätzungen für den Erwartungswert  $\mu = E(X)$  und für die Standardabweichung  $\sigma = SD(X)$  von  $X$  kennengelernt (vgl. auch Rechen-Box 1):

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_i x_i \quad (\text{Schätzung von } \mu) \quad (1)$$

$$\hat{\sigma} = s_x = \sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} S_{xx}} \quad (\text{Schätzung von } \sigma) \quad (2)$$

Wir wollen jetzt die untere und obere Konfidenzgrenze für  $\mu$  zur Sicherheit  $1 - \alpha$  symmetrisch zur Schätzung  $\bar{x}$  wie folgt wählen

$$\hat{\mu}_{un,\alpha} = \bar{x} - \hat{d}_\alpha, \quad \hat{\mu}_{ob,\alpha} = \bar{x} + \hat{d}_\alpha, \quad (3)$$

wobei die Bandbreite

$$\hat{d}_\alpha = c_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{n}} \quad (4)$$

als ein Vielfaches der Standardabweichung  $SD(\bar{X})$  des Mittelwerts  $\bar{X}$  (vgl. **2.6.5**) ange setzt wird. Der Faktor  $c_\alpha$  ist hierbei noch so zu bestimmen, daß die gewünschte Sicherheit von  $1 - \alpha$  erreicht wird. Hierzu betrachten wir zunächst die untere Grenze und untersuchen, wann diese Grenze unterhalb von  $\mu$  liegt:

$$\bar{x} - \hat{d}_\alpha < \mu \quad \Leftrightarrow \quad t := \frac{(\bar{x} - \mu)\sqrt{n}}{\hat{\sigma}} < c_\alpha. \quad (5)$$

Der Wert  $t$  ist die Realisierung der zugehörigen Zufallsvariablen (vgl. **2.6.5**)

$$t(X) := \frac{(\bar{X} - \mu)\sqrt{n}}{S_X} \quad \text{mit} \quad S_X = \sqrt{\frac{1}{n-1} \sum_i (X_i - \bar{X})^2}, \quad (6)$$

und somit ergibt sich der Faktor  $c_\alpha$  als *oberes  $\alpha$ -Quantil* der Verteilung von  $T$ , d.h.

$$P\{t(X) \geq c_\alpha\} = \alpha \quad (7)$$

Da die Verteilung von  $t(X)$  im allgemeinen *unbekannt* ist, läßt sich das Quantil  $c_\alpha$  nicht konkret aus (7) bestimmen. Wir betrachten deshalb zunächst den Spezialfall, daß  $X$  eine *Normalverteilung* hat.

### **Exakte Konfidenzgrenzen bei normalverteiltem X**

Wenn die Zufallsvariable  $X$  eine  $N(\mu, \sigma^2)$ -Verteilung besitzt, so hat die Zufallsvariable  $t(X)$  eine sogenannte  *$t_m$ -Verteilung mit Freiheitsgrad  $m = n - 1$*  (vgl. 3.4.5). Wählen wir nun den Faktor  $c_\alpha$  als oberes  $\alpha$ -Quantil  $t_{m;\alpha}$  der  $t_m$ -Verteilung (vgl. Anhang T) so hat die zugehörige untere Konfidenzgrenze

$$\mu_{un,\alpha} = \bar{x} - t_{m;\alpha} \cdot \frac{\hat{\sigma}}{\sqrt{n}} \quad (\text{untere Konfidenzgrenze}) \quad (8)$$

die *exakte* Sicherheit  $1 - \alpha$ . Analog hat obere Konfidenzgrenze

$$\mu_{ob,\alpha} = \bar{x} + t_{m;\alpha} \cdot \frac{\hat{\sigma}}{\sqrt{n}} \quad (\text{obere Konfidenzgrenze}) \quad (9)$$

ebenfalls die *exakte* Sicherheit  $1 - \alpha$ . Und das *zweiseitige* Konfidenzintervall

$$(\hat{\mu}_{un,\frac{\alpha}{2}} = \hat{\mu}_{ob,\frac{\alpha}{2}}) \quad (\text{zweiseitige Konfidenzgrenze}) \quad (10)$$

hat auch die *exakte* Sicherheit  $1 - \alpha$ .

### **Asymptotische Konfidenzgrenzen bei beliebiger Verteilung von X**

Wir betrachten jetzt wieder den allgemeinen Fall mit einer *beliebigen* Verteilung (statt einer Normalverteilung) von  $X$ . In diesem Fall ist die Zufallsvariable  $t(X)$  *approximativ*  $N(0, 1)$ -verteilt und mit  $c_\alpha$  als  $\alpha$ -Quantil  $z_\alpha$  von  $N(0, 1)$  haben die Konfidenzgrenzen in (3) die *approximative* Sicherheit  $1 - \alpha$ , wobei die Approximation für  $n \rightarrow \infty$  beliebig genau wird. Das Quantil  $t_{m;\alpha}$  ist einerseits stets *größer* als  $z_\alpha$  und *strebt* andererseits *für wachsendes n gegen  $z_\alpha$* . Deshalb haben die Grenzen in (8) und (9) auch bei *beliebiger* Verteilung von  $X$  noch die *approximative* Sicherheit  $1 - \alpha$ . Der Vorteil der Grenzen (8) und (9) gegenüber denen mit  $c_\alpha = z_\alpha$  liegt darin, daß sie geringfügig *weiter* vom Mittelwert  $\bar{x}$  entfernt sind (was *zusätzliche* Sicherheit bedeutet) und bei *normalverteiltem X* sogar die

exakte Sicherheit  $1 - \alpha$  besitzen.

Die erforderlichen Rechenschritte sind in der *Konfidenz-Box 3* zusammengefaßt.

<b>Konfidenz-Box 3:</b> Für den <i>Erwartungswert</i> $m = E(X)$ einer Zufallsvariablen $X$ mit Standardabweichung $\sigma = SD(X)$ .	
Irrtums-Wahrscheinlichkeit:	$\alpha$
Sicherheit:	$1 - \alpha$
Stichprobenumfang:	$n$ ( <i>Anzahl der Wiederholungen</i> )
Beobachtete Stichprobe:	$x_1, x_2, \dots, x_n$ ( <i>unabhängige Wiederholungen von X</i> )
Schätzung von $\mu$ :	$\hat{\mu} = \bar{x}$ ( <i>Mittelwert</i> )
Schätzung von $\sigma$ :	$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2}$ ( <i>empir. Standardabweichung, vgl. 2.6.5 und Rechen-Box 1</i> )
Freiheitsgrad:	$m = n - 1$
<b>Einseitige Konfidenzgrenzen für <math>\mu</math>:</b>	
bei nicht zu kleinem $n$ (mindestens 10):	Sicherheit $\approx 1 - \alpha$
bei normalverteiltem $X$ :	Sicherheit $= 1 - \alpha$ (exakt)
<i>einseitige untere Grenze</i>	<i>einseitige obere Grenze</i>
$\hat{\mu}_{un,\alpha} = \bar{x} - \hat{d}_\alpha$	$\hat{\mu}_{ob,\alpha} = \bar{x} + \hat{d}_\alpha$
$\hat{d}_\alpha = t_{m;\alpha} \cdot \frac{\hat{\sigma}}{\sqrt{n}}$	( <i>Bandbreite</i> )
<b>Zweiseitiges Konfidenzintervall für <math>\mu</math>:</b> $(\hat{\mu}_{un,\frac{\alpha}{2}}, \hat{\mu}_{ob,\frac{\alpha}{2}})$	

**Bleigehalt im Apfelsaft:** Ein Getränkehersteller läßt regelmäßig den Bleigehalt  $X$  [in Mikrogramm pro Liter] seiner Apfelsaftproduktion überprüfen, um sicherzustellen, daß keine Ware mit einem zu hohen Bleigehalt ausgeliefert wird. Bei einer Überprüfung von  $n = 10$  zufällig ausgewählten Proben ergab sich der mittlere Bleigehalt zu  $\bar{x} = 536 [\mu\text{g/l}]$  und die empirische Standardabweichung zu  $\hat{\sigma} = 79 [\mu\text{g/l}]$ . Die *einseitige obere* 95%-Konfidenzgrenze für den erwarteten (durchschnittlichen) Bleigehalt  $\mu = E(X)$  ergibt sich wie folgt:

$$m = 9; \quad t_{9;5\%} = 1,833; \quad \hat{d}_{5\%} = 45,795; \quad \mu_{ob,5\%} = 581,8 [\mu\text{g/l}]. \quad \square$$

### 3.4.2 Tests für den Erwartungswert

Für die Zufallsvariable  $X$  soll jetzt ihr Erwartungswert  $\mu = E(X)$  mit einem festen Referenzwert  $\mu_0$  verglichen werden, wobei das zugehörige Testproblem wie folgt ein- oder zweiseitig gestellt sein kann:

$H_0: \mu \leq \mu_0,$	$H_1: \mu > \mu_0$	(einseitig mit oberer Alternative),
$H_0: \mu \geq \mu_0,$	$H_1: \mu < \mu_0$	(einseitig mit unterer Alternative),
$H_0: \mu = \mu_0,$	$H_1: \mu \neq \mu_0$	(zweiseitig).

In der Praxis hängt es von der interessierenden Fragestellung ab, welches dieser drei Testprobleme relevant ist, und wir geben hierfür einige typische Beispiele.

- Ist  $X$  ein Schadstoffgehalt (z.B. in einem Nahrungsmittel oder in der Luft) und  $\mu_0$  der zulässige gesetzliche Grenzwert, so ist der Konsument am einseitigen Testproblem interessiert
 
$$\begin{aligned} H_0: \mu \leq \mu_0 && (\text{Grenzwert wird nicht überschritten}), \\ H: \mu > \mu_0 && (\text{Grenzwert wird überschritten}). \end{aligned}$$
- Ist  $X$  die Lebensdauer eines technischen Produkts und  $\mu_0$  die vom Hersteller garantierte Mindestlebensdauer, so ist für die Qualitätskontrolle das einseitige Testproblem von Interesse:
 
$$\begin{aligned} H_0: \mu \geq \mu_0 && (\text{Mindestlebensdauer wird erreicht}), \\ H: \mu < \mu_0 && (\text{Mindestlebensdauer wird nicht erreicht}). \end{aligned}$$
- Ist  $X$  der Wirkstoffgehalt eines Medikaments und  $\mu_0$  der Sollgehalt laut Packungsbeilage, so interessiert man sich für das zweiseitige Testproblem
 
$$\begin{aligned} H_0: \mu = \mu_0 && (\text{korrekte Dosierung}), \\ H: \mu \neq \mu_0 && (\text{falsche Dosierung: Über- oder Unterdosierung}). \end{aligned}$$

#### Einseitiger Test mit oberer Alternative

Wir beginnen mit den *einseitigen* Hypothesen und *oberer Alternative*

$$\begin{array}{lll} \text{Nullhypothese} & H_0: \mu \leq \mu_0 & (\text{Referenzwert nicht überschritten}) \\ \text{Alternative} & H: \mu > \mu_0 & (\text{Referenzwert überschritten}). \end{array} \quad \text{vs. (1)}$$

Den Ausgangspunkt bilden wieder die Realisierungen  $x_1, x_2, \dots, x_n$  von  $n$  unabhängigen Wiederholungen  $X_1, X_2, \dots, X_n$  der Zufallsvariablen  $X$  mit den zugehörigen Schätzungen  $\hat{\mu}$  und  $\hat{\sigma}$  aus 3.4.1 (1)-(2), vgl. auch *Rechen-Box 1*.

Als Testwert verwendet man die Abweichung des Mittelwerts vom Referenzwert  $\mu_0$  geteilt durch seine (geschätzte) Standardabweichung  $\frac{1}{\sqrt{n}}\hat{\sigma}$ :

$$t := \frac{(\bar{x} - \mu_0)\sqrt{n}}{\hat{\sigma}} \quad (\text{Testwert}). \quad (2)$$

Da große Werte von  $t$  gegen die Nullhypothese sprechen, lehnt man die Nullhypothese ab, wenn der Testwert einen kritischen Wert übersteigt. Als kritischer Wert benutzt man das obere  $\alpha$ -Quantil  $t_{m;\alpha}$  (vgl. Anhang T) der  $t_m$ -Verteilung mit Freiheitsgrad  $m = n - 1$ , weil der Testwert unter dem Spezialfall  $\mu = \mu_0$  der Nullhypothese approximativ eine  $t_m$ -Verteilung besitzt, und dies ergibt den folgenden Test.

$$\text{Einseitiger } t\text{-Test: Ablehnung von } H_0: \mu \leq \mu_0 \Leftrightarrow t \geq t_{m,\alpha}. \quad (3)$$

Das Fehlerrisiko 1. Art dieses Tests ist *approximativ* gleich  $\alpha$ , wobei die Approximation für wachsenden Stichprobenumfang  $n$  beliebig genau wird. Wenn die Zufallsvariable  $X$  *normalverteilt* ist, so ist das Fehlerrisiko sogar *für jedes  $n$  exakt* gleich  $\alpha$ .

Die Testentscheidung läßt sich auch unter Verwendung des *P-Werts* herbeiführen, der (approximativ) die Wahrscheinlichkeit dafür angibt, daß die zugehörige *zufällige Testvariable* unter dem Spezialfall  $\mu = \mu_0$  der Nullhypothese *mindestens* so groß ist wie der *beobachtete* Wert  $t$ :

$$\Phi_m(-t) \approx P\{\text{zufällige Testvariable} \geq \text{beobachteter Testwert } t \mid \mu = \mu_0\} \quad (4)$$

wobei  $\Phi_m$  die *Verteilungsfunktion* der  $t_m$ -Verteilung ist (vgl. 3.4.5). Es gilt (vgl. Abb. 1)

$$t \geq t_{m,\alpha} \Leftrightarrow \Phi_m(-t) \leq \alpha. \quad (5)$$

Schließlich kann der Test auch unter Verwendung der *unteren Konfidenzgrenze* zur Sicherheit  $1 - \alpha$  entschieden werden: Die Nullhypothese wird abgelehnt, wenn  $\mu_0$  unterhalb der unteren Grenze liegt:

$$t \geq t_{m,\alpha} \Leftrightarrow \mu_0 \leq \hat{\mu}_{un,\alpha}. \quad (6)$$

Die verschiedenen Entscheidungs-Kriterien sind mit den erforderlichen Rechenschritten in der Test-Box 3 zusammengestellt.

**Bleigehalt im Apfelsaft (Fortsetzung 1):** Anhand der Stichprobe mit

$$n = 10 \quad \bar{x} = 536 [\mu\text{g/l}] \quad \hat{\sigma} = 79 [\mu\text{g/l}]$$

soll zum Niveau  $\alpha = 5\%$  überprüft werden, ob der Richtwert von  $\mu_0 = 500 [\mu\text{g/l}]$  eingehalten wird ( $H_0: \mu \leq \mu_0$ ) oder nicht ( $H: \mu > \mu_0$ ). Der  $t$ -Test lehnt die Nullhypothese nicht ab, weil

$$\begin{aligned} t &= 1,441 &< & 1,833 = t_{9,5\%} \\ \Phi_m(-t) &= 9,17\% &> & 5,0\% = \alpha \\ \hat{\mu}_{un,5\%} &= 490,2 &< & 500 = \mu_0. \end{aligned}$$

Wir haben hier alle *drei* Entscheidungskriterien angegeben, obwohl es prinzipiell reicht, *eine* der drei äquivalenten Bedingungen zu überprüfen.

Da der Test die Nullhypothese  $H_0: \mu \leq \mu_0$  nicht abgelehnt hat, ist ein Fehler 2. Art möglich, dessen Fehlerrisiko wir später untersuchen.  $\square$

### **Einseitiger Test mit unterer Alternative**

Für die *einseitigen* Hypothesen mit *unterer* Alternative,

$$\begin{aligned} \text{Nullhypothese } H_0: \mu \geq \mu_0 & \quad (\text{Referenzwert nicht unterschritten}) & \text{vs.} & \quad (7) \\ \text{Alternative } H: \mu < \mu_0 & \quad (\text{Referenzwert unterschritten}), \end{aligned}$$

wird ebenfalls der Testwert  $t$  aus (2) verwendet. Da jetzt hohe *negative* Werte von  $t$  gegen die Nullhypothese sprechen, ergibt sich analog zu (3) folgender

$$\boxed{\text{Einseitiger } t\text{-Test: Ablehnung von } H_0: \mu \geq \mu_0 \Leftrightarrow t \leq -t_{m;\alpha}.} \quad (8)$$

Die Testentscheidung lässt sich sowohl unter Verwendung des *P-Werts* (vgl. Abb. 1)

$$\boxed{\Phi_m(t) \approx P\{\text{zufällige Testvariable} \leq \text{beobachteter Testwert } t \mid \mu = \mu_0\}} \quad (9)$$

als auch mit der *oberen Konfidenzgrenze* zur Sicherheit  $1 - \alpha$  entscheiden (vgl. Test-Box 3).

**Zweiseitiger Test:** Der Test für die *zweiseitigen* Hypothesen,

$$\begin{aligned} \text{Nullhypothese } H_0: \mu = \mu_0 & \quad (\text{Referenzwert eingehalten}) & \text{vs.} & \quad (10) \\ \text{Alternative } H: \mu \neq \mu_0 & \quad (\text{Referenzwert nicht eingehalten}), \end{aligned}$$

lässt sich wieder auf die beiden einseitigen Tests jeweils zum halben Niveau zurückführen (vgl. Abb. 1). Die Details sind in Test-Box 3 angegeben.

**Test-Box 3:** Für den Erwartungswert  $\mu = E(X)$  einer Zufallsvariablen  $X$  mit Standardabweichung  $\sigma = SD(X)$ .

einseitig oben	einseitig unten	zweiseitig
$H_0: \mu \leq \mu_0$ vs. $H: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ vs. $H: \mu < \mu_0$	$H_0: \mu = \mu_0$ vs. $H: \mu \neq \mu_0$
Fehlerrisiko 1. Art: Stichprobenumfang: Beobachtete Stichprobe:	$\alpha$ $n$ $x_1, x_2, \dots, x_n$	(Testniveau) (Anzahl der Wiederholungen) (unabhängige Wiederholungen von $X$ )
Schätzung von $\mu$ :	$\hat{\mu} = \bar{x}$	(Mittelwert)
Schätzung von $\sigma$ :	$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2}$	(empirische Standardabweichung, vgl. 2.6.5 und Rechen-Box 1)
Freiheitsgrad:	$m = n - 1$	

### Student's t-Test:

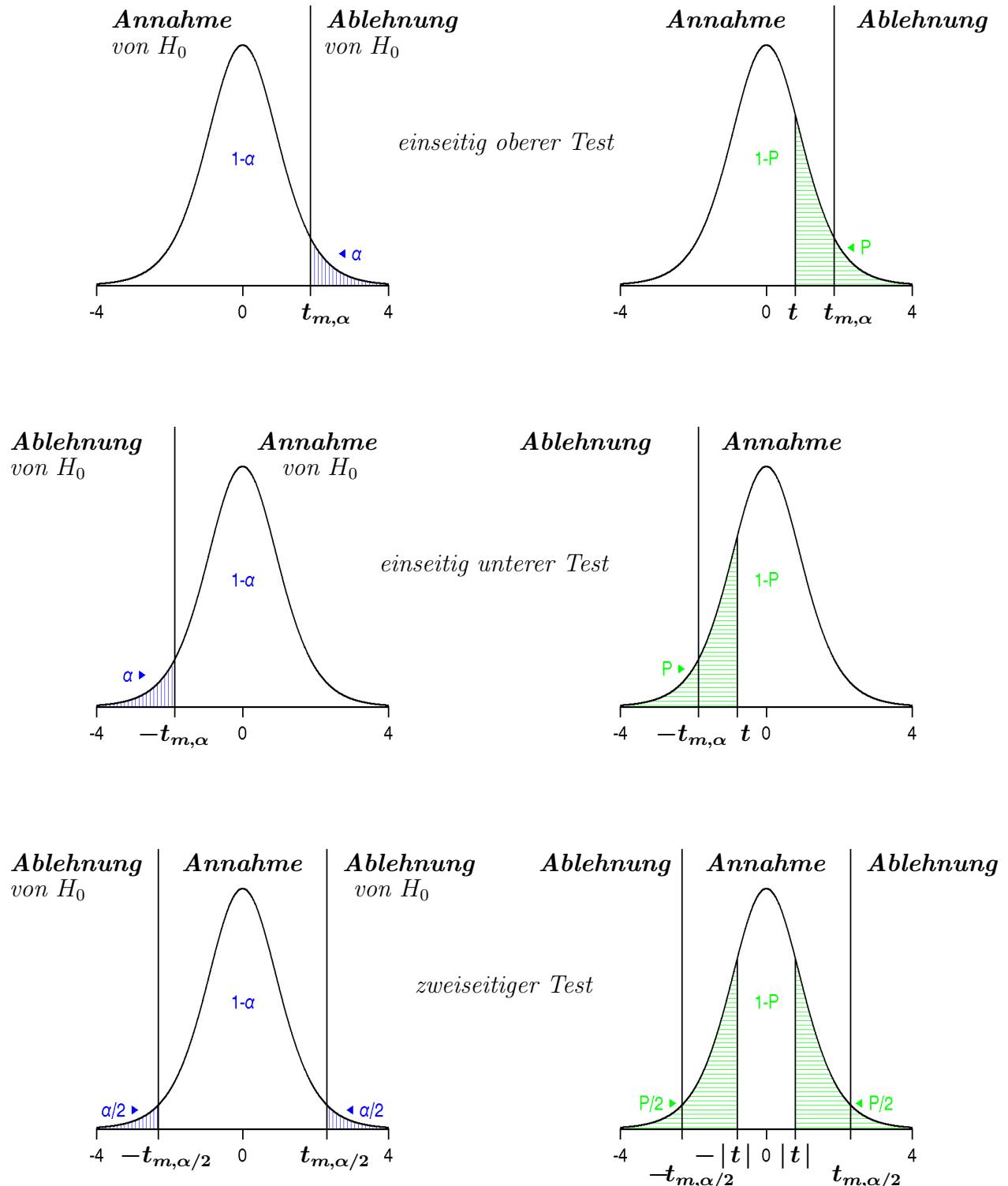
- bei nicht zu kleinem  $n$  (mindestens 10):      Fehlerrisiko 1. Art  $\approx \alpha$   
 bei normalverteiltem  $X$ :                          Fehlerrisiko 1. Art  $= \alpha$  (exakt)

$$\text{Testwert: } t = \frac{(\bar{x} - \mu_0) \sqrt{n}}{\hat{\sigma}}$$

Entscheidung: Ablehnung von  $H_0$  (Annahme von  $H$ ), falls gilt

einseitig oben	einseitig unten	zweiseitig
Testwert $t \geq t_{m;\alpha}$ bzw. $P\text{-Wert } \Phi_m(-t) \leq \alpha$ bzw. $\mu_0 \leq \hat{\mu}_{un,\alpha}$	Testwert $t \leq -t_{m;\alpha}$ bzw. $P\text{-Wert } \Phi_m(t) \leq \alpha$ bzw. $\hat{\mu}_{ob,\alpha} \leq \mu_0$	Testwert $ t  \geq t_{m;\frac{\alpha}{2}}$ bzw. $P\text{-Wert } 2\Phi_m(- t ) \leq \alpha$ bzw. $\mu_0 \leq \hat{\mu}_{un,\frac{\alpha}{2}}$ oder $\hat{\mu}_{ob,\frac{\alpha}{2}} \leq \mu_0$

**Erläuterung zur Test-Box 3:** Das  $\alpha$ -Quantil  $t_{m;\alpha}$  ist im Anhang T tabelliert. Die Verteilungsfunktion  $\Phi_m$  der  $t_m$ -Verteilung kann per Computer berechnet oder nach 3.4.5 approximiert bzw. aus Abb. 3 ungefähr abgelesen werden. Zur Berechnung der Konfidenzgrenzen  $\hat{\mu}_{un,\alpha}$  und  $\hat{\mu}_{ob,\alpha}$  vgl. Konfidenz-Box 3.



**Abb. 1:** Ablehnungsbereiche und  $P$ -Wert beim  $t$ -Test (beachte die Analogie zu 3.2, Abb. 4)

**links:** Dichte der  $t_m$ -verteilten Testvariablen (für  $\mu = \mu_0$ ) mit Ablehnungsbereichen der Nullhypothese beim ein- und zweiseitigen Test zum Fehlerrisiko 1. Art  $\alpha = 5\%$ .

**rechts:** Der zum Testwert  $t$  gehörige  $P$ -Wert  $P$  als markierte Fläche.

**Arzneimittelkontrolle** (*Kinder-Osius-Timm (1982), Beispiel 7.18*): Bei einer schmerzstillenden Tablette ist der Gehalt  $X$  [in mg] des Wirkstoffs Acetylsalicylsäure (produktionsbedingt) eine Zufallsvariable mit Erwartungswert  $\mu$  und Standardabweichung  $\sigma$ . Eine *Verbraucherorganisation* will zum Niveau  $\alpha = 5\%$  überprüfen, ob der *Sollwert* von  $\mu_0 = 300 \text{ mg}$  eingehalten wird oder nicht, d.h. es liegt das *zweiseitige* Testproblem vor mit

$$\begin{array}{ll} \text{Nullhypothese} & H_0: \mu = \mu_0 \quad (\text{Sollwert eingehalten: Ware einwandfrei}), \\ \text{Hypothese} & H: \mu \neq \mu_0 \quad (\text{Sollwert nicht eingehalten: Ware nicht einwandfrei}). \end{array}$$

Die Analyse einer zufälligen Stichprobe des Marktes ergab

$$n = 20; \quad \hat{\mu} = \bar{x} = 286,0 \text{ mg}; \quad \hat{\sigma} = s_x = 20,97.$$

Der  $t$ -Test zum Niveau  $\alpha = 5\%$  mit  $n - 1 = 19$  Freiheitsgraden lehnt hier  $H_0$  ab, weil

$$|t| = 2,986 \geq t_{19,\alpha/2} = 2,093.$$

Der mit Computer *exakt* berechnete P-Wert  $P = 2\Phi_{19}(-2,986) = 0,76\%$  stimmt hier gut überein mit der *Approximation* nach **3.4.5** (1):

$$P \approx 2 \cdot \Phi(d \cdot \sqrt{c}) = 2 \cdot \Phi(-2,67) = 0,76\% \quad \text{mit} \quad c = 0,0212; \quad d = -18,34. \quad \square$$

## Tipps für *Microsoft Excel*

**Funktionen:** Die *Quantile* der  $t$ -Verteilung (vgl. Anhang **T**) lassen sich mit folgenden *Statistik-Funktionen* in *Excel* berechnen:

$$t_{m;\frac{\alpha}{2}} = \text{TINV}(\alpha; m), \quad t_{m;\alpha} = \text{TINV}(2\alpha; m).$$

Man beachte dabei den Wechsel von  $\frac{\alpha}{2}$  zu  $\alpha$  bzw. von  $\alpha$  zu  $2\alpha$ .

Statt der *Verteilungsfunktion*  $\Phi_m(x)$  der  $t_m$ -Verteilung ist in *Excel* die *komplementäre* Funktion implementiert – allerdings nur für *positive* Argumente  $x$

$$\begin{aligned} 1 - \Phi_m(x) &= \text{TVERT}(x; m; 1) \quad \text{für } x \geq 0 \quad \text{bzw.} \\ \Phi_m(x) &= 1 - \text{TVERT}(x; m; 1) \quad \text{für } x \geq 0. \end{aligned}$$

Das *letzte* Argument der Funktion **TVERT** muß hier = 1 sein (es steht für *einseitig*).

Unter Verwendung der Symmetrie-Eigenschaft

$$\Phi_m(-x) = 1 - \Phi_m(x)$$

lässt sich dann  $\Phi_m(x)$  für *positives und negatives*  $x$  durch folgende Formel in *Excel* berechnen:

$$\Phi_m(x) = \text{WENN}(x \geq 0; 1 - \text{TVERT}(x; m; 1); \text{TVERT}(-x; m; 1)).$$

### 3.4.3 Fehlerrisiko 2. Art und Testschärfe (Power)

Wenn man die Nullhypothese *nicht* abgelehnt hat, so ist ein Fehler 2. Art möglich, dessen Risiko man (zumindest für relevante Werte aus der Alternative) berechnen sollte und zwar eigentlich schon bei der Planung des Versuchs (vgl. auch 3.4.4). Beim *einseitigen oberen* Testproblem

$$H_0: \mu \leq \mu_0 \quad \text{vs.} \quad H: \mu > \mu_0 \quad (1)$$

hängen das Fehlerrisiko 2. Art  $\beta = \beta(\mu_1)$  und die dazu komplementäre Testschärfe  $\gamma = \gamma(\mu_1)$  sowohl von dem konkreten Wert  $\mu_1 > \mu_0$  aus der Alternative als auch von der Standardabweichung  $\sigma$  von  $X$  ab. Je *kleiner* der Unterschied  $\mu_1 - \mu_0$  und je *größer* die Standardabweichung  $\sigma$  ist, desto *größer* ist das Fehlerrisiko  $\beta(\mu_1)$ . Die Berechnung von  $\beta(\mu_1)$  basiert auf der *approximativen* Normalverteilung der Test-Variablen und die erforderlichen Schritte sind in der *Power-Box 3* zusammengestellt.

Für das *einseitig untere* Testproblem wird die Schärfe analog zum *einseitig oberen* Test bestimmt und beim *zweiseitigen* Testproblem ergibt sich die Schärfe aus dem zugehörigen einseitigen Test zum *halben* Niveau. Die Details sind in der *Power-Box 3* angegeben.

<b>Power-Box 3:</b> Für Tests über den <i>Erwartungswert</i> $\mu = E(X)$ einer Zufallsvariablen $X$ mit Standardabweichung $\sigma = SD(X)$ .		
<i>einseitig oben</i>	<i>einseitig unten</i>	<i>zweiseitig</i>
$H_0: \mu \leq \mu_0$ <i>vs.</i> $H: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ <i>vs.</i> $H: \mu < \mu_0$	$H_0: \mu = \mu_0$ <i>vs.</i> $H: \mu \neq \mu_0$
Fehlerrisiko 1. Art (Testniveau): $\alpha$ Stichprobenumfang: $n$ Wert aus der <i>Alternative</i> $H$ : $\mu_1$	Standardabweichung: $\sigma$ Freiheitsgrad: $m = n - 1$	
<i>einseitig oben</i>	<i>einseitig unten</i>	<i>zweiseitig</i>
$t_m = t_{m;\alpha}$	$t_m = t_{m;\alpha}$	$t_m = t_{m;\frac{\alpha}{2}}$
<i>Hilfsgrößen:</i>	$\Delta =  \mu_1 - \mu_0  > 0$ $u = t_m - \frac{\Delta\sqrt{n}}{\sigma}$	(Abweichung von $H_0$ )
<i>Fehlerrisiko 2. Art:</i> <i>Testschärfe (Power):</i>	$\beta(\mu_1) \approx \Phi(u)$ $\gamma(\mu_1) \approx \Phi(-\mu)$	mit $\Phi = \text{Verteilungsfunktion}$ von $N(0, 1)$

**Erläuterung zur Power-Box 3:** Das  $\alpha$ -Quantil  $t_{m;\alpha}$  und die Verteilungsfunktion  $\Phi$  der  $N(0, 1)$ -Verteilung sind im Anhang T tabelliert oder können per Computer berechnet werden. – Wenn die betrachtete Zufallsvariable  $X$  normalverteilt ist, so lässt sich das Fehlerrisiko 2. Art mit fortgeschrittenen Methoden sogar *exakt* (statt *approximativ* wie in der Power-Box 3) berechnen, vgl. z.B. Kinder-Osius-Timm (1982), 4.5.2 und 7.3.1.

**Bleigehalt im Apfelsaft (Fortsetzung 2):** Wir wollen jetzt das Fehlerrisiko  $\beta$  für den Fall bestimmen, daß der Richtwert um 10% überschritten wird, d.h. für  $\mu_1 = 550[\mu\text{g/l}]$  bzw.  $\Delta = 50$ . Da die Standardabweichung  $\sigma$  nicht bekannt ist, verwenden wir in der Power-Box 3 ihre *Schätzung*, d.h.  $\sigma = 79$ . Damit ergibt sich  $u = -0,168$  und somit das relativ hohe Fehlerrisiko  $\beta(\mu_1) = \Phi(u) \approx 43\%$ .  $\square$

### 3.4.4 Versuchsplanung: erforderlicher Stichprobenumfang

Es sei  $\mu_1$  ein für die konkrete Anwendung *relevanter* Wert aus der Alternative und  $\Delta = |\mu_1 - \mu_0| > 0$  die zugehörige Abweichung von der Nullhypothese. Dann fällt das Fehlerrisiko  $\beta(\mu_1)$  bzw. die Schärfe  $\gamma(\mu_1)$  wächst, wenn man den Stichprobenumfang  $n$  erhöht. Im Rahmen einer *Versuchsplanung* kann man wieder für einen vorgegebenen Wert  $\beta$  den erforderlichen Mindestumfang  $n$  bestimmen (vgl. Planungs-Box 3), bei denen  $\beta(\mu_1)$  höchstens gleich der Vorgabe  $\beta$  ist. Typischerweise verwendet man für die Vorgabe  $\beta$  wieder das *doppelte* Fehlerrisiko 1. Art, also  $\beta = 2\alpha$  (z.B.  $\beta = 10\%$  bei  $\alpha = 5\%$ ). Die Einzelheiten sind in der Planungs-Box 3 zusammengefasst. Da die dortigen Berechnungen auf vereinfachten Approximation beruhen, sollte man den resultierenden Stichprobenumfang sicherheitshalber um 1 erhöhen.

**Bleigehalt im Apfelsaft (Fortsetzung 3):** Soll für  $\mu_1 = 550[\mu\text{g/l}]$  bzw.  $\Delta = 50$  das Fehlerrisiko  $\beta = 10\%$  betragen, so erhält man den erforderlichen Mindestumfang bei einer Standardabweichung von  $\sigma = 79$  nach Planungs-Box 3 aus

$$a = 5\%; \quad z_\alpha = 1,645; \quad b = 10\%; \quad z_\beta = 1,282$$

zu  $n \approx 21,4$ . Dies ergibt aufgerundet 22 (also mehr als doppelt soviel wie der verwendete Umfang von 10) und sicherheitshalber sogar 23.  $\square$

**Planungs-Box 3:** Für Test über den *Erwartungswert*  $\mu = E(X)$  einer Zufallsvariablen  $X$  mit Standardabweichung  $\sigma = SD(X)$ .

einseitig oben	einseitig unten	zweiseitig
$H_0: \mu \leq \mu_0$ vs. $H: \mu > \mu_0$	$H_0: \mu \geq \mu_0$ vs. $H: \mu < \mu_0$	$H_0: \mu = \mu_0$ vs. $H: \mu \neq \mu_0$

<b>Vorgaben:</b>	Fehlerrisiko 1. Art:	$\alpha$
	Standardabweichung:	$\sigma$
	Wert aus der <i>Alternative</i> $H$ :	$\mu_1$
	Fehlerrisiko 2. Art für $\mu = \mu_1$ :	$\beta$
Schärfe	für $\mu = \mu_1$ :	bzw. $\gamma = 1 - \beta$

**Hilfsgrößen:**  $\Delta = |\mu_1 - \mu_0| > 0$  (*Abweichung von  $H_0$* )

einseitig oben	einseitig unten	zweiseitig
$z = z_\alpha$	$z = z_\alpha$	$z = z_{\frac{\alpha}{2}}$

*Erforderlicher Mindestumfang*

$$n \approx \left[ \frac{(z + z_\beta)\sigma}{\Delta} \right]^2$$

(aufzurunden auf ganze Zahl, dann sicherheitshalber um 1 erhöhen)

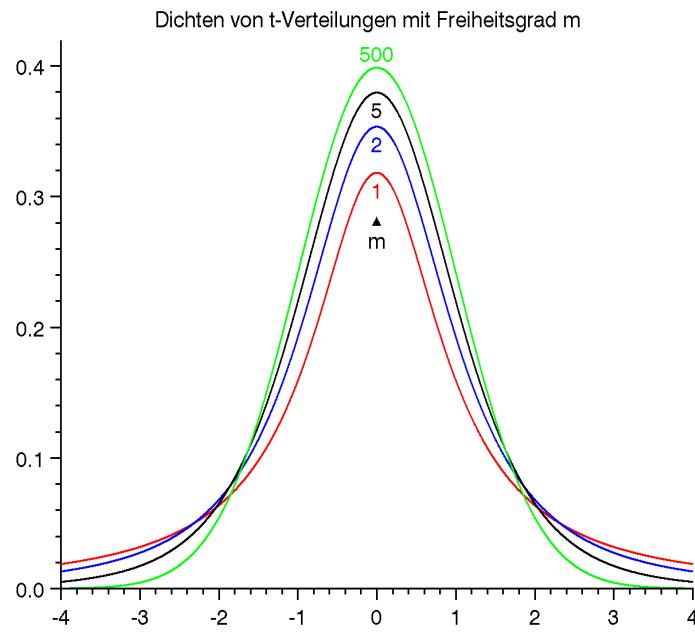
**Erläuterung zur Planungs-Box 3:** Die Quantile  $z_\alpha$  und  $z_\beta$  sind im Anhang T tabelliert oder können per Computer berechnet werden. – Halbiert man die Abweichung  $\Delta$ , so vervierfacht sich der erforderliche Umfang  $n$ .

Für *normalverteiltes*  $X$  lässt sich der Mindestumfang auch *exakt* angeben – vgl. z.B. Kinder-Osius-Timm (1982), 4.5.2 – und stimmt typischerweise mit dem *um 1 erhöhten Umfang* der Planungs-Box 3 überein.

### 3.4.5 Die Student'sche $t$ -Verteilung

#### Die Dichtefunktion

Die *Dichtefunktion*  $\varphi_m$  der  $t_m$ -Verteilung (Abb. rechts) ist eine um den Nullpunkt symmetrische Glocke, die für wachsenden Freiheitsgrad (FG)  $m$  gegen die Dichtefunktion  $\varphi$  der  $N(0, 1)$ -Verteilung strebt. Im Rahmen der Zeichengenauigkeit entspricht  $m = 500$  bereits der  $N(0, 1)$ -Verteilung.



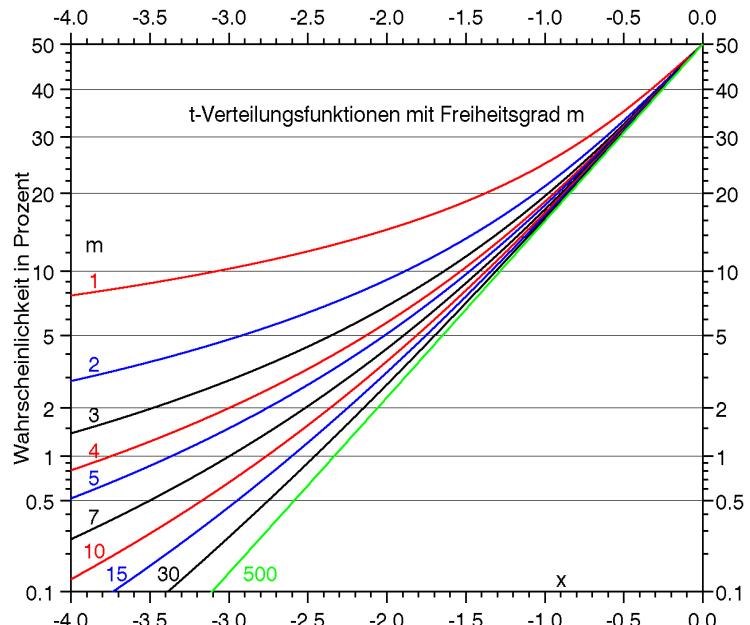
#### Die Verteilungsfunktion

Die *Verteilungsfunktion*  $\Phi_m$  der  $t_m$ -Verteilung auf einer Wahrscheinlichkeits-Skala (Abb. rechts) für *negative* Argumente  $x$ .

Für *positive* Argumente  $x$  gilt (wegen der Symmetrie der Dichtefunktion):

$$\Phi_m(x) = 1 - \Phi_m(-x).$$

Für  $m = 500$  ergibt sich bei der Zeichengenauigkeit bereits die  $N(0, 1)$ -Verteilungsfunktion  $\Phi$ .



Die Verteilungsfunktion  $\Phi_m$  der  $t_m$ -Verteilung mit  $m$  Freiheitsgraden lässt sich mit geeigneten Computer-Programmen (z.B. der Tabellenkalkulationen *Excel*) berechnen. Wenn nur ein Taschenrechner und eine Tabelle der Verteilungsfunktion  $\Phi$  der Standard-Normalverteilung  $N(0, 1)$  zur Verfügung steht, so kann man die folgende *Approximation* verwenden:

$$\Phi_m(a) \approx \Phi(d \cdot \sqrt{c}) \quad \text{mit} \quad (1)$$

$$c = \frac{\ln(1 + \frac{a^2}{m})}{m - \frac{5}{6}} \quad \text{und} \quad d = \pm \left( m - \frac{2}{3} + \frac{1}{10m} \right) \quad \text{mit} \quad \begin{cases} + & \text{für } a > 0 \\ - & \text{für } a < 0 \end{cases} \quad (2)$$

wobei  $d$  dasselbe Vorzeichen wie  $a$  hat. Für  $a = 0$  ist  $c = 0$  und  $\Phi_m = \frac{1}{2}$  für alle  $m$ . Diese Approximation wird bei wachsendem Freiheitsgrad  $m$  immer genauer und hat schon bei  $m = 2$  bzw.  $m = 4$  einen absoluten Fehler von weniger als 1% bzw. 0,1% .

**Für Fortgeschrittene:** Die Dichte  $\varphi_m$  der  $t_m$ -Verteilung ist gegeben durch

$$\varphi_m(x) = a_m \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}$$

wobei die Konstante  $a_m$  durch die *Eulersche Gammafunktion*  $\Gamma$  definiert ist

$$a_m = \Gamma\left(\frac{m+1}{2}\right) / \left(\Gamma\left(\frac{m}{2}\right)\sqrt{n\pi}\right).$$

□

### 3.5 Vergleich zweier Erwartungswerte

Nachdem wir bisher Konfidenzgrenzen und Tests für den Erwartungswert *einer* Zufallsvariable behandelt haben, betrachten wir jetzt *zwei* voneinander *unabhängige* Zufallsvariablen  $X$  und  $Y$ , deren Erwartungswerte  $\mu_X$  und  $\mu_Y$  wir miteinander *vergleichen* wollen. Typische Anwendungsbeispiele hierfür sind:

- Unterscheidet sich das Geburtsgewicht  $X$  der Jungen im Mittel vom Geburtsgewicht  $Y$  der Mädchen?
- Stimmen die Ergebnisse  $X$  und  $Y$  nach zwei verschiedenen Meßverfahren im Mittel überein?
- Unterscheiden sich zwei Medikamente hinsichtlich der gewünschten Wirkung, wie z.B. der (mittleren) Senkung des Blutdrucks?

Hierbei soll der *Unterschied*  $\Delta\mu = \mu_X - \mu_Y$  beider Erwartungswerte *geschätzt* werden (mit Angabe von Konfidenzgrenzen), und man will Hypothesen über die Erwartungswerte *testen*:

$$\begin{array}{lll} H_0: \mu_X \leq \mu_Y & \text{vs.} & H: \mu_X > \mu_Y \quad (\text{einseitig}) \\ H_0: \mu_X = \mu_Y & \text{vs.} & H: \mu_X \neq \mu_Y \quad (\text{zweiseitig}) . \end{array}$$

#### 3.5.1 Schätzung des Unterschieds mit Konfidenzgrenzen

Ausgangspunkt der statistischen Analyse sind  $n_X$  *unabhängige* Wiederholungen  $X_1, X_2, \dots, X_{n_X}$  der Zufallsvariablen  $X$  und *davon unabhängige* weitere  $n_Y$  unabhängige Wiederholungen  $Y_1, Y_2, \dots, Y_{n_Y}$  der Zufallsvariablen  $Y$ . Für die zugehörigen Realisierungen (beobachteten Werte)  $x_1, x_2, \dots, x_{n_X}$  und  $y_1, y_2, \dots, y_{n_Y}$  bestimmen wir zunächst für die  $x$ -Werte und  $y$ -Werte getrennt die jeweiligen Mittelwerte  $\bar{x}$  und  $\bar{y}$  als Schätzung der *Erwartungswerte*  $\mu_X = E(X)$  und  $\mu_Y = E(Y)$

$$\hat{\mu}_X = \bar{x}, \quad \hat{\mu}_Y = \bar{y} \quad (\text{Schätzung von } \mu_X, \mu_Y), \quad (1)$$

sowie die *empirischen* Standardabweichungen  $s_x$  und  $s_y$  als Schätzung der *Standardabweichungen*  $\sigma_X = SD(X)$ ,  $\sigma_Y = SD(Y)$

$$\hat{\sigma}_X = s_x, \quad \hat{\sigma}_Y = s_y \quad (\text{Schätzung von } \sigma_X, \sigma_Y), \quad (2)$$

vgl. hierzu auch **2.6.5** und *Rechen-Box 1*. Für den interessierenden Unterschied beider Erwartungswerte

$$\Delta\mu = \mu_X - \mu_Y \quad (\text{Differenz, Unterschied}) \quad (3)$$

ergibt sich dann als zugehörige Schätzung

$$\Delta\hat{\mu} = \hat{\mu}_X - \hat{\mu}_Y = \bar{x} - \bar{y} \quad (\text{geschätzte Differenz, geschätzter Unterschied}). \quad (4)$$

Zur Konstruktion von *Konfidenzgrenzen* für den Unterschied  $\Delta\mu$  benötigen wir wieder Information über die *Streuung* der Schätzung  $\Delta\hat{\mu}$ . Die Varianz der zugehörigen Differenz  $\bar{X} - \bar{Y}$  der Mittelwerte ist (nach 2.6.4) gleich der Summe beider Varianzen von  $\bar{X}$  und  $\bar{Y}$ , also (nach 2.6.5)

$$\sigma_{\Delta}^2 = \text{Var}(\bar{X} - \bar{Y}) = \frac{1}{n_X} \sigma_X^2 + \frac{1}{n_Y} \sigma_Y^2. \quad (5)$$

Ersetzt man nun die (typischerweise unbekannten) Varianzen  $\sigma_X^2$  und  $\sigma_Y^2$  durch ihre Schätzungen, so ergibt sich als Schätzung der Standardabweichung  $\sigma_{\Delta}$

$$\hat{\sigma}_{\Delta} = \sqrt{\frac{1}{n_X} \hat{\sigma}_X^2 + \frac{1}{n_Y} \hat{\sigma}_Y^2} \quad (\text{Schätzung der Standardabweichung von } \Delta\mu). \quad (6)$$

Unter Verwendung des  $\alpha$ -Quantils  $t_{m;\alpha}$  der  $t_m$ -Verteilung (vgl. Anhang T) mit dem Freiheitsgrad  $m = n_X + n_Y - 2$  (Gesamtanzahl der Beobachtungen minus 2) erhält man folgende *einseitigen* Konfidenzgrenzen für  $\Delta\mu$  zur *approximativen* Sicherheit  $1 - \alpha$ :

$$\Delta\hat{\mu}_{un,\alpha} = (\bar{x} - \bar{y}) - \hat{d}_{\alpha} \quad (\text{untere Konfidenzgrenze für } \Delta\mu), \quad (7)$$

$$\Delta\hat{\mu}_{ob,\alpha} = (\bar{x} - \bar{y}) + \hat{d}_{\alpha} \quad (\text{obere Konfidenzgrenze für } \Delta\mu), \quad \text{mit} \quad (8)$$

$$\hat{d}_{\alpha} = t_{m;\alpha} \cdot \hat{\sigma}_{\Delta} \quad (\text{Bandbreite}). \quad (9)$$

Das *zweiseitige* Konfidenzintervall ergibt sich dann wieder aus den einseitigen Grenzen zur halben Irrtums-Wahrscheinlichkeit:

$$(\Delta\hat{\mu}_{un,\frac{\alpha}{2}}, \Delta\hat{\mu}_{ob,\frac{\alpha}{2}}) \quad (\text{zweiseitiges Konfidenzintervall}). \quad (10)$$

Diese Konfidenzgrenzen haben allerdings nur die *approximative* Sicherheit  $1 - \alpha$ , wobei die Approximation für wachsende Stichprobenumfänge  $n_X$  und  $n_Y$  immer besser wird. Konfidenzgrenzen mit einer *exakten* Sicherheit lassen sich (bei unbekannten Standardabweichungen  $\sigma_X$  und  $\sigma_Y$ ) auf diese Weise nur unter zusätzlichen Annahmen (vgl. unten) an die Verteilung von  $X$  und  $Y$  herleiten.

**Konfidenz-Box 4:** Vergleich der Erwartungswerte  $\mu_X$  und  $\mu_Y$  von zwei unabhängigen Zufallsvariablen  $X$  und  $Y$  mit Standardabweichungen  $\sigma_X$  und  $\sigma_Y$ .

Irrtums-Wahrscheinlichkeit:	$\alpha$
Sicherheit:	$1 - \alpha$
Stichprobenumfänge:	$n_X$ ( $x$ -Werte), $n_Y$ ( $y$ -Werte)
Beobachtete Stichproben:	$x_1, x_2, \dots, x_{n_X}$ (unabhängige Wiederholungen von $X$ ) $y_1, y_2, \dots, y_{n_Y}$ (unabhängige Wiederholungen von $Y$ )
Schätzung von $\mu_X$ und $\mu_Y$ :	$\hat{\mu}_X = \bar{x}$ (Mittelwerte: $x$ - bzw. $y$ -Werte) $\hat{\mu}_Y = \bar{y}$
Schätzung von $\sigma_X$ und $\sigma_Y$ :	$\hat{\sigma}_X = s_x$ (empirische Standardabweichung: $\hat{\sigma}_Y = s_y$ $x$ - bzw. $y$ -Werte, vgl. Rechen-Box 1)
Schätzung des Unterschieds $\Delta\mu = \mu_X - \mu_Y$ :	$\Delta\hat{\mu} = \hat{\mu}_X - \hat{\mu}_Y$
Schätzung der Standardabweichung von $\Delta\mu$ :	$\hat{\sigma}_\Delta = \sqrt{\frac{1}{n_X} \hat{\sigma}_X^2 + \frac{1}{n_Y} \hat{\sigma}_Y^2}$
Freiheitsgrad:	$m = n_X + n_Y - 2$

### Einseitige asymptotische Konfidenzgrenzen für $\Delta\mu$

Anwendbar bei nicht zu kleinen  $n_X, n_Y$  (beide mindestens 10): Sicherheit  $\approx 1 - \alpha$

einseitige untere Grenze	einseitige obere Grenze
$\Delta\hat{\mu}_{un,\alpha} = \Delta\hat{\mu} - \hat{d}_\alpha$	$\Delta\hat{\mu}_{ob,\alpha} = \Delta\hat{\mu} + \hat{d}_\alpha$
$\hat{d}_\alpha = t_{m;\alpha} \cdot \hat{\sigma}_\Delta$	(Bandbreite)
<b>Zweiseitiges Konfidenzintervall für <math>\Delta\mu</math>:</b> $(\Delta\hat{\mu}_{un,\frac{\alpha}{2}}, \Delta\hat{\mu}_{ob,\frac{\alpha}{2}})$	

### Exakte Grenzen bei Normalverteilungen mit homogenen Varianzen

Wir gehen jetzt davon aus, daß sowohl  $X$  als auch  $Y$  normalverteilt sind und daß ihre Varianzen bzw. Standardabweichungen übereinstimmen:

$$\sigma_X = \sigma_Y \quad (\text{Homogene Standardabweichungen bzw. Varianzen}). \quad (\text{H})$$

Unter Ausnutzung der Homogenität der Varianzen läßt sich die Standardabweichung der Mittelwerts-Differenz  $\bar{X} - \bar{Y}$  wie folgt schätzen (der Index  $H$  steht für *Homogenität*)

$$\hat{\sigma}_{\Delta H} = \sqrt{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right) \frac{(n_X - 1)\hat{\sigma}_X^2 + (n_Y - 1)\hat{\sigma}_Y^2}{m}} \quad \text{mit } m = n_X + n_Y - 2. \quad (11)$$

Verwendet man nun in (9) (bzw. in der Konfidenz-Box 5) die Schätzung  $\hat{\sigma}_{\Delta H}$  statt  $\hat{\sigma}_\Delta$ , d.h. die folgende Bandbreite

$$\hat{d}_\alpha = t_{m;\alpha} \cdot \hat{\sigma}_{\Delta H} \quad (\text{Bandbreite bei homogenen Varianzen}), \quad (9)_H$$

so haben die zugehörigen Grenzen aus (7) und (8) (bzw. aus der Konfidenz-Box 5) sogar die *exakte Sicherheit*  $1 - \alpha$ .

Bei konkreten Daten stellt sich nun die Frage, ob man Bandbreite (9)<sub>H</sub> oder die aus (9) verwenden soll. Im ersten Fall erhält man zwar Grenzen mit *exakter* Sicherheit, aber dies gilt nur unter zusätzlichen Voraussetzungen (*normalverteilte* Zufallsvariablen  $X$  und  $Y$  mit *homogenen* Varianzen), auf deren Gültigkeit man (z.B. aufgrund von Vorinformationen) vertrauen muß. Glücklicherweise unterscheiden sich die Schätzungen  $\hat{\sigma}_{\Delta H}$  und  $\hat{\sigma}_\Delta$  bei nicht zu kleinen Umfängen  $n_X$  und  $n_Y$  aber nur geringfügig, sodaß es keine wesentliche Rolle spielt, welche der beiden Schätzungen man verwendet. Im *balancierten* Design, d.h. bei *gleichen* Umfängen  $n_X = n_Y$  stimmen die Schätzungen sogar *exakt* überein:

$$\hat{\sigma}_{\Delta H} = \hat{\sigma}_\Delta \quad \text{falls} \quad n_X = n_Y. \quad (12)$$

In diesem Fall ist es dann egal, welche der beiden Darstellungen (9) oder (9)<sub>H</sub> man verwendet. Dies spricht auch dafür, bei der Versuchsplanung (vgl. 3.5.4) ein *balanciertes* Design zu verwenden.

**Geburtsgewicht und Geschlecht:** Für die *StatLab-Auswahl 1985* soll der Unterschied der (erwarteten) Geburtsgewichte [in Pound] zwischen beiden Geschlechtern geschätzt werden. Aus den Daten

$$\begin{array}{lll} \text{Jungen: } & n_X = 50 & \hat{\mu}_X = \bar{x} = 7,644 \\ & & \hat{\sigma}_X = s_x = 0,9881 \\ \text{Mädchen: } & n_Y = 50 & \hat{\mu}_Y = \bar{y} = 7,130 \\ & & \hat{\sigma}_Y = s_y = 1,2987 \end{array}$$

ergibt sich der geschätzte Unterschied der (erwarteten) Geburtsgewichte

$$\Delta\hat{\mu} = \hat{\mu}_X - \hat{\mu}_Y = 0,514 \text{ [Pound]} \quad \text{mit} \quad \hat{\sigma}_\Delta = 0,2308 \text{ [Pound].}$$

Das *zweiseitige* 95%-Konfidenzintervall für den Unterschied  $\Delta\mu$  ergibt sich aus  $\alpha = 5\%$ , dem Quantil  $t_{98,\alpha/2} = 1,984$  und der Bandbreite  $\hat{d}_\alpha = 0,458$  zu

$$\Delta\hat{\mu}_{un,\alpha/2} = 0,056 \text{ [Pound]} \quad \Delta\hat{\mu}_{ob,\alpha/2} = 0,972 \text{ [Pound].}$$

Da die untere Grenze  $\Delta\hat{\mu}_{un,\alpha/2}$  *positiv* ist, ist das erwartete Geburtsgewicht der Jungen mit der Sicherheit von 95% *größer* als das der Mädchen.  $\square$

### 3.5.2 Testen von Hypothesen

Wir wollen jetzt die beiden Erwartungswerte miteinander *vergleichen*, wobei das zugehörige Testproblem ein- oder zweiseitig gestellt werden kann.

*Einseitiger Test:*

$$\begin{array}{lll} H_0: \mu_X \leq \mu_Y & \text{bzw.} & \Delta\mu \leq 0 \\ H_1: \mu_X > \mu_Y & \text{bzw.} & \Delta\mu > 0 \end{array} \quad \begin{array}{l} (\text{kein positiver Unterschied}), \\ (\text{positiver Unterschied}). \end{array}$$

*Zweiseitiger Test:*

$$\begin{array}{lll} H_0: \mu_X = \mu_Y & \text{bzw.} & \Delta\mu = 0 \\ H_1: \mu_X \neq \mu_Y & \text{bzw.} & \Delta\mu \neq 0 \end{array} \quad \begin{array}{l} (\text{kein Unterschied}), \\ (\text{Unterschied}). \end{array}$$

Das duale einseitige Testproblem mit  $\mu_Y \leq \mu_X$  als Nullhypothese und  $\mu_Y > \mu_X$  als Alternative muß nicht separat betrachtet werden, weil es nach *Vertauschen* von  $X$  mit  $Y$  wieder von obiger (einseitigen) Form ist.

Ob man *ein-* oder *zweiseitig* testet, hängt von der konkreten Anwendungssituation ab. Will man einen *positiven* Unterschied nachweisen, so ist der *einseitige* Test angemessen. Ist man aber nur am Nachweis eines Unterschieds interessiert (egal ab positiv oder negativ), so wird man *zweiseitig* testen.

Als *Testwert* verwendet man den geschätzten Unterschied (d.h. die Differenz beider Mittelwerte) geteilt durch seine (geschätzte) Standardabweichung  $\hat{\sigma}_\Delta$  (vgl. 3.5.1):

$$t = \frac{\Delta\hat{\mu}}{\hat{\sigma}_\Delta} = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\hat{\sigma}_\Delta} = \frac{\bar{x} - \bar{y}}{\hat{\sigma}_\Delta} \quad (\text{Testwert}). \quad (1)$$

Beim *einseitigen* Testproblem sprechen große Werte von  $t$  *gegen* die Nullhypothese, die man daher ablehnt, wenn der Testwert einen kritischen Wert übersteigt. Als kritischer Wert benutzt man das obere  $\alpha$ -Quantil  $t_{m;\alpha}$  (vgl. Anhang T) der  $t_m$ -Verteilung mit Freiheitsgrad  $m = n_X + n_Y - 2$ , weil der Testwert unter dem Spezialfall  $\mu_X = \mu_Y$  der Nullhypothese approximativ eine  $t_m$ -Verteilung besitzt. Damit ergibt sich (analog 3.4.2) folgender Test.

<i>Einseitiger t-Test:</i>	Ablehnung von $H_0: \mu_X \leq \mu_Y \Leftrightarrow t \geq t_{m;\alpha}$ .	(2)
----------------------------	---	-----

Das Fehlerrisiko 1. Art dieses Tests ist *approximativ* gleich  $\alpha$ , wobei die Approximation für wachsende Umfänge  $n_X$  und  $n_Y$  beliebig genau wird. Die Testentscheidung lässt sich auch unter Verwendung des *P-Werts* entscheiden, der (approximativ) die Wahrscheinlichkeit dafür angibt, daß die zugehörige *zufällige Testvariable* unter dem Spezialfall  $\mu_X = \mu_Y$  der

Nullhypothese *mindestens* so groß ist wie der *beobachtete* Wert  $t$ ,

$$\Phi_m(-t) \approx P\{ \text{zufällige Testvariable} \geq \text{beobachteter Testwert } t \mid \mu_X = \mu_Y \} \quad (3)$$

mit  $\Phi_m$  als *Verteilungsfunktion* der  $t_m$ -Verteilung (vgl. 3.4.5). Es gilt (vgl. 3.4 Abb. 1)

$$t \geq t_{m;\alpha} \Leftrightarrow \Phi_m(-t) \leq \alpha. \quad (4)$$

Schließlich kann der Test auch unter Verwendung der *unteren Konfidenzgrenze* für den Unterschied  $\Delta\mu$  zur Sicherheit  $1 - \alpha$  entschieden werden: Die Nullhypothese wird abgelehnt, wenn die untere Grenze nicht negativ ist:

$$t \geq t_{m;\alpha} \Leftrightarrow \Delta\hat{\mu}_{un,\alpha} \geq 0. \quad (5)$$

Die verschiedenen Entscheidungs-Kriterien sind mit den erforderlichen Rechenschritten in der Test-Box 4 zusammengestellt.

Beim *zweiseitigen* Testproblem sprechen sowohl *hohe positive* als auch *hohe negative* Testwerte gegen die Nullhypothese und daraus ergibt sich (analog 3.4.2) folgender Test zum Niveau  $\alpha$  (vgl. 3.4, Abb. 4):

$$\text{Zweiseitiger } t\text{-Test:} \quad \text{Ablehnung von } H_0: \mu_X = \mu_Y \Leftrightarrow |t| \geq t_{m;\frac{\alpha}{2}}. \quad (6)$$

Weitere äquivalente Entscheidungskriterien sind in der Test-Box 4 angegeben.

Bei diesen Tests ist das Fehlerrisiko 1 Art nur *approximativ* gleich der Vorgabe  $\alpha$ , wobei die Approximation für wachsende Stichprobenumfänge  $n_X$  und  $n_Y$  immer besser wird. Test mit *exaktem* Fehlerrisiko  $\alpha$  lassen sich – ebenso wie exakte Konfidenzgrenzen – nur unter zusätzlichen Annahmen an die Verteilung von  $X$  und  $Y$  (vgl. unten) oder unter Verwendung anderer Verfahren (z.B. *Rangtests*) herleiten.

### **Exakte $t$ -Tests bei Normalverteilungen mit homogenen Varianzen**

Wir gehen jetzt wie schon in 3.5.1 davon aus, daß sowohl  $X$  als auch  $Y$  *normalverteilt* sind mit *homogenen Varianzen* bzw.  $\sigma_X = \sigma_Y$ . Für diesen Fall haben wir bereits in 3.5.1 (11) eine andere Schätzung  $\hat{\sigma}_{\Delta H}$  für die Standardabweichung der Mittelwert-Differenz angegeben. Mit dieser Schätzung – statt  $\hat{\sigma}_\Delta$  – erhält man folgenden Testwert

$$t = \frac{\Delta\hat{\mu}}{\hat{\sigma}_{\Delta H}} = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\hat{\sigma}_{\Delta H}} = \frac{\bar{x} - \bar{y}}{\hat{\sigma}_{\Delta H}} \quad (\text{Testwert bei homogenen Varianzen}). \quad (7)$$

Bei Verwendung dieses Testwerts ist das Fehlerrisiko 1. Art der obigen  $t$ -Tests im vorliegenden Spezialfall sogar *exakt* gleich  $\alpha$ . Die Details sind in der Test-Box 4 zusammengestellt.

### **Exakter vs. asymptotischer t-Test**

Die Testwerte des asymptotischen und exakten  $t$ -Tests unterscheiden sich nur durch die unterschiedlichen Schätzungen  $\hat{\sigma}_\Delta$  und  $\hat{\sigma}_{\Delta H}$  im Nenner. Wie schon in **3.5.1** festgestellt wurde, weichen diese Schätzungen jedoch typischerweise nur gering voneinander ab (insbesondere bei höheren Stichprobenumfängen) und stimmen bei gleichen Stichprobenumfängen  $n_X = n_Y$  (*balanciertes Design*) sogar genau überein. Insbesondere stimmt der *exakte* Test im *balancierten* Design mit dem *asymptotischen* Test überein. – Im *nicht-balancierten* Design sollte man den *asymptotischen* Test nur bei *nicht zu kleinen* Stichprobenumfängen  $n_X$  und  $n_Y$  anwenden (Faustregel: beide mindestens 10). Sicherheitsshalber kann man zusätzlich auch den *exakten t-Test* verwenden, der aber (bis auf extreme Ausnahmen) typischerweise zur gleichen Entscheidung führen wird. Bei *kleinen* Stichprobenumfängen (z.B. weniger als 10) sollte man den *exakten* Test durchführen, sofern nicht berechtigte Zweifel an der Homogenität der Standardabweichungen oder der Normalverteilungsannahme bestehen. Um ganz sicher zu gehen, kann man auch auf einen verteilungsfreien exakten *Rangtest* (vgl. z.B. Sachs 1999) ausweichen, worauf hier aber nicht eingegangen wird.

**Geburtsgewicht und Geschlecht (Fortsetzung):** Anhand der *StatLab-Auswahl 1985* soll zum Niveau  $\alpha = 5\%$  überprüft werden, ob es einen *geschlechtspezifischen Unterschied* beim Geburtsgewicht gibt. Die *zugehörigen* zweiseitigen Hypothesen lauten:

$$\begin{array}{ll} \text{Nullhypothese} & H_0: \mu_X = \mu_Y \quad (\text{kein geschlechtspezifischer Unterschied}), \\ \text{Hypothese} & H: \mu_X \neq \mu_Y \quad (\text{geschlechtspezifischer Unterschied vorhanden}). \end{array}$$

Der (*asymptotische*)  $t$ -Test mit  $m = 98$  Freiheitsgraden lehnt die Nullhypothese *ab*, weil

$$\begin{array}{lllll} \text{Testwert} & |t| & = & 2,227 & \geq 1,984 = t_{98, \frac{\alpha}{2}} \text{ bzw.} \\ \text{P-Wert} & 2\Phi_{98}(-|t|) & = & 2,82\% & \leq 5\% = \alpha \text{ bzw.} \\ \text{untere Konfidenzgrenze} & \Delta\hat{\mu}_{un,\alpha/2} & = & 0,056 & \geq 0. \end{array}$$

Zur Illustration haben wir alle *drei* äquivalenten Entscheidungskriterien angegeben, obwohl es reicht, *eine* dieser Bedingungen zu prüfen. – Damit ist die Hypothese  $H$  (*geschlechtspezifischer Unterschied*) mit dem Fehlerrisiko 5% *abgesichert*: das männliche Geburtsgewicht ist im Mittel größer als das weibliche. Man beachte, daß wegen der gleichen Stichprobenumfänge  $n_X = n_Y = 50$  der *exakte t-Test* hier mit dem *asymptotischen* übereinstimmt. – Den Unterschied  $\Delta\mu$  der (erwarteten) Geburtsgewichte haben wir zusammen mit Konfidenzgrenzen bereits in **3.5.1** angegeben.  $\square$

**Test-Box 4:** Vergleich der Erwartungswerte  $\mu_X$  und  $\mu_Y$  von zwei *unabhängigen* Zufallsvariablen  $X$  und  $Y$  mit Standardabweichungen  $\sigma_X$  und  $\sigma_Y$ .

<i>einseitiger Test</i>	<i>zweiseitiger Test</i>
$H_0: \mu_X \leq \mu_Y \quad \text{vs.} \quad H: \mu_X > \mu_Y$	$H_0: \mu_X = \mu_Y \quad \text{vs.} \quad H: \mu_X \neq \mu_Y$
Fehlerrisiko 1. Art: $\alpha$	<i>(Testniveau)</i>
Stichprobenumfänge: $n_X$ bzw. $n_Y$	<i>(x- bzw. y-Werte)</i>
Beobachtete Stichproben: $x_1, x_2, \dots, x_{n_X}$ $y_1, y_2, \dots, y_{n_Y}$	<i>(unabhängige Wiederholungen von X)</i> <i>(unabhängige Wiederholungen von Y)</i>
Schätzung von $\mu_X$ und $\mu_Y$ : $\hat{\mu}_X = \bar{x}$ $\hat{\mu}_Y = \bar{y}$	<i>(Mittelwerte: x- bzw. y-Werte)</i>
Schätzung von $\sigma_X$ und $\sigma_Y$ : $\hat{\sigma}_X = s_x$ $\hat{\sigma}_Y = s_y$	<i>(empirische Standardabweichung: x- bzw. y-Werte, vgl. Rechen-Box 1)</i>

**Asymptotischer t-Test:** Fehlerrisiko 1. Art =  $\alpha$

Anwendbar, wenn  $n_X$  und  $n_Y$  nicht zu klein (beide mindestens 10)

Testwert:	$t = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\hat{\sigma}_\Delta}$	mit	$\hat{\sigma}_\Delta = \sqrt{\frac{1}{n_X} \hat{\sigma}_X^2 + \frac{1}{n_Y} \hat{\sigma}_Y^2}$
Freiheitsgrad:	$m = n_X + n_Y - 2$		

Entscheidung: Ablehnung von  $H_0$  (Annahme von  $H$ ), falls gilt

<i>einseitiger Test</i>	<i>zweiseitiger Test</i>
Testwert $t \geq t_{m;\alpha}$ bzw.	Testwert $ t  \geq t_{m;\frac{\alpha}{2}}$ bzw.
P-Wert $\Phi_m(-t) \leq \alpha$ bzw.	P-Wert $2\Phi_m(- t ) \leq \alpha$ bzw.
untere Grenze $\Delta\hat{\mu}_{un,\frac{\alpha}{2}} \geq 0$	$\Delta\hat{\mu}_{un,\frac{\alpha}{2}} \geq 0$ oder $\Delta\hat{\mu}_{ob,\frac{\alpha}{2}} \leq 0$

**Exakter t-Test:** Fehlerrisiko 1. Art =  $\alpha$

Anwendbar, für **normalverteilte**  $X, Y$  mit **homogenen** Standardabweichungen  $\sigma_X = \sigma_Y$

Testwert: $t = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\hat{\sigma}_{\Delta H}}$	mit $\hat{\sigma}_{\Delta H} = \sqrt{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right) \frac{(n_X - 1)\hat{\sigma}_X^2 + (n_Y - 1)\hat{\sigma}_Y^2}{m}}$
---	---

Entscheidung: Ablehnung von  $H_0$  (Annahme von  $H$ ), falls gilt

<i>einseitiger Test</i>	<i>zweiseitiger Test</i>
$t \geq t_{m;\alpha}$ bzw. $\Phi_m(-t) \leq \alpha$	$ t  \geq t_{m;\frac{\alpha}{2}}$ bzw. $2\Phi_m(- t ) \leq \alpha$

### 3.5.3 Fehlerrisiko 2. Art und Testschärfe (Power)

Wie wollen jetzt wieder das Risiko für einen Fehler 2. Art für relevante Werte aus der Alternative berechnen. Beim *einseitigen* Testproblem

$$H_0: \mu_X \leq \mu_Y \quad \text{vs.} \quad H: \mu_X > \mu_Y \quad (1)$$

hängen das Fehlerrisiko 2. Art  $\beta = \beta(\mu_X, \mu_Y)$  und die dazu komplementäre Testschärfe  $\gamma = \gamma(\mu_X, \mu_Y)$  sowohl von den konkreten Werten  $\mu_X > \mu_Y$  aus der Alternative als auch von den Standardabweichungen  $\sigma_X$  und  $\sigma_Y$  ab. Je *kleiner* der Unterschied  $\Delta\mu = \mu_X - \mu_Y$  ist und je *größer* die Standardabweichungen sind, desto größer ist das Fehlerrisiko  $\beta(\mu_X, \mu_Y)$ . Die Berechnung von  $\beta(\mu_X, \mu_Y)$  basiert auf der *approximativen* Normalverteilung der Test-Variablen und die erforderlichen Rechenschritte sind in der *Power-Box 4* zusammengestellt. Man beachte, daß das Fehlerrisiko  $\beta(\mu_X, \mu_Y)$  nur noch über den (absoluten) *Unterschied*  $\Delta = |\mu_X - \mu_Y|$  von den beiden Erwartungswerten  $\mu_X$  und  $\mu_Y$  abhängt.

Für das *zweiseitige* Testproblem ergibt sich die Schärfe aus dem zugehörigen einseitigen Test zum *halben* Niveau. Die Details sind in der *Power-Box 4* angegeben.

<b>Power-Box 4:</b> Für Test der Erwartungswerte $\mu_X$ und $\mu_Y$ von zwei <i>unabhängigen</i> Zufallsvariablen $X$ und $Y$ mit Standardabweichungen $\sigma_X$ und $\sigma_Y$ .	
<i>einseitiger Test</i>	<i>zweiseitiger Test</i>
$H_0: \mu_X \leq \mu_Y \quad \text{vs.} \quad H: \mu_X > \mu_Y$	$H_0: \mu_X = \mu_Y \quad \text{vs.} \quad H: \mu_X \neq \mu_Y$
Fehlerrisiko 1. Art: Stichprobenumfänge: Freiheitsgrad: Wertpaare aus der <i>Alternative</i> $H$ :	$\alpha$ $n_X, n_Y$ $m = n_X + n_Y - 2$ $\mu_X > \mu_Y$
<i>einseitiger Test</i>	<i>zweiseitiger Test</i>
$t_m = t_{m;\alpha}$	$t_m = t_{m;\frac{\alpha}{2}}$
Hilfsgrößen:	$\Delta =  \mu_X - \mu_Y  > 0$ <i>Abweichung von <math>H_0</math></i>
	$\sigma_\Delta = \sqrt{\frac{1}{n_X} \sigma_X^2 + \frac{1}{n_Y} \sigma_Y^2}$ <i>(Standardabweichung von <math>\bar{x} - \bar{y}</math>)</i>
	$u = t_m - \frac{\Delta}{\sigma_\Delta}$
Fehlerrisiko 2. Art: Testschärfe (Power):	$\beta(\mu_X, \mu_Y) \approx \Phi(u)$ <i>mit <math>\Phi = Verteilungsfunktion</math></i> $\gamma(\mu_X, \mu_Y) \approx \Phi(-u)$ <i>von <math>N(0, 1)</math></i>

**Geburtsgröße und Geschlecht:** Anhand der *StatLab-Auswahl 1985* soll zum Niveau  $\alpha = 5\%$  überprüft werden, ob es einen *geschlechtspezifischen Unterschied* bei der Geburtsgröße gibt. Bezeichnet  $X$  die Geburtsgröße [in Zoll] der Jungen und  $Y$  die der Mädchen, so lauten die zweiseitigen Hypothesen:

$$\begin{array}{lll} \text{Nullhypothese} & H_0: \mu_X = \mu_Y & (\text{kein geschlechtspezifischer Unterschied}), \\ \text{Hypothese} & H: \mu_X \neq \mu_Y & (\text{geschlechtspezifischer Unterschied vorhanden}). \end{array}$$

Aus den Daten

$$\begin{array}{lll} \text{Jungen:} & n_X = 50 & \hat{\mu}_X = \bar{x} = 20,548 \\ \text{Mädchen:} & n_Y = 50 & \hat{\mu}_Y = \bar{y} = 20,128 \end{array} \quad \begin{array}{ll} \hat{\sigma}_X = s_x = 0,9509 \\ \hat{\sigma}_Y = s_y = 1,1693 \end{array}$$

ergibt sich der geschätzte Unterschied der (erwarteten) Geburtsgröße:

$$\Delta\hat{\mu} = \hat{\mu}_X - \hat{\mu}_Y = 0,420 \text{ [Zoll]} \quad \text{mit} \quad \hat{\sigma}_\Delta = 0,2131 \text{ [Zoll].}$$

Der (asymptotische)  $t$ -Test mit  $m = 98$  Freiheitsgraden lehnt die Nullhypothese *nicht ab*, weil

$$\begin{array}{lll} \text{Testwert} & |t| & = 1,971 < 1,984 = t_{98, \frac{\alpha}{2}} \\ \text{P-Wert} & 2\Phi_{98}(-|t|) & = 5,16\% > 5\% = \alpha. \end{array} \quad \text{bzw.}$$

Man beachte, daß wegen der gleichen Stichprobenumfänge  $n_X = n_Y = 50$  der *exakte t-Test* hier mit dem *asymptotischen* übereinstimmt.

*Analyse des Fehlerrisikos 2. Art:* Da hier ein Fehler 2. Art vorliegen kann, ist die Nullhypothese zunächst nur mit dem unbekannten Fehlerrisiko 2. Art  $\beta$  „abgesichert“, das wir jetzt exemplarisch analysieren wollen. Falls die unbekannten Standardabweichungen den geschätzten entsprechen – d.h.  $\sigma_X = 0,9509$  und  $\sigma_Y = 1,1693$  – so ergibt sich z.B. für den Unterschied  $\Delta = |\mu_X - \mu_Y| = 0,5$  (der etwas über dem beobachteten Unterschied  $\Delta\hat{\mu} = 0,42$  liegt) noch ein relativ hohes Fehlerrisiko

$$\beta \approx \Phi(-0,3614) \approx 36\% \quad \text{wobei} \quad \sigma_\Delta = 0,2131.$$

□

### 3.5.4 Versuchsplanung erforderlicher Stichprobenumfang

Es sei wieder  $\mu_X > \mu_Y$  ein für die konkrete Anwendung *relevantes* Wertepaar aus der Alternative und  $\Delta = |\mu_X - \mu_Y| > 0$  die zugehörige Abweichung von der Nullhypothese. Dann fällt das Fehlerrisiko  $\beta(\mu_X, \mu_Y)$  bzw. die Schärfe  $\gamma(\mu_X, \mu_Y)$  wächst, wenn man die Stichprobenumfänge  $n_X$  und  $n_Y$  erhöht. Im Rahmen einer *Versuchsplanung* kann man wieder für einen vorgegebenen Wert  $\beta$  die erforderlichen Mindestumfänge  $n_X$  und  $n_Y$  bestimmen, bei denen  $\beta(\mu_X, \mu_Y)$  höchstens gleich der Vorgabe  $\beta$  ist. Hierbei ist es (aus statistischer Sicht) optimal, gleiche Stichprobenumfänge  $n_X = n_Y$  zu verwenden (*balanciertes Design*).

Die erforderlichen Rechenschritte sind in der *Planungs-Box 4* zusammengestellt. Da diese Berechnungen auf vereinfachter Approximation beruhen, sollte man die resultierenden Stichprobenumfänge sicherheitshalber um 1 erhöhen.

Typischerweise verwendet man für die Vorgabe  $\beta$  wieder das *doppelte* Fehlerrisiko 1. Art, also  $\beta = 2\alpha$  (z.B.  $\beta = 10\%$  bei  $\alpha = 5\%$ ).

<b>Planungs-Box 4:</b> Für Tests zum Vergleich der Erwartungswerte $\mu_X$ und $\mu_Y$ von zwei <i>unabhängigen</i> Zufallsvariablen $X, Y$ mit Standardabweichungen $\sigma_X, \sigma_Y$	
<i>einseitiger Test</i>	<i>zweiseitiger Test</i>
$H_0: \mu_X \leq \mu_Y \quad \text{vs.} \quad H: \mu_X > \mu_Y$	$H_0: \mu_X = \mu_Y \quad \text{vs.} \quad H: \mu_X \neq \mu_Y$
<b>Vorgaben:</b>	<p>Fehlerrisiko 1. Art (Testniveau): <math>\alpha</math></p> <p>Wertepaar aus der <i>Alternative</i> <math>H: \mu_X &gt; \mu_Y</math></p> <p>Fehlerrisiko 2. Art für <math>\mu_X &gt; \mu_Y</math>: <math>\beta</math> bzw.</p> <p>Schärfe für <math>\mu_X &gt; \mu_Y</math>: <math>\gamma = 1 - \beta</math></p>
<b>Hilfsgrößen:</b>	$\Delta =  \mu_X - \mu_Y  > 0$ ( <i>Abweichung von <math>H_0</math></i> ) $\bar{\sigma} = \sqrt{\frac{1}{2}(\sigma_X^2 + \sigma_Y^2)}$ ( <i>mittlere Standardabweichung</i> )
<i>einseitiger Test</i>	<i>zweiseitiger Test</i>
$z = z_\alpha$	$z = z_{\frac{\alpha}{2}}$
<b>Mindestumfang</b>	
pro Gruppe: $n_X = n_Y \approx 2 \left[ \frac{(z + z_\beta) \cdot \bar{\sigma}}{\Delta} \right]^2$	(aufzurunden auf ganze Zahl, sicherheitshalber um 1 erhöhen)
insgesamt: $n = n_X + n_Y = 2 n_X$	

**GeburtsgröÙe und Geschlecht (Fortsetzung):** Der erforderliche Mindestumfang  $n_X = n_Y$ , bei dem der zweiseitige Test zum Niveau  $\alpha = 5\%$  für den Unterschied  $\Delta = 0,5$  ein Fehlerrisiko 2. Art von (höchstens)  $\beta = 10\%$  hat, ergibt sich wie folgt:

$$z_\alpha = 1.960 \quad z_\beta = 1.282 \quad \Rightarrow \quad n_X = n_Y \approx 95.5.$$

Aufgerundet und (sicherheitshalber) um 1 erhöht erhält man  $n_X = n_Y = 97$ .

**Vergleich von zwei Meßverfahren:** Es soll zum Niveau  $\alpha = 5\%$  überprüft werden, ob zwei verschiedene Methoden zur Bestimmung des Bleigehalts [in mg/l] im Apfelsaft zu unterschiedlichen Ergebnissen führen (vgl. auch *Kinder-Osius-Timm (1982), Beispiel 7.19*). Die zweiseitigen Hypothesen lauten

Nullhypothese  $H_0: \mu_X = \mu_Y$  (kein Unterschied beider Methoden),

Hypothese  $H: \mu_X \neq \mu_Y$  (systematischer Unterschied beider Methoden).

Bei jeweils  $n_X = 25$  und  $n_Y = 20$  unabhängigen Wiederholungs-Messungen an einer Apfelsaftprobe ergaben sich für die insgesamt  $n = 45$  Meßwerte:

$$\begin{array}{ll} \hat{\mu}_X = \bar{x} = 0,520 & \hat{\sigma}_X = s_x = 0,04713 \\ \hat{\mu}_Y = \bar{y} = 0,545 & \hat{\sigma}_Y = s_y = 0,03987. \end{array}$$

Aufgrund von Vorinformationen kann davon ausgegangen werden, daß  $X$  und  $Y$  die Voraussetzungen für den exakten  $t$ -Test (Normalverteilung von  $X$  und  $Y$  mit homogenen Standardabweichungen) erfüllen. Deshalb wird der *exakte t-Test* angewandt. Der geschätzte Unterschied der Erwartungswerte ist

$$\Delta\hat{\mu} = \hat{\mu}_X - \hat{\mu}_Y = -0,0250 \text{ [mg/l]} \quad \text{mit} \quad \hat{\sigma}_{\Delta H} = 0,0132 \text{ [mg/l].}$$

Der *exakte t-Test* mit  $m = 43$  Freiheitsgraden lehnt die Nullhypothese *nicht ab*, weil

$$\begin{array}{lll} \text{Testwert} & |t| & = 1,891 < 2,017 = t_{43, \frac{\alpha}{2}} \\ \text{P-Wert} & 2\Phi_{98}(-|t|) & = 6,54\% > 5\% = \alpha. \end{array}$$

Der *asymptotische t-Test* kommt übrigens zum gleichen Ergebnis: hier ist  $|t| = 1,927$  und  $\hat{\sigma}_{\Delta} = 0,0130$ .

*Analyse des Fehlerrisikos 2. Art:* Da ein Fehler 2. Art vorliegen kann, ist die Nullhypothese zunächst nur mit dem unbekannten Fehlerrisiko 2. Art  $\beta$  „abgesichert“, das hier exemplarisch analysiert werden soll. Hierbei gehen wir davon aus, daß die unbekannten Standardabweichungen den geschätzten entsprechen, d.h.  $\sigma_X = 0,04713$  und  $\sigma_Y = 0,03987$ . Für den Unterschied  $\Delta = |\mu_X - \mu_Y| = 0,05 \text{ [mg/l]}$  (der einer Abweichung von ca. 10% der beobachteten Mittelwerte entspricht) ergibt sich das geringe Fehlerrisiko:

$$\beta \approx \Phi(-1,837) \approx 3,3\% \quad \text{wobei} \quad \sigma_{\Delta} = 0,0130.$$

*Versuchsplanung:* Will man jedoch den (geringeren) Unterschied von  $\Delta = 0,03 \text{ [mg/l]}$  beim zweiseitigen Test zum Niveau  $\alpha = 5\%$  mit einem Fehlerrisiko 2. Art von (höchstens)  $\beta = 10\%$  entdecken, so ergeben sich die erforderlichen Umfänge beim balancierten Design wie folgt

$$z_{\alpha} = 1,960 \quad z_{\beta} = 1,282 \quad \Rightarrow \quad n_X = n_Y \approx 44,5.$$

Aufgerundet und (sicherheitshalber) um 1 erhöht erhält man  $n_X = n_Y = 46$ .  $\square$

### 3.6 Lineare Regressionsanalyse

Eine Regressionsanalyse untersucht die *Abhangigkeit* einer Zufallsvariablen  $Y$  (*Zielvariable*) von einer *Einflusvariablen*  $X$ . Fur einen *vorgegebenen* Wert  $x$  von  $X$  ist der zugehorige *Erwartungswert* von  $Y$  eine Funktion von  $x$ , die mit  $f(x) := E\{Y | X = x\}$  bezeichnet wird. Diese sogenannte *Regressionsfunktion* soll *modelliert* und dann *geschatzt* werden.

**Lineares Regressionsmodell:** Hier wird die Regressionsfunktion durch eine *lineare* Funktion modelliert

$$f(x) = a + bx \quad (\text{lineares Regressionsmodell}) . \quad (1)$$

Unter Verwendung der Abweichung  $\varepsilon = Y - f(x)$  erhalt man die Darstellung

$$Y = f(x) + \varepsilon, \quad (2)$$

wobei der Erwartungswert der (zuflligen) Abweichung  $\varepsilon$  Null ist:  $E(\varepsilon) = 0$ .

**Beobachtete Daten:** Die typische Form der Datenerhebung besteht darin, da man sich  $n$  Werte  $x_1, x_2, \dots, x_n$  von  $X$  *vorgibt* (die nicht *paarweise verschieden* sein mussen) und dann die Realisierungen  $y_1, y_2, \dots, y_n$  der zugehorigen *voneinander unabhangigen* Zufallsvariablen  $Y_1, Y_2, \dots, Y_n$  *beobachtet*. Ein Beispiel hierfr (vgl. 1.5.7) sind die *beobachteten* Blutdruckwerte ( $y$ -Werte) fur die *vorgegebenen* Altersgruppen ( $x$ -Werte).

Eine andere Form der Datenerhebung liegt vor, wenn die *Paare*  $(x_1, y_1), \dots, (x_n, y_n)$  Realisierungen von *unabhangigen* Wiederholungen des *Paars*  $(X, Y)$  sind. Dies ist z.B. bei den StatLab-Daten in 1.5.5 der Fall mit den Geburtsgroen als  $x$ -Werte und den Geburtsgewichten als  $y$ -Werte. Obwohl in einer solchen Situation die  $x$ -Werte *nicht* fest vorgegeben sind, kann man die Regressionsanalyse trotzdem so durchfuhren, als ob sie *vorher fixiert* gewesen waren. Man spricht dann von einer *auf die beobachteten  $x$ -Werte bedingten* Analyse. Hierbei sind die  $x$ -Werte nicht notwendig alle unterschiedlich. Zur gemeinsamen Behandlung beider Datenerhebungen gehen wir bei der folgenden Analyse von  $x$ -Werten  $x_1, x_2, \dots, x_n$  aus, unter denen sich mindestens *zwei verschiedene*  $x$ -Werte befinden mussen, und betrachten sie als *vorgegeben*.

#### Aufgaben der Regressionsanalyse

- Berechnung von *Schatzungen*  $\hat{a}$  und  $\hat{b}$  fur die Parameter  $a$  und  $b$ .
- *Testen* von Hypothesen uber den Anstiegsparameter  $b$ .
- Angabe von *Konfidenzgrenzen* fur den Anstiegsparameter  $b$ .
- *Prognose* von  $f(x_0)$  fur einen interessierenden Wert  $x_0$ , d.h. Schatzen von  $f(x_0)$  und Angabe von *Konfidenzgrenzen* fur  $f(x_0)$ .
- *Uberprufen* des linearen Regressionsmodells.

### 3.6.1 Ungewichtete Regression bei homogenen Varianzen

Wir betrachten zunächst den Fall, daß die voneinander unabhängigen Zufallsvariablen  $Y_1, Y_2, \dots, Y_n$  alle die *gleiche Varianz*  $\sigma_Y^2$  besitzen,

$$\sigma_Y^2 = \text{Var} (Y_1) = \text{Var} (Y_2) = \dots = \text{Var} (Y_n) \quad (\text{homogene Varianzen}),$$

wobei die gemeinsame Varianz  $\sigma_Y^2$  *unbekannt* ist und auch *geschätzt* werden soll.

#### 3.6.1 a Schätzung der Parameter und Modellüberprüfung

**Schätzung:** Als *Schätzungen* für  $a$  und  $b$  verwendet man die Koeffizienten der *Regressionsgeraden* der beobachteten  $(x, y)$ -Werte, d.h. (vgl. Abschnitt 1.4)

$$\hat{b} = \frac{Sxy}{Sxx} \quad \text{mit} \quad Sxy = \sum_i (x_i - \bar{x})(y_i - \bar{y}) \quad (1)$$

$$Sxx = \sum_i (x_i - \bar{x})^2$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \quad \text{bzw.} \quad \bar{y} = \hat{a} + \hat{b}\bar{x}. \quad (2)$$

Die Varianz  $\sigma_Y^2$  wird durch die mittlere Streuung der  $y$ -Werte um die Regressionsgerade wie folgt geschätzt

$$\hat{\sigma}_Y^2 = \frac{1}{m} \cdot SQ(\hat{a}, \hat{b}) = \frac{1}{m} \cdot \sum_i [y_i - (\hat{a} + \hat{b}x_i)]^2 \quad \text{mit} \quad m = n - 2. \quad (3)$$

Man beachte, daß hier durch den sogenannten *Freiheitsgrad*  $m$  und nicht durch die Anzahl  $n$  aller Punkte dividiert wird. Hierdurch erreicht man (analog zur Division durch  $n - 1$  bei der Varianz in Rechen-Box 1), daß die Schätzung „im Mittel“ um die unbekannte Varianz  $\sigma_Y^2$  streut (d.h.  $\sigma_Y^2$  als *Erwartungswert* hat). Rein formal ergibt sich der *Freiheitsgrad* als Umfang  $n$  *reduziert* um die Anzahl der geschätzten Parameter der Regressionsfunktion, also hier *zwei* (für  $\hat{a}$  und  $\hat{b}$ ).

Die für (3) erforderliche Streuung  $SQ(\hat{a}, \hat{b})$  um die Regressionsgerade kann berechnet werden durch

$$\begin{aligned} SQ(\hat{a}, \hat{b}) &= \sum_i [y_i - (\hat{a} + \hat{b}x_i)]^2 \\ &= Syy - \hat{b} \cdot Sxy \end{aligned} \quad \text{mit} \quad Syy = \sum_i (y_i - \bar{y})^2. \quad (4)$$

Wenn *alle* Punkte auf der Regressionsgeraden liegen, so ist  $SQ(\hat{a}, \hat{b}) = 0$  und somit auch  $\hat{\sigma}_Y = 0$ . In diesem (extrem selten auftretenden) Fall, kann zwar die Regressionsgerade geschätzt werden, aber weitere statistische Analysen (Konfidenzgrenzen, Tests) sind nicht

möglich oder nicht sinnvoll. Bei den folgenden Betrachtungen gehen wir daher davon aus, daß  $SQ(\hat{a}, \hat{b}) > 0$  und somit  $\hat{\sigma}_Y > 0$  gilt.

**Konfidenzgrenzen für den Anstieg:** Die Konstruktion von Konfidenzgrenzen für den Anstiegs-Parameter  $b$  basiert auf der Standardabweichung der Schätzung  $\hat{b}$

$$\sigma_{\hat{b}} = \frac{\sigma_Y}{\sqrt{Sxx}}, \quad (5)$$

die – durch Ersetzen von  $\sigma_Y$  durch  $\hat{\sigma}_Y$  – geschätzt werden kann durch

$$\hat{\sigma}_{\hat{b}} = \frac{\hat{\sigma}_Y}{\sqrt{Sxx}}. \quad (6)$$

Die einseitigen Konfidenzgrenzen zur approximativen Sicherheit  $1 - \alpha$  ergeben sich jetzt (analog 3.5.1) unter Verwendung des  $\alpha$ -Quantils  $t_{m;\alpha}$  der  $t_m$ -Verteilung (vgl. Anhang T) mit dem Freiheitsgrad  $m = n - 2$  zu

$$\hat{b}_{un,\alpha} = \hat{b} - \hat{d}_\alpha \quad (\text{einseitige untere Konfidenzgrenze für } b), \quad (7)$$

$$\hat{b}_{ob,\alpha} = \hat{b} + \hat{d}_\alpha \quad (\text{einseitige obere Konfidenzgrenze für } b), \quad \text{mit} \quad (8)$$

$$\hat{d}_\alpha = t_{m;\alpha} \cdot \hat{\sigma}_{\hat{b}} \quad (\text{Bandbreite}). \quad (9)$$

Wenn die Zielvariable  $Y$  normalverteilt ist, so haben diese Konfidenzgrenzen sogar die exakte Sicherheit  $1 - \alpha$ . Das zweiseitige Konfidenzintervall ergibt sich wieder aus den einseitigen Grenzen zur halben Irrtums-Wahrscheinlichkeit:

$$(\hat{b}_{un,\frac{\alpha}{2}}, \hat{b}_{ob,\frac{\alpha}{2}}) \quad (\text{zweiseitiges Konfidenzintervall}). \quad (10)$$

Konfidenzgrenzen für den Parameter  $a$  werden erst in 3.6.1 d angegeben.

**Modellüberprüfung:** Zur Überprüfung des linearen Regressionsmodells stellt man die beobachteten  $(x, y)$ -Werte zusammen mit der Regressionsgeraden dar und beurteilt, ob die Punkte zufällig oder systematisch von der Geraden abweichen. Eine starke (aber zufällige) Streuung der  $y$ -Werte um die Gerade ist kein Indiz für die Ungültigkeit des linearen Regressionsmodells, sondern lediglich ein Hinweis auf eine große Varianz  $\sigma_Y^2$  von  $Y$ . Zur besseren Beurteilung der Abweichungen untersucht man im Rahmen einer Residuenanalyse die skalierten Residuen

$$r_i = \frac{y_i - \hat{f}(x_i)}{\hat{\sigma}_Y} = \frac{y_i - (\hat{a} + \hat{b} x_i)}{\hat{\sigma}_Y} \quad (\text{skaliertes Residuum}) \quad (11)$$

auf Zufälligkeit. Ein Residuenplot stellt die Punkte  $(x_i, r_i)$  dar (z.B. in 3.6.1 e, Abb. 1), und man beurteilt deren Streuung um die horizontale Gerade  $r = 0$  auf Zufälligkeit, d.h. grob (und vereinfachend) gesprochen sollten die Residuen sich wie unabhängige und

$N(0, 1)$ -verteilte Zufallszahlen verhalten. Insbesondere ist dann (nach 2.5.4) nur für rund 5% aller Residuen ein Absolutbetrag  $|r_i| \geq 2$  zu erwarten und sogar nur für 0,3% ein Absolutbetrag  $|r_i| \geq 3$ . Wenn im Residuenplot *systematische* Abweichungen oder *Trends* erkennbar sind, so spricht dies *gegen* das lineare Regressionsmodell.

### 3.6.1 b Testen des Anstiegs

**Test auf signifikanten Anstieg:** Die naheliegendste Frage ist, ob der  $x$ -Wert überhaupt einen (linearen) Einfluß auf (den Erwartungswert von)  $Y$  hat oder nicht, d.h. ob der Anstiegsparameter  $b$  von Null verschieden ist oder nicht. Dies führt zur Nullhypothese

$$H_0: b = 0 \quad (x \text{ hat } \textit{keinen} \text{ linearen Einfluß auf } Y) \quad (1)$$

mit den ein- oder zweiseitigen alternativen Hypothesen

$$\begin{aligned} H: b > 0 & \quad (\textit{positiver linearer Einfluß von } x \text{ auf } Y) \\ H: b < 0 & \quad (\textit{negativer linearer Einfluß von } x \text{ auf } Y) \\ H: b \neq 0 & \quad (x \text{ hat einen linearen } \textit{Einfluß} \text{ auf } Y) \end{aligned} \quad (2)$$

Es liegt nahe, die Entscheidung über diese Hypothesen davon abhängig zu machen, wie stark die Schätzung  $\hat{b}$  (unter Berücksichtigung ihrer Streuung) vom Wert 0 abweicht. Als *Testwert* verwendet man die *standardisierte* Schätzung von  $b$ , die sich auch als streng monoton wachsende Transformation des *Korrelationskoeffizienten*  $r_{xy}$  (vgl. 1.5.3) darstellen läßt:

$$t = \frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} = \frac{\hat{b} \cdot \sqrt{S_{xx}}}{\hat{\sigma}_Y} = \frac{r_{xy} \sqrt{m}}{\sqrt{1 - r_{xy}^2}} \quad (\textit{Testwert}). \quad (3)$$

Wenn die Zielvariable  $Y$  *normalverteilt* ist, so ist die Verteilung des Testwerts *unter der Nullhypothese*  $H_0: b = 0$  eine  $t$ -Verteilung mit  $m = n - 2$  Freiheitsgraden. Für *nicht-normalverteiltes*  $Y$  läßt sich die Verteilung des Testwerts unter der Nullhypothese zumindest durch eine  $t_m$ -Verteilung *approximieren*, wobei die Approximation für wachsenden Stichprobenumfang  $n$  immer genauer wird. Hieraus ergeben sich (analog zu Test-Box 4) die ein- bzw. zweiseitigen  $t$ -Tests in Test-Box 5a, wobei die jeweils letzte Formulierung für die Ablehnung von  $H_0$  nur den *Korrelationskoeffizienten*  $r_{xy}$  bzw. das *Bestimmtheitsmaß*  $B_{xy} = r_{xy}^2$  (vgl. 1.5.2) verwendet. – Ein weiteres Entscheidungskriterium unter Verwendung von *Konfidenzgrenzen* findet man in der *Konfidenz- und Test-Box 5b* für den Spezialfall  $b_0 = 0$ .

Bei den *einseitigen* Alternativen  $H: b > b_0$  bzw.  $H: b < b_0$  wird in der Test-Box 5a statt  $H_0: b = b_0$  die *erweiterte* Nullhypothese  $H_0: b \leq b_0$  bzw.  $H_0: b \geq b_0$  (also die *Negation*

der Alternative) verwendet, wodurch sich am Rechengang aber nichts ändert.

Die in **1.5.2 (5)** und **1.5.3 (5)** genannten *Faustregeln* entsprechen daher hier der Ablehnung der *zweiseitigen Nullhypothese* für  $t_{m;\alpha} \approx 2$  und somit  $\alpha \approx 5\%$  ab  $n \geq 8$ .

**Testen eines Referenz-Anstiegs:** Ein etwas allgemeinere *Nullhypothese* als (1) liegt vor, wenn man überprüfen will, ob der Anstieg  $b$  einen vorgegebenen *Referenzwert*  $b_0$  hat:

$$H_0: b = b_0 \quad (\text{Anstieg entspricht dem Referenz-Anstieg } b_0) \quad (4)$$

Dies entspricht der Nullhypothese (1) für  $b_0 = 0$ . Die ein- bzw. zweiseitigen alternativen Hypothesen lauten

$$H_0: b > b_0 \quad (\text{Anstieg ist } \textit{größer} \text{ als der Referenz-Anstieg } b_0) \quad (5)$$

$$H_0: b < b_0 \quad (\text{Anstieg ist } \textit{kleiner} \text{ als der Referenz-Anstieg } b_0)$$

$$H_0: b \neq b_0 \quad (\text{Anstieg ist } \textit{verschieden} \text{ vom Referenz-Anstieg } b_0)$$

Als *Testwert* verwendet man jetzt die standardisierte Abweichung der Schätzung  $\hat{b}$  vom Referenzwert  $b_0$

$$t = \frac{\hat{b} - b_0}{\hat{\sigma}_{\hat{b}}} = \frac{(\hat{b} - b_0) \cdot \sqrt{S_{xx}}}{\hat{\sigma}_Y} \quad (\text{Testwert}). \quad (6)$$

Im Spezialfall  $b_0 = 0$  reduzieren sich die Hypothesen (4)-(5) zu (1)-(2), und der Testwert (6) stimmt mit (3) überein. Man beachte aber, daß sich der Testwert (6) – im Gegensatz zu (3) – *nicht* als Funktion des Korrelationskoeffizienten schreiben läßt.

Die Testentscheidung über die Hypothesen (5) und (6) wird wie im obigen Spezialfall  $b_0 = 0$  durchgeführt (vgl. Konfidenz- und Test-Box 5b).

Bei den *einseitigen* Alternativen  $H: b > b_0$  bzw.  $H: b < b_0$  wird in der Test-Box 5a statt  $H_0: b = b_0$  die *erweiterte Nullhypothese*  $H_0: b \leq b_0$  bzw.  $H_0: b \geq b_0$  (also die *Negation* der Alternative) verwendet, wodurch sich am Rechengang aber nichts ändert.

<b>Test-Box 5a (ungewichtete Regression): Test auf signifikanten Anstieg</b>				
Stichprobenumfang: $n$		( $x, y$ )-Werte: $(x_1, y_1), \dots, (x_n, y_n)$		
Schätzungen (zu nicht erläuterten Größen vgl. Rechen-Box 2:)				
Regressionsgerade $y = \hat{a} + \hat{b}x$ :	$\hat{b} = S_{xy}/S_{xx}$ ,	$\hat{a} = \bar{y} - \hat{b}\bar{x}$		
Streuung um die Gerade:	$SQ(\hat{a}, \hat{b}) = S_{yy} - \hat{b} \cdot S_{xy}$			
Freiheitsgrad:	$m = n - 2$			
Standardabweichung der $y$ -Werte:	$\hat{\sigma}_Y = \sqrt{\frac{1}{m}SQ(\hat{a}, \hat{b})}$			
Standardabweichung von $\hat{b}$ :	$\hat{\sigma}_{\hat{b}} = \frac{\hat{\sigma}_Y}{\sqrt{S_{xx}}}$			
<i>einseitig oben</i>	<i>einseitig unten</i>	<i>zweiseitig</i>		
$H_0: b \leq 0$ vs. $H: b > 0$	$H_0: b \geq 0$ vs. $H: b < 0$	$H_0: b = 0$ vs. $H: b \neq 0$		
<b>Student's t-Test:</b>				
bei nicht zu kleinem $n$ (mindestens 10): Fehlerrisiko 1. Art $\approx \alpha$				
bei normalverteiltem $Y$ : Fehlerrisiko 1. Art = $\alpha$ (exakt)				
Testwert: $t = \frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} = \frac{\hat{b} \cdot \sqrt{S_{xx}}}{\hat{\sigma}_Y} = \frac{r_{xy}\sqrt{m}}{\sqrt{1 - r_{xy}^2}}$				
kritische Werte: $t_{m;\alpha}$ ( $\alpha$ -Quantil von $t_m$ ) für den Testwert				
$r_{m;\alpha} = \frac{t_{m;\alpha}}{\sqrt{m + t_{m;\alpha}^2}}$	für den Korrelationskoeffizienten			
$\Phi_m$ bezeichnet die Verteilungsfunktion der $t_m$ -Verteilung.				
Entscheidung: Ablehnung von $H_0$ (Annahme von $H$ ), falls gilt				
<i>einseitig oben</i>	<i>einseitig unten</i>	<i>zweiseitig</i>		
Testwert $t \geq t_{m;\alpha}$ bzw. $P$ -Wert $\Phi_m(-t) \leq \alpha$ bzw. $r_{xy} \geq r_{m;\alpha}$	Testwert $t \leq -t_{m;\alpha}$ bzw. $P$ -Wert $\Phi_m(t) \leq \alpha$ bzw. $r_{xy} \leq -r_{m;\alpha}$	Testwert $ t  \geq t_{m;\frac{\alpha}{2}}$ bzw. $P$ -Wert $2\Phi_m(- t ) \leq \alpha$ bzw. $ r_{xy}  \geq r_{m;\frac{\alpha}{2}}$ $B_{xy} \geq r_{m;\frac{\alpha}{2}}^2$		

Konfidenz- und Test-Box 5b ( <i>ungewichtete Regression</i> ):	Parameter $b$
Berechnungen (zu nicht erläuterten Größen vgl. <i>Rechen-Box 2, Test-Box 5a</i> ) :	
Freiheitsgrad:	$m = n - 2$
Standardabweichung der $y$ -Werte:	$\hat{\sigma}_Y = \sqrt{\frac{1}{m} SQ(\hat{a}, \hat{b})}$
Standartabweichung von $\hat{b}$ :	$\hat{\sigma}_{\hat{b}} = \frac{\hat{\sigma}_Y}{\sqrt{S_{xx}}}$

<b>Einseitige Konfidenzgrenzen für <math>b</math>:</b>							
bei nicht zu kleinem $n$ (mindestens 10):	Sicherheit $\approx 1 - \alpha$						
bei normalverteiltem $Y$ :	Sicherheit $= 1 - \alpha$ (exakt)						
<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center; padding: 5px;">einseitige untere Grenze</td> <td style="text-align: center; padding: 5px;">einseitige obere Grenze</td> </tr> <tr> <td style="text-align: center; padding: 5px;"><math>\hat{b}_{un,\alpha} = \hat{b} - \hat{d}_{\alpha}</math></td> <td style="text-align: center; padding: 5px;"><math>\hat{b}_{ob,\alpha} = \hat{b} + \hat{d}_{\alpha}</math></td> </tr> <tr> <td colspan="2" style="text-align: center; padding: 5px;"><math>\hat{d}_{\alpha} = t_{m;\alpha} \cdot \hat{\sigma}_{\hat{b}}</math> (Bandbreite)</td></tr> </table>		einseitige untere Grenze	einseitige obere Grenze	$\hat{b}_{un,\alpha} = \hat{b} - \hat{d}_{\alpha}$	$\hat{b}_{ob,\alpha} = \hat{b} + \hat{d}_{\alpha}$	$\hat{d}_{\alpha} = t_{m;\alpha} \cdot \hat{\sigma}_{\hat{b}}$ (Bandbreite)	
einseitige untere Grenze	einseitige obere Grenze						
$\hat{b}_{un,\alpha} = \hat{b} - \hat{d}_{\alpha}$	$\hat{b}_{ob,\alpha} = \hat{b} + \hat{d}_{\alpha}$						
$\hat{d}_{\alpha} = t_{m;\alpha} \cdot \hat{\sigma}_{\hat{b}}$ (Bandbreite)							

<b>Student's t-Test:</b>		
bei nicht zu kleinem $n$ (mindestens 10):		Fehlerrisiko 1. Art $\approx \alpha$
bei normalverteiltem $Y$ :		Fehlerrisiko 1. Art $= \alpha$ (exakt)
einseitig oben	einseitig unten	zweiseitig
$H_0: b \leq b_0$ vs. $H: b > b_0$	$H_0: b \geq b_0$ vs. $H: b < b_0$	$H_0: b = b_0$ vs. $H: b \neq b_0$
Testwert: $t = \frac{\hat{b} - b_0}{\hat{\sigma}_{\hat{b}}} = \frac{(\hat{b} - b_0) \cdot \sqrt{S_{xx}}}{\hat{\sigma}_Y}$		
Entscheidung: Ablehnung von $H_0$ (Annahme von $H$ ), falls gilt		
einseitig oben	einseitig unten	zweiseitig
Testwert $t \geq t_{m;\alpha}$ bzw. $P$ -Wert $\Phi_m(-t) \leq \alpha$ bzw. $b_0 \leq \hat{b}_{un,\alpha}$	Testwert $t \leq -t_{m;\alpha}$ bzw. $P$ -Wert $\Phi_m(t) \leq \alpha$ bzw. $\hat{b}_{ob,\alpha} \leq b_0$	Testwert $ t  \geq t_{m;\frac{\alpha}{2}}$ bzw. $P$ -Wert $2\Phi_m(- t ) \leq \alpha$ bzw. $b_0 \leq \hat{b}_{un,\frac{\alpha}{2}}$ oder $\hat{b}_{ob,\frac{\alpha}{2}} \leq b_0$

### 3.6.1 c Fehlerrisiko 2. Art, Testschärfe und Versuchsplanung

**Fehlerrisiko 2. Art und Testschärfe:** Wir wollen zunächst das Fehlerrisiko 2. Art für den *einseitigen* Test auf einen *Referenzanstieg*  $b_0$  bestimmen, d.h. für

$$H_0: b \leq b_0 \quad \text{vs.} \quad H: b > b_0. \quad (1)$$

Der Test auf *signifikanten Anstieg* wird hier nicht separat betrachtet, weil er als Spezialfall für  $b_0 = 0$  bereits im allgemeineren Testproblem (1) enthalten ist. Das Fehlerrisiko 2. Art  $\beta = \beta(b_1)$  und die dazu komplementäre Testschärfe  $\gamma = \gamma(b_1)$  hängen einerseits vom konkreten Wert  $b_1 > b_0$  aus der Alternative und andererseits sowohl von der (gemeinsamen) Standardabweichung  $\sigma_Y$  als auch von der Quadratsumme  $Sxx$  (und somit von der *Streuung* der  $x$ -Werte) ab. Je *kleiner* der Unterschied  $b_1 - b_0$  ist und je *größer* die Standardabweichung  $\sigma_Y$  ist, desto größer ist das Fehlerrisiko  $\beta(b_1)$ . Der Einfluß von  $Sxx$  auf das Fehlerrisiko  $\beta(b_1)$  ist dagegen genau umgekehrt: je *größer*  $Sxx$  ist, desto *kleiner* ist  $\beta(b_1)$ . Aus der Darstellung

$$Sxx = (n - 1) \cdot s_x^2 \quad \text{mit} \quad s_x = \text{empirische Standardabweichung } x\text{-Werte} \quad (2)$$

ergibt sich daher, daß das Fehlerrisiko  $\beta(b_1)$  fällt wenn entweder der Umfang  $n$  oder die empirische Standardabweichung  $s_x$  anwachsen.

Die in der *Power-Box 5* zusammengestellte Berechnung von  $\beta(b_1)$  basiert auf der *approximativen* Normalverteilung der Test-Variablen. Für eine konkrete Bestimmung von  $\beta(b_1)$  wird noch die *unbekannte* Standardabweichung  $\sigma_Y$  benötigt, für die man typischerweise ihre *Schätzung*  $\hat{\sigma}_Y$  verwendet. – Für das andere *einseitige* sowie das *zweiseitige* Testproblem ergeben sich das Fehlerrisiko 2. Art und die Schärfe wieder ganz analog (vgl. *Power-Box 5*).

**Versuchsplanung:** Die Versuchsplanung ist hier aufwendiger als bei den bisherigen Tests, weil neben dem Umfang  $n$  auch die Platzierung der  $x$ -Werte im interessierenden Bereich optimal zu wählen ist. Will man, daß für eine *vorgegebene Abweichung*  $\Delta = |b_1 - b_0|$  das Fehlerrisiko 2. Art (höchstens) einen vorgegebenen Wert  $\beta$  erreicht, so ermittelt man zunächst (approximativ) die zugehörige

erforderliche Quadratsumme  $Sxx$

$$Sxx \approx \left[ \frac{(z + z_\beta) \cdot \sigma_Y}{\Delta} \right]^2 \quad \text{mit} \quad z = \begin{cases} z_\alpha & \text{beim einseitigen Test} \\ z_{\alpha/2} & \text{beim zweiseitigen Test} \end{cases} \quad (3)$$

Um diese erforderliche Quadratsumme  $Sxx$  durch eine geeignete Wahl der  $x$ -Werte (mindestens) zu erreichen, gibt es verschiedene Möglichkeiten, von denen wir nur das einfachste Design (äquidistante  $x$ -Werte) besprechen wollen. Danach behandeln wir noch den Fall, daß die  $x$ -Werte nicht *vorgegeben*, sondern *zufällig* sind.

**Power-Box 5a:** (ungewichtete Regression):

Für Tests über den *Anstiegs-Parameter*  $b$  der Regressionsgeraden

<i>einseitig oben</i>	<i>einseitig unten</i>	<i>zweiseitig</i>
$H_0: b \leq b_0$ <i>vs.</i> $H: b > b_0$	$H_0: b \geq b_0$ <i>vs.</i> $H: b < b_0$	$H_0: b = b_0$ <i>vs.</i> $H: b \neq b_0$
Fehlerrisiko 1. Art (Testniveau): $\alpha$		Standardabweichung: $\sigma_Y$
Stichprobenumfang: $n$		Freiheitsgrad: $m = n - 2$
Streuung der $x$ -Werte: $Sxx = (n - 1) \cdot s_x^2$		(vgl. <i>Rechen-Box 1</i> )
Werte aus der Alternative $H$ : $b_1$		
<i>einseitig oben</i>	<i>einseitig unten</i>	<i>zweiseitig</i>
$t_m = t_{m;\alpha}$	$t_m = t_{m;\alpha}$	$t_m = t_{m;\frac{\alpha}{2}}$
Hilfsgrößen:	$\Delta =  b_1 - b_0  > 0$	(Abweichung von $H_0$ )
	$\delta = \frac{\Delta \cdot \sqrt{Sxx}}{\sigma_Y}$	
	$u = t_m - \delta$	
Fehlerrisiko 2. Art: $\beta(b_1) \approx \Phi(u)$	mit $\Phi = \text{Verteilungsfunktion}$	
Testschärfe (Power): $\gamma(b_1) \approx \Phi(-u)$		von $N(0, 1)$

**Äquidistante  $x$ -Werte:** Beim äquidistanten Design wählt man insgesamt  $n_x$  verschiedene  $x$ -Werte, die äquidistant angeordnet sind, d.h. je zwei benachbarte  $x$ -Werte haben den gleichen Abstand  $d$  voneinander. Ordnet man die  $x$ -Werte nach ihrer Größe, so erhält man

$$x_1, \quad x_2 = x_1 + d, \quad x_3 = x_1 + 2d \quad \dots \quad x_{n_x} = x_1 + (n_x - 1)d. \quad (4)$$

Der hierdurch „abgedeckte“ Bereich der  $x$ -Werte hat die Länge

$$L = x_n - x_1 = (n_x - 1)d \quad (\text{Spannweite der } x\text{-Werte}). \quad (5)$$

Für jeden  $x$ -Wert soll jetzt eine feste Anzahl  $k$  von Wiederholungen der zugehörigen Zufallsvariablen  $Y$  beobachtet werden. Konkret bedeutet dies, daß jeder  $x$ -Wert insgesamt  $k$  mal vervielfacht wird, und somit insgesamt

$$n = k \cdot n_x \quad (\text{Gesamtumfang}) \quad (6)$$

Punkte beobachtet werden sollen. Zum Beispiel ergibt sich für  $k = 3$  die folgende Reihe von insgesamt  $3n_x$   $x$ -Werten:

$$x_1, x_1, x_1, \quad x_2, x_2, x_2, \quad \dots \dots \dots \quad x_{n_x}, x_{n_x}, x_{n_x}.$$

Für eine solche Reihe von äquidistanten  $x$ -Werten mit  $k$  Wiederholungen lässt sich die zugehörige Summe  $Sxx$  nach längeren Umformungen wie folgt darstellen:

$$Sxx = \frac{k \cdot n_x \cdot (n_x + 1) \cdot L^2}{12(n_x - 1)} . \quad (7)$$

Die Länge  $L$  des abzudeckenden Bereichs wird durch die konkrete Anwendung vorgegeben. Interessiert man sich z.B. für einen Altersbereich von 20 bis 100 Jahren, so ist  $L = 80$ . Durch Einsetzen von (7) in (3) ergibt sich dann für *vorgegebenes*  $L$  der Zusammenhang

$$\frac{k \cdot n_x \cdot (n_x + 1)}{n_x - 1} \approx 12 \left[ \frac{(z + z_\beta) \cdot \sigma_Y}{\Delta \cdot L} \right]^2 \quad \text{mit} \quad z = \begin{cases} z_\alpha & \text{einseitig} \\ z_{\frac{\alpha}{2}} & \text{zweiseitig} \end{cases} , \quad (8)$$

aus dem man *eine* der beiden Größen  $k$  oder  $n_x$  bestimmen kann, wenn man die jeweils *andere* vorgibt.

Gibt man sich die Anzahl  $n_x$  der verschiedenen  $x$ -Werte vor (z.B.  $n_x = 17$  Altersstufen im Abstand von 5 Jahren, d.h.  $x_1 = 20, x_2 = 25, \dots, x_{17} = 100$ ), so erhält man aus (8) die zugehörige

*erforderliche Anzahl k der Wiederholungen pro x-Wert der  $n_x$  äquidistanten x-Werte*

$$k \approx \frac{12(n_x - 1)}{n_x(n_x + 1)} \quad \left[ \frac{(z + z_\beta) \cdot \sigma_Y}{\Delta \cdot L} \right]^2 \quad \text{mit} \quad z = \begin{cases} z_\alpha & \text{einseitig} \\ z_{\frac{\alpha}{2}} & \text{zweiseitig} \end{cases} . \quad (9)$$

Dieses  $k$  ist erst ganzzahlig *aufzurunden* und der zugehörige Umfang  $n$  ergibt sich aus (6). Man kann diesen Umfang auch direkt bestimmen aus:

*erforderlicher Umfang n bei vorgegebener Anzahl  $n_x$  äquidistanter x-Werte:*

$$n \approx \frac{12(n_x - 1)}{n_x + 1} \quad \left[ \frac{(z + z_\beta) \cdot \sigma_Y}{\Delta \cdot L} \right]^2 \quad \text{mit} \quad z = \begin{cases} z_\alpha & \text{einseitig} \\ z_{\frac{\alpha}{2}} & \text{zweiseitig} \end{cases} , \quad (10)$$

wobei dieses  $n$  auf das nächsthöhere Vielfache von  $n_x$  aufzurunden ist (also z.B. bei  $n_x = 10$  auf die nächsthöhere *durch 10 teilbare* ganze Zahl).

Gibt man sich umgekehrt die Anzahl  $k$  der Wiederholungen pro  $x$ -Wert vor (z.B.  $k = 1$  für *einen* oder  $k = 5$  für je *fünf* Wiederholungen), so ergibt sich zusammen mit der Vereinfachung  $n_x - 1 \approx n_x + 1$  (für *nicht zu kleines*  $n_x$ ) der

erforderliche Umfang  $n$  bei nicht zu kleiner Anzahl  $n_x$  äquidistanter  $x$ -Werte:

$$n \approx 12 \left[ \frac{(z + z_\beta) \cdot \sigma_Y}{\Delta \cdot L} \right]^2 \quad \text{mit} \quad z = \begin{cases} z_\alpha & \text{einseitig} \\ z_{\frac{\alpha}{2}} & \text{zweiseitig} \end{cases}. \quad (11)$$

Dieser Umfang  $n$  ist auf das nächsthöhere *Vielfache* von  $k$  aufzurunden (also z.B. bei  $k = 5$  auf die nächsthöhere *durch 5 teilbare* ganze Zahl). Die zugehörige Quadratsumme  $S_{xx}$  nach (7) ist dann (eventuell sogar deutlich) *größer* als der erforderliche Wert in der rechten Seite von (3).

Für *nicht zu kleines*  $n_x$  liefern (10) und (11) annähernd den gleichen Gesamtumfang  $n$ . Aber für *kleines*  $n_x$  kann der Unterschied bedeutend sein. Im Extremfall mit nur  $n_x = 2$  verschiedenen  $x$ -Werten ist der (noch nicht aufgerundete) Umfang  $n$  aus (11) genau das *Dreifache* des Umfangs  $n$  aus (10). Allerdings empfiehlt sich dieses extreme Design nicht, weil man mit ihm die Gültigkeit des linearen Modells nicht überprüfen kann. Typischerweise sollte  $n_x$  *mindestens* 5 (besser: 10) betragen, damit man auch die Gültigkeit des linearen Modells beurteilen kann.

**Zufällige  $x$ -Werte:** Wenn die Datenerhebung so erfolgt, daß die  $(x, y)$ -Paare Realisierungen von unabhängigen Wiederholungen des Zufalls-Paars  $(X, Y)$  sind, so sind die  $x$ -Werte *zufällig* und lassen sich nicht vorher festlegen. Ersetzt man in (2) die empirische Standardabweichung  $s_x$  durch die Standardabweichung  $\sigma_X$  der Zufallsvariablen  $X$ , so ergibt sich *approximativ*

$$S_{xx} \approx (n - 1) \cdot \sigma_X^2. \quad (12)$$

Zusammen mit (3) ergibt sich der

erforderliche Umfang  $n$  bei zufälligen  $x$ -Werten

$$n \approx 1 + \left[ \frac{(z + z_\beta) \cdot \sigma_Y}{\Delta \cdot \sigma_X} \right]^2 \quad \text{mit} \quad z = \begin{cases} z_\alpha & \text{beim einseitigen Test} \\ z_{\frac{\alpha}{2}} & \text{beim zweiseitigen Test} \end{cases}, \quad (13)$$

der wieder *ganzzahlig aufzurunden* ist. zieht man nun eine Stichprobe vom diesem Umfang  $n$ , so ist aber nicht sichergestellt, daß die resultierende Quadratsumme  $S_{xx}$  den erforderlichen Wert aus (3) auch erreicht, weil die empirische Standardabweichung  $s_x$  (zufallsbedingt) auch *kleiner* als die zur Berechnung verwendete Standardabweichung  $\sigma_X$  ausfallen kann. Vorsichtshalber kann man dann die Datenerhebung solange fortsetzen, bis  $S_{xx}$  den in (3) geforderten Wert erreicht oder übersteigt.

### 3.6.1 d Prognosen

Wir wollen uns jetzt mit einer weiteren wichtigen Aufgabe der Regressionsanalyse beschäftigen: der Prognose. Für einen vorgegebenen Wert  $x_0$  soll der zugehörige *Erwartungswert*  $f(x_0)$  von  $Y$  geschätzt (prognostiziert) werden und die Schätzung soll durch Konfidenzgrenzen ergänzt werden. Als Schätzung verwendet man natürlich den zu  $x_0$  gehörigen  $y$ -Wert auf der geschätzten Regressionsgeraden, d.h. den geschätzten Funktionswert

$$\hat{f}(x_0) = \hat{a} + \hat{b} x_0 \quad (\text{Prognose, geschätzter Funktionswert}). \quad (1)$$

Diese Schätzung streut – um den zugehörigen Erwartungswert  $f(x_0)$  von  $Y$  – mit der Varianz

$$\sigma^2(x_0) = \sigma_Y^2 \cdot h(x_0) \quad \text{wobei} \quad h(x_0) = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}. \quad (2)$$

Hierbei ist  $h(x_0)$  eine *quadratische* Funktion (Parabel) in  $x_0$ , die für  $x_0 = \bar{x}$  ihr Minimum annimmt. Eine *Schätzung* der Standardabweichung  $\sigma(x_0)$  erhält man, indem man die Standardabweichung  $\sigma_Y$  durch ihre Schätzung  $\hat{\sigma}_Y$  ersetzt:

$$\hat{\sigma}(x_0) = \hat{\sigma}_Y \cdot \sqrt{h(x_0)} \quad (\text{geschätzte Standardabweichung der Prognose}). \quad (3)$$

Die *einseitigen* Konfidenzgrenzen für  $f(x_0)$  zur *approximativen* Sicherheit  $1 - \alpha$  ergeben sich jetzt (analog 3.5.1) unter Verwendung des  $\alpha$ -Quantils  $t_{m;\alpha}$  der  $t_m$ -Verteilung (vgl. Anhang T) mit dem Freiheitsgrad  $m = n - 2$  zu

$$\hat{f}_{un,\alpha}(x_0) = \hat{f}(x_0) - \hat{d}_\alpha \quad (\text{einseitige untere Konfidenzgrenze für } b), \quad (4)$$

$$\hat{f}_{ob,\alpha}(x_0) = \hat{f}(x_0) + \hat{d}_\alpha \quad (\text{einseitige obere Konfidenzgrenze für } b), \text{ mit} \quad (5)$$

$$\hat{d}_\alpha = t_{m;\alpha} \cdot \hat{\sigma}(x_0) \quad (\text{Bandbreite}). \quad (6)$$

Wenn die Zielvariable  $Y$  *normalverteilt* ist, so haben diese Konfidenzgrenzen sogar die *exakte* Sicherheit  $1 - \alpha$ . Berechnet man diese Grenzen für  $\frac{\alpha}{2}$  statt  $\alpha$ , so erhält man ein *zweiseitiges Konfidenzintervall* (vgl. *Prognose-Box 5a*).

Speziell für  $x_0 = 0$  ist  $f(0) = a$ , und man erhält in diesem Fall aus (3) die (geschätzte) Standardabweichung von  $\hat{a}$  sowie aus (4)-(6) Konfidenzgrenzen für Achsenabschnittsparameter  $a$ .

**Prognose-Box 5a (ungewichtete Regression):**

Prognose des Funktionswerts  $f(x_0) = \hat{a} + \hat{b}x_0$  für vorgegebenes  $x_0$

Stichprobenumfang: $n$	$(x, y)$ -Werte: $(x_1, y_1), \dots, (x_n, y_n)$
------------------------	--

Berechnungen (zu nicht erläuterten Größen vgl. Rechen-Box 2):

$$\text{Regressionsgerade } y = \hat{a} + \hat{b}x : \quad \hat{b} = S_{xy}/S_{xx}, \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$\text{Streuung um die Gerade: } SQ(\hat{a}, \hat{b}) = S_{yy} - \hat{b} \cdot S_{xy}$$

$$\text{Freiheitsgrad: } m = n - 2$$

$$\text{Standardabweichung der } y\text{-Werte: } \hat{\sigma}_Y = \sqrt{\frac{1}{m}SQ(\hat{a}, \hat{b})}$$

**Prognose des Funktionswertes  $f(x_0)$ :**

$$\text{Prognose: } \hat{f}(x_0) = \hat{a} + \hat{b}x_0$$

$$\text{Standardabweichung der Prognose: } \hat{\sigma}(x_0) = \hat{\sigma}_Y \cdot \sqrt{h(x_0)} \quad \text{mit}$$

$$h(x_0) = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}$$

**Einseitige Konfidenzgrenzen des Funktionswerts  $f(x_0)$ :**

bei nicht zu kleinem  $n$  (mindestens 10): Sicherheit  $\approx 1 - \alpha$

bei normalverteiltem  $Y$ : Sicherheit =  $1 - \alpha$  (exakt)

einseitige untere Grenze	einseitige obere Grenze
--------------------------	-------------------------

$$\hat{f}_{un,\alpha}(x_0) = \hat{f}(x_0) - \hat{d}_\alpha$$

$$\hat{f}_{ob,\alpha}(x_0) = \hat{f}(x_0) + \hat{d}_\alpha$$

$$\hat{d}_\alpha = t_{m;\alpha} \cdot \hat{\sigma}(x_0) \quad (\text{Bandbreite})$$

**Zweiseitiges Konfidenzintervall für  $f(x_0)$ :**  $(\hat{f}_{un,\frac{\alpha}{2}}(x_0), \hat{f}_{ob,\frac{\alpha}{2}}(x_0))$ **3.6.1 e Anwendungen**

Prinzipiell können alle Beispiele für (ungewichtete) Regressionsgeraden aus den Abschnitten 1.5 und 1.6 mit den Methoden der Regressionsanalyse eingehender untersucht werden, und wir betrachten zuerst zwei Anwendungen für die StatLab-Daten.

### **StatLab-Auswahl 1985: Zigarettenkonsum der Mutter und Geburtsgewicht**

Die Abhängigkeit des Geburtsgewicht  $Y$  des Kindes vom Zigarettenkonsum  $X$  [in Zigaretten/Tag] der Mutter soll durch ein lineares Regressionsmodell beschrieben und der Einfluß soll zum Niveau  $\alpha = 5\%$  überprüft werden. Bei Müttern, die nie geraucht oder mit dem Rauchen aufgehört haben, wird der Wert  $X = 0$  verwendet.

Für die  $n = 100$  Kinder der StatLab-Auswahl 1985 erhält man aus

$$\begin{array}{lll} \bar{x} = 5,280 & \bar{y} = 7,387 \\ S_{xx} = 7796,1600 & S_{yy} = 137,0931 & S_{xy} = -320,2360 \end{array}$$

die folgenden Schätzungen

$$\begin{array}{lll} \hat{a} = 7,6039 & \hat{b} = -0,0411 & r_{xy} = -0,310 \\ \hat{\sigma}_Y = 1,1246 & \hat{\sigma}_{\hat{b}} = 0,0127. & \end{array}$$

Der zweiseitige Test von  $H_0: b = 0$  gegen  $H: b \neq 0$  lehnt die Nullhypothese ab, weil

$$\begin{aligned} |t| &= 3,225 \geq t_{98;2,5\%} = 1,984 & \text{bzw.} & P = 0,17\% \leq 5\% & \text{bzw.} \\ |r_{xy}| &= 0,310 \geq r_{98;2,5\%} = 0,197. \end{aligned}$$

Folglich ist der (negative) Einfluß des Zigarettenkonsums auf das Geburtsgewicht mit einer Sicherheit von 95% erwiesen, obwohl „nur“  $B_{xy} = 9,6\%$  der Streuung der  $y$ -Werte durch die Regressionsgerade „erklärt“ wird (eine *vollständigere* statistische Analyse des Geburtsgewichts – auf die hier verzichtet wird – würde natürlich neben dem Zigarettenkonsum der Mutter noch weitere Einflußvariablen berücksichtigen).

Das zweiseitige 95%-Konfidenzintervall für  $b$  ergibt sich zu

$$\hat{b}_{un} = -0,066 \quad \hat{b}_{ob} = -0,016 \quad \text{mit} \quad \hat{d}_{2,5\%} = 0,025.$$

Die Darstellung der beobachteten Punkte zusammen mit der Regressionsgeraden  $y = \hat{f}(x)$  in Abb. 1 (*links*) und der zugehörige Residuenplot in Abb. 1 (*rechts*) geben keinen Anlaß, um an dem linearen Regressionsmodell zu zweifeln, obwohl ein einzelner „Ausreißerpunkt“ (mit  $x = 20$  Zig./Tag und extremem Untergewicht  $y = 2,3$  Pound) auch ein sehr extremes Residuum von rund  $-4$  hat.

Zusätzlich sind in Abb. 1 (*links*) die zweiseitigen 95%-Konfidenzgrenzen  $\hat{f}_{un}(x)$  und  $\hat{f}_{ob}(x)$  für das prognostizierte (erwartete) Geburtsgewicht  $f(x)$  als Funktion des Zigarettenkonsums  $x$  eingezeichnet. Man erkennt deutlich, daß diese Konfidenzintervalle größer werden, je weiter  $x$  vom Mittelwert  $\bar{x}$  entfernt ist. Konkret ergibt sich für  $x = 45$  Zigaretten/Tag das erwartete Geburtsgewicht und sein 95%-Konfidenzintervall nach **3.6.1 d** zu (vgl. auch Abb. 1)

$$\hat{f}(45) = 5,76 \quad \hat{f}_{un}(45) = 4,73 \quad \hat{f}_{ob}(45) = 6,78.$$

Und für  $x = 0$  Zigaretten/Tag erhält man analog ein 95%-Konfidenzintervall für das erwartete Geburtsgewicht  $a = f(0)$

$$\hat{a}_{un} = \hat{f}_{un}(0) = 7,34 \quad \hat{a}_{ob} = \hat{f}_{ob}(0) = 7,86.$$

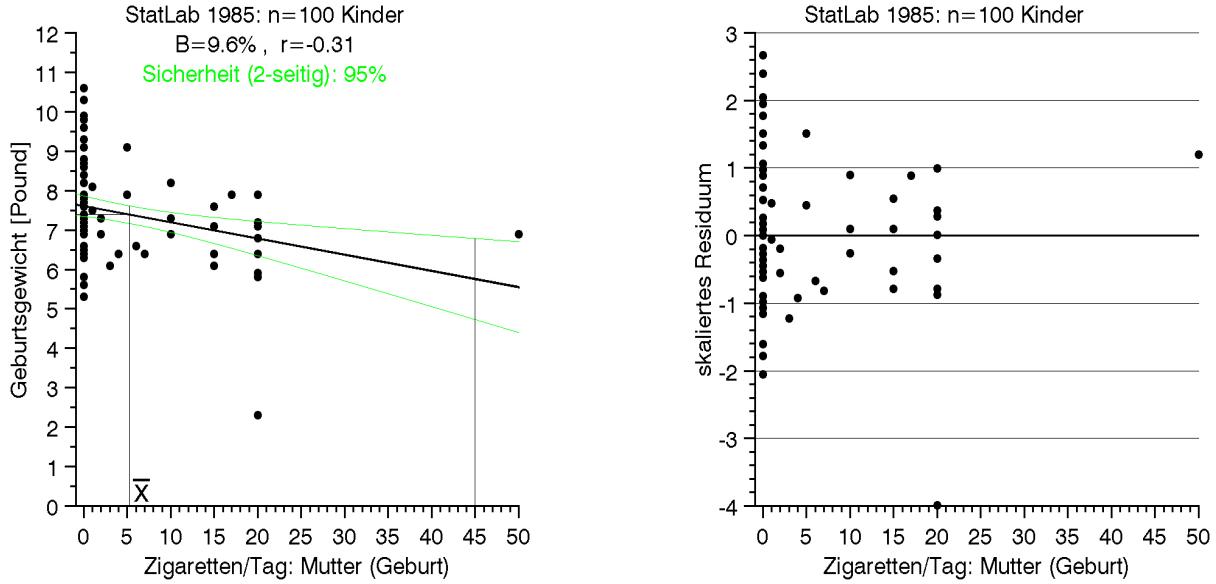


Abb. 1: *Zigarettenkonsum der Mutter und Geburtsgewicht*

- links:* Punkte mit geschätzter Regressionsgeraden  $\hat{f}$  und den Funktionen  $\hat{f}_{un}$  und  $\hat{f}_{ob}$  der zweiseitigen 95%-Konfidenzgrenzen (vgl. Text)
- rechts:* zugehöriger Residuenplot

### *StatLab-Auswahl 1985: Alter der Mutter und Geburtsgewicht*

Die Abhängigkeit des Geburtsgewichts  $Y$  des Kindes vom Alter  $X$  [in Jahren] der Mutter soll durch ein lineares Regressionsmodell beschrieben und der Einfluß soll zum Niveau  $\alpha = 5\%$  überprüft werden.

Für die  $n = 100$  Kinder der StatLab-Auswahl 1985 erhält man aus

$$\begin{aligned}\bar{x} &= 26,880 & \bar{y} &= 7,387 \\ S_{xx} &= 3050,5600 & S_{yy} &= 137,0931 & S_{xy} &= 119,4440\end{aligned}$$

die folgenden Schätzungen

$$\begin{aligned}\hat{a} &= 6,3345 & \hat{b} &= 0,0392 & r_{xy} &= 0,185 \\ \hat{\sigma}_Y &= 1,1624 & \hat{\sigma}_{\hat{b}} &= 0,0210\end{aligned}$$

Die Nullhypothese  $H_0: b = 0$  wird beim zweiseitigen Test gegen  $H: b \neq 0$  nicht abgelehnt, weil

$$|t| = 1,860 < t_{98;2,5\%} = 1,984 \quad \text{bzw.} \quad P = 6,58\% > 5\% \quad \text{bzw.}$$

$$|r_{xy}| = 0,185 < r_{98;2,5\%} = 0,197.$$

Das zweiseitige 95%-Konfidenzintervall für  $b$  ist gegeben durch

$$\hat{b}_{un} = -0,0026 \quad \hat{b}_{ob} = 0,0809 \quad \text{mit} \quad \hat{d}_{2,5\%} = 0,0418$$

und enthält die Null (weil die Nullhypothese  $H_0: b = 0$  nicht abgelehnt wurde).

Die beobachteten Punkte zusammen mit der Regressionsgeraden  $y = \hat{f}(x)$  – und der Mittelwertsgeraden  $y = \bar{y}$  – sowie den zweiseitigen 95%-Konfidenzgrenzen  $\hat{f}_{un}(x)$  und  $\hat{f}_{ob}(x)$  sind in Abb. 2 dargestellt. Auffällig ist wieder der „Ausreißer“ mit  $x = 24$  Jahren und dem extremen Untergewicht  $y = 2,3$  Pound.

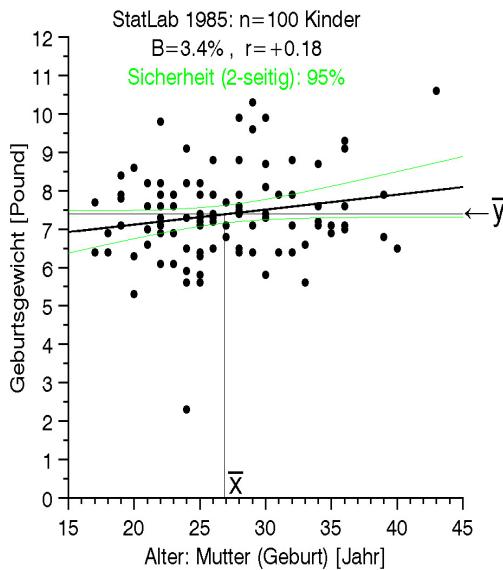


Abb. 2: Alter der Mutter und Geburtsgewicht

Punkte mit geschätzter Regressionsgeraden  $\hat{f}$ , Mittelwertsgeraden  $y = \bar{y}$  und den Funktionen  $\hat{f}_{un}$  und  $\hat{f}_{ob}$  der zweiseitigen 95%-Konfidenzgrenzen (vgl. Text)

Da bei Nichtablehnung der Nullhypothese ein Fehler 2. Art vorliegen kann, ist die Nullhypothese – d.h. das Alter der Mutter hat *keinen* Einfluß auf das (erwartete) Geburtsgewicht – nicht mit *quantifizierter* Sicherheit bestätigt. Wir können aber das Fehlerrisiko 2. Art für hypothetische Werte des Anstiegs  $b$  bestimmen. Wenn sich über dem Altersbereich von 20 bis 40 Jahren das Geburtsgewicht z.B. um 2 [Pound] – d.h. rund 1 kg – erhöht, so entspricht dies dem Anstieg  $b = 0,1$ . Gehen wir weiter davon aus, daß die unbekannte Standardabweichung  $\sigma_Y$  der *beobachteten* entspricht, d.h.  $\sigma_Y = \hat{\sigma}_Y = 1,1624$ , so ergibt sich aus der *Power-Box 5a*

$$\delta = 4,752 \quad u = -2,762 \quad \beta(0,1) \approx 0,3\%.$$

Für diesen „deutlichen“ Anstieg ist das Fehlerrisiko 2. Art also extrem gering. Ist der Anstieg aber mit  $b = 0,05$  nur *halb* so groß, so ist das Risiko  $\beta(0,05) \approx 35\%$  bereits inakzeptabel hoch. Wir wollen daher für diesen Anstieg den erforderlichen Mindestumfang  $n$  bestimmen, bei dem das Fehlerrisiko 2. Art nur  $\beta = 10\%$  beträgt. Zur Illustration geben wir die Berechnungen sowohl für ein Design mit *äquidistanten*  $x$ -Werten als auch für *zufällige*  $x$ -Werte an. Bei einer Datenerhebung mit *zufälligen*  $x$ -Werten (die ja bei den

StatLab-Daten vorlag) ergibt sich unter Verwendung von  $\sigma_X = s_x = 5,55$  aus (12) der (aufgerundete) Umfang von  $n = 152$ .

Bei einer Datenerhebung mit vorgebbaren  $x$ -Werten wählen wir zunächst den interessierenden Altersbereich von 15 bis 40 Jahre, der die Länge  $L = 25$  hat. Wenn wir uns für insgesamt  $n_x = 6$  Altersstufen im Abstand von je 5 Jahren entscheiden (d.h.  $x_1 = 15, \dots, x_6 = 40$ ), so ergibt sich aus (9) und  $z_\beta = 1,282$  die erforderliche Anzahl von Kindern für jede Altergruppe der Mutter zu  $k \approx 12,98$  – aufgerundet also  $k = 13$ . Insgesamt sind daher  $n = 78$  Kinder erforderlich, und das sind sogar *weniger* als die oben analysierten 100 Kinder. Beim balancierten Design mit  $n = 78$  ist das Fehlerrisiko 2. Art also geringer ( $\beta = 10\%$ ) als oben ( $\beta = 35\%$ ) beim Umfang  $n = 100$ . Dies liegt daran, daß hier nicht die Größe  $n$  des *Stichprobenumfangs* entscheidend ist, sondern die *Quadratsumme Sxx*. Und im *balancierten* Design ist  $Sxx = 5687,50$  (nach (7)) wesentlich größer als beim *zufälligen* Design oben mit  $Sxx = 3050,56$ . Dies unterstreicht noch einmal die Bedeutung der Versuchsplanung. Typischerweise interessiert man sich bei Studien (wie z.B. der StatLab-Studie) für *mehrere* Fragestellungen (also nicht nur für den Einfluß des Alters der Mutter auf das Geburtsgewicht) und man kann das Studiendesign dann nur für eine *Hauptfragestellung* optimieren.

Zur Illustration betrachten wir noch ein äquidistantes Design mit insgesamt  $n_x = 26$  Altersstufen im Abstand von je *einem* Jahr. Hier sind (aufgerundet)  $k = 4$  Kinder pro Altersstufe erforderlich und somit insgesamt  $n = 104$  Kinder. Dies sind deutlich mehr als oben ( $n = 78$ ) bei  $n_x = 6$  Altersstufen. Hieran erkennt man, daß eine Konzentration auf *weniger*  $x$ -Werte (wie oben) für das *Fehlerrisiko* günstiger ist. Für die *Modellüberprüfung* dagegen ist es sinnvoller, nicht zu wenig  $x$ -Werte zur Verfügung zu haben (wie auch das folgende Beispiel zeigt).

### **Blickwinkel und Reaktionszeit bei Rhesusaffen**

Wir greifen das Beispiel aus **1.7.1** noch einmal auf, um zu demonstrieren, wie wichtig die *Überprüfung* des linearen Regressionsmodells an Hand grafischer Darstellungen ist. Hierzu „vergessen“ wir zunächst die Analyse aus **1.7.1** und betrachten ein lineares Regressionsmodell für die Reaktionszeit  $Y$  in Abhängigkeit vom *Absolutbetrag X des Blickwinkels* (d.h. wir ignorieren sein Vorzeichen). Die zugehörige Regressionsgerade „erklärt“  $B_{xy} = 86\%$  der Streuung der  $y$ -Werte, und die Nullhypothese  $H_0: b = 0$  wird beim zweiseitigen Test zum Niveau  $\alpha = 5\%$  abgelehnt, weil

$$|r_{xy}| = 0,93 \geq r_{15;2,5\%} = 0,4821 \quad \text{bzw.} \quad B_{xy} = 86\% \geq r_{15;2,5\%}^2 = 23,24\%.$$

Ein Blick auf *Abb. 3 (oben)* läßt jedoch *systematische* Abweichungen der beobachteten Punkte von der Geraden erkennen, die im Residuenplot besonders auffällig sind. Und *Abb. 3 (unten)* zeigt auch klar die Überlegenheit des *quadratischen Regressionsmodells* aus **1.7.1** gegenüber dem hier betrachteten *linearen* Regressionsmodell. Es ist unmittelbar einleuchtend, daß ein Residuenplot aussagekräftiger wird, je mehr (verschiedene)

$x$ -Werte zu Verfügung stehen. Mit z.B. nur 5  $x$ -Werten ( $-90, -45, 0, 45, 90$ ) und jeweils 4 Wiederholungen pro  $x$ -Wert ließe sich die Abweichung vom linearen Modell nicht so überzeugend erkennen.

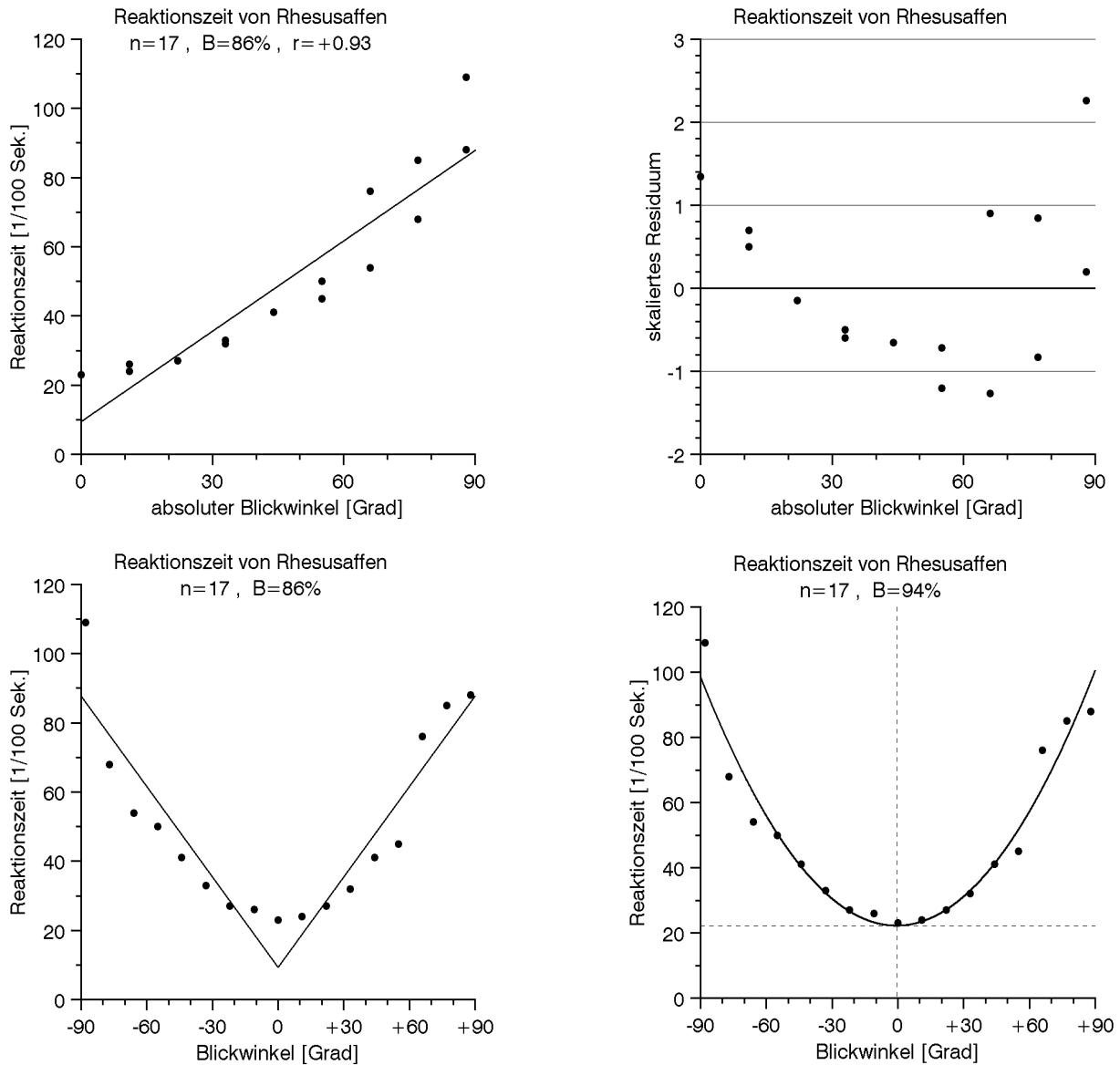


Abb. 3: (oben) beobachtete Punkte mit Regressionsgerade und Residuenplot  
(unten) lineares und quadratisches Regressionsmodell (aus 1.7.1)

### ☒ Tipps für Microsoft Excel

Wenn man sich eine Arbeitsmappe für die Regressionsanalyse anlegt, so ist es vorteilhaft, zuerst die *gewichtete* Regressionsanalyse aus 3.6.2 zu behandeln. Die *ungewichtete* Analyse ergibt sich daraus als Spezialfall mit *konstanten* Gewichten  $w_i = 1$ .

### 3.6.2 Gewichtete Regressionsanalyse bei *proportionalen* Varianzen

Die bisherigen Betrachtungen sollen jetzt auf den allgemeineren Fall übertragen werden, bei dem die voneinander unabhängigen Zufallsvariablen  $Y_1, Y_2, \dots, Y_n$  nicht notwendig die *gleiche Varianz*  $\sigma^2$  besitzen müssen. Hierbei dürfen die Varianzen allerdings nicht völlig beliebig sein, sondern müssen sich um *bekannte* Faktoren  $w_i > 0$  voneinander unterscheiden. Genauer soll gelten

$$\text{Var}(Y_i) = \frac{\sigma_Y^2}{w_i} \quad \text{für } i = 1, \dots, n \quad (\text{proportionale Varianzen}) \quad (1)$$

wobei  $\sigma_Y^2$  *unbekannt* ist und auch *geschätzt* werden soll. Dies bedeutet nicht anderes, als daß die *Quotienten* je zweier Varianzen *bekannt* sind:

$$\frac{\text{Var}(Y_i)}{\text{Var}(Y_k)} = \frac{w_k}{w_i} \quad \text{für alle } i \text{ und } k \quad (\text{proportionale Varianzen}) \quad (2)$$

Der Faktor  $w_i$  ist *umgekehrt proportional* zur Varianz von  $Y_i$  und wird deshalb auch als *Gewichtung* interpretiert: je *größer*  $w_i$  ist, desto *weniger* streut die Beobachtung  $Y_i$  und sie ist deshalb *gewichtiger*. Eine typische Situation für proportionale Varianzen liegt vor, wenn jede Beobachtung  $Y_i$  bereits ein Mittelwert von  $n_i$  unabhängigen Wiederholungen darstellt. Dann ist  $w_i = n_i$  und  $\sigma_Y^2$  ist die zugehörige Varianz der zugrundeliegenden einzelnen Wiederholung.

Die Regressionsanalyse mit proportionalen Varianzen wird auch als *gewichtete Regressionsanalyse* bezeichnet. Formal unterscheidet sich diese Analyse von der *ungewichteten* Regressionsanalyse aus **3.6.1** lediglich dadurch, daß alle Berechnungen mit den entsprechenden *gewichteten Summen* aus **1.5.6** an Stelle der *ungewichteten* Summen aus **1.5.4** durchzuführen sind. Aus diesem Grunde geben wir die Rechen-Boxen hier nur an und verweisen für die Erläuterungen auf die entsprechenden Abschnitte über *ungewichtete* Regression in **3.6.1**. Die nicht durch diese Boxen abgedeckte *Residuenanalyse* wird noch kurz abgehandelt, aber auf Ausführungen zur *Versuchsplanung* verzichten wir hier. – Man beachte, daß die proportionalen Varianzen auch den bisher betrachteten Fall mit gleichen Varianzen einschließt, wenn *alle* Gewichte *übereinstimmen*, also z.B. gleich 1 sind.

<b>Test-Box 5c (gewichtete Regression):</b>		<b>Test auf signifikanten Anstieg</b>
<i>Stichprobenumfang:</i> $n$	<i>gewichtete-Werte:</i> $(x_i, y_i), w_i$	für $i = 1, \dots, n$
Schätzungen (zu nicht erläuterten Größen vgl. <i>Rechen-Box 3:</i> )		
<i>gewichtete Regressionsgerade</i> $y = \hat{a} + \hat{b}x$ :	$\hat{b} = Swxy/Swxx$ ,	$\hat{a} = \bar{y} - \hat{b}\bar{x}$
<i>gewichtete Streuung um die Gerade:</i>	$WSQ(\hat{a}, \hat{b}) = Swyy - \hat{b} \cdot Swxy$	
<i>Freiheitsgrad:</i>	$m = n - 2$	
<i>Standardabweichung der y-Werte:</i>	$\hat{\sigma}_Y = \sqrt{\frac{1}{m} WSQ(\hat{a}, \hat{b})}$	
<i>Standardabweichung von <math>\hat{b}</math>:</i>	$\hat{\sigma}_{\hat{b}} = \frac{\hat{\sigma}_Y}{\sqrt{Swxx}}$	
<i>einseitig oben</i>	<i>einseitig unten</i>	<i>zweiseitig</i>
$H_0: b \leq 0$ vs. $H: b > 0$	$H_0: b \geq 0$ vs. $H: b < 0$	$H_0: b = 0$ vs. $H: b \neq 0$
<b>Student's t-Test:</b>		
bei nicht zu kleinem $n$ (mindestens 10):	Fehlerrisiko 1. Art $\approx \alpha$	
bei normalverteiltem $Y$ :	Fehlerrisiko 1. Art $= \alpha$ (exakt)	
Testwert:	$t = \frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} = \frac{\hat{b} \cdot \sqrt{Swxx}}{\hat{\sigma}_Y} = \frac{r_{xy} \sqrt{m}}{\sqrt{1 - r_{xy}^2}}$	
kritische Werte:	$t_{m;\alpha}$ ( $\alpha$ -Quantil von $t_m$ )	für den Testwert
	$r_{m;\alpha} = \frac{t_{m;\alpha}}{\sqrt{m + t_{m;\alpha}^2}}$	für den Korrelationskoeffizienten
$\Phi_m$ bezeichnet die Verteilungsfunktion der $t_m$ -Verteilung.		
Entscheidung: Ablehnung von $H_0$ (Annahme von $H$ ), falls gilt		
<i>einseitig oben</i>	<i>einseitig unten</i>	<i>zweiseitig</i>
Testwert $t \geq t_{m;\alpha}$ bzw. $P$ -Wert $\Phi_m(-t) \leq \alpha$ bzw. $r_{xy} \geq r_{m;\alpha}$	Testwert $t \leq -t_{m;\alpha}$ bzw. $P$ -Wert $\Phi_m(t) \leq \alpha$ bzw. $r_{xy} \leq -r_{m;\alpha}$	Testwert $ t  \geq t_{m;\frac{\alpha}{2}}$ bzw. $P$ -Wert $2\Phi_m(- t ) \leq \alpha$ bzw. $ r_{xy}  \geq r_{m;\frac{\alpha}{2}}$ bzw. $B_{xy} \geq r_{m;\frac{\alpha}{2}}^2$

Konfidenz- und Test-Box 5d ( <i>gewichtete Regression</i> ):	Parameter <b>b</b>
Berechnungen (zu nicht erläuterten Größen vgl. <i>Rechen-Box 3, Test-Box 5c</i> ):	
Standardabweichung der $y$ -Werte:	$\hat{\sigma}_Y = \sqrt{\frac{1}{m} WSQ(\hat{a}, \hat{b})}$
Standartabweichung von $\hat{b}$ :	$\hat{\sigma}_{\hat{b}} = \frac{\hat{\sigma}_Y}{\sqrt{Swxx}}$
Freiheitsgrad:	$m = n - 2$

**Einseitige Konfidenzgrenzen für b:**bei nicht zu kleinem  $n$  (mindestens 10): Sicherheit  $\approx 1 - \alpha$ bei normalverteiltem  $Y$ : Sicherheit  $= 1 - \alpha$  (exakt)

einseitige untere Grenze	einseitige obere Grenze
$\hat{b}_{un,\alpha} = \hat{b} - \hat{d}_{\alpha}$	$\hat{b}_{ob,\alpha} = \hat{b} + \hat{d}_{\alpha}$
$\hat{d}_{\alpha} = t_{m;\alpha} \cdot \hat{\sigma}_{\hat{b}}$ (Bandbreite)	

**Student's t-Test:**bei nicht zu kleinem  $n$  (mindestens 10): Fehlerrisiko 1. Art  $\approx \alpha$ bei normalverteiltem  $Y$ : Fehlerrisiko 1. Art  $= \alpha$  (exakt)

einseitig oben	einseitig unten	zweiseitig
$H_0: b \leq b_0$ vs. $H: b > b_0$	$H_0: b \geq b_0$ vs. $H: b < b_0$	$H_0: b = b_0$ vs. $H: b \neq b_0$

Testwert: 
$$t = \frac{\hat{b} - b_0}{\hat{\sigma}_{\hat{b}}} = \frac{(\hat{b} - b_0) \cdot \sqrt{Swxx}}{\hat{\sigma}_Y}$$

Entscheidung: Ablehnung von  $H_0$  (Annahme von  $H$ ), falls gilt

einseitig oben	einseitig unten	zweiseitig
Testwert $t \geq t_{m;\alpha}$ bzw. $P$ -Wert $\Phi_m(-t) \leq \alpha$ bzw. $b_0 \leq \hat{b}_{un,\alpha}$	Testwert $t \leq -t_{m;\alpha}$ bzw. $P$ -Wert $\Phi_m(t) \leq \alpha$ bzw. $\hat{b}_{ob,\alpha} \leq b_0$	Testwert $ t  \geq t_{m;\frac{\alpha}{2}}$ bzw. $P$ -Wert $2\Phi_m(- t ) \leq \alpha$ bzw. $b_0 \leq \hat{b}_{un,\frac{\alpha}{2}}$ oder $\hat{b}_{ob,\frac{\alpha}{2}} \leq b_0$

**Power-Box 5b (gewichtete Regression):**

Für Tests über den Anstiegs-Parameter  $b$  der Regressionsgeraden

einseitig oben	einseitig unten	zweiseitig
$H_0: b \leq b_0$	$H_0: b \geq b_0$	$H_0: b = b_0$
<i>vs.</i>	<i>vs.</i>	<i>vs.</i>
$H: b > b_0$	$H: b < b_0$	$H: b \neq b_0$
Fehlerrisiko 1. Art (Testniveau): $\alpha$		Standardabweichung: $\sigma_Y$
Stichprobenumfang:	$n$	Freiheitsgrad: $m = n - 2$
Streuung der $x$ -Werte:	$Swxx = (n - 1) \cdot s_x^2$	(vgl. Rechen-Box 3)
Werte aus der Alternative $H$ :	$b_1$	
einseitig oben	einseitig unten	zweiseitig
$t_m = t_{m;\alpha}$	$t_m = t_{m;\alpha}$	$t_m = t_{m;\frac{\alpha}{2}}$
<i>Hilfsgrößen:</i>		$\Delta =  b_1 - b_0  > 0$ (Abweichung von $H_0$ )
		$\delta = \frac{\Delta \cdot \sqrt{Swxx}}{\sigma_Y}$
		$u = t_m - \delta$
Fehlerrisiko 2. Art:		$\beta(b_1) \approx \Phi(u)$ mit $\Phi = \text{Verteilungsfunktion}$
Testschärfe (Power):		$\gamma(b_1) \approx \Phi(-u)$ von $N(0, 1)$

**Residuenanalyse:** Bei der Residuenanalyse werden die folgenden *skalierten Residuen*

$$r_i = \frac{\sqrt{w_i} [y_i - \hat{f}(x_i)]}{\hat{\sigma}_Y} = \frac{\sqrt{w_i} [y_i - (\hat{a} + \hat{b} x_i)]}{\hat{\sigma}_Y} \quad (\text{skaliertes Residuum}) \quad (3)$$

verwendet, die das Gewicht  $w_i$  der Beobachtung  $y_i$  mit berücksichtigen.

**Prognose-Box 5b (gewichtete Regression):**

*Prognose des Funktionswerts*  $f(x_0) = \hat{a} + \hat{b}x_0$  für vorgegebenes  $x_0$

Stichprobenumfang:  $n$  gewichtete Werte:  $(x_i, y_i)$ ,  $w_i$  für  $i = 1, \dots, n$

Schätzungen (zu nicht erläuterten Größen vgl. Rechen-Box 3:)

$$\text{Regressionsgerade } y = \hat{a} + \hat{b}x : \quad \hat{b} = \frac{Swxy}{Swxx}, \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

$$\text{gewichtete Streuung um die Gerade: } WSQ(\hat{a}, \hat{b}) = Swyy - \hat{b} \cdot Swxy$$

$$\text{Standardabweichung der } y\text{-Werte: } \hat{\sigma}_Y = \sqrt{\frac{1}{m} WSQ(\hat{a}, \hat{b})}$$

$$\text{Freiheitsgrad: } m = n - 2$$

**Prognose des Funktionswertes  $f(x_0)$ :**

$$\text{Prognose: } \hat{f}(x_0) = \hat{a} + \hat{b}x_0$$

$$\text{Standardabweichung der Prognose: } \hat{\sigma}(x_0) = \hat{\sigma}_Y \cdot \sqrt{h(x_0)} \quad \text{mit}$$

$$h(x_0) = \frac{1}{\sum w} + \frac{(x_0 - \bar{x})^2}{Swxx}$$

**Einseitige Konfidenzgrenzen des Funktionswertes  $f(x_0)$ :**

bei nicht zu kleinem  $n$  (mindestens 10): Sicherheit  $\approx 1 - \alpha$

bei normalverteiltem  $Y$ : Sicherheit =  $1 - \alpha$  (exakt)

einseitige untere Grenze

einseitige obere Grenze

$$\hat{f}_{un,\alpha}(x_0) = \hat{f}(x_0) - \hat{d}_\alpha$$

$$\hat{f}_{ob,\alpha}(x_0) = \hat{f}(x_0) + \hat{d}_\alpha$$

$$\hat{d}_\alpha = t_{m,\alpha} \cdot \hat{\sigma}(x_0) \quad (\text{Bandbreite})$$

**Zweiseitiges Konfidenzintervall für  $f(x_0)$ :**  $(\hat{f}_{un,\frac{\alpha}{2}}(x_0), \hat{f}_{ob,\frac{\alpha}{2}}(x_0))$ **3.6.2 a Anwendungen**

Prinzipiell können alle Beispiele für *gewichtete* Regressionsgeraden aus den Abschnitten **1.5** und **1.6** mit den Methoden der *gewichteten* Regressionsanalyse eingehender untersucht werden, und wir greifen hier nur ein Beispiel wieder auf.

### Blutdruck und Alter

Wir betrachten noch einmal das Beispiel aus **1.5.7**, bei dem die Abhängigkeit des systolischen Blutdrucks  $Y$  vom Alter (bei Männern) durch ein lineares Regressionsmodell beschrieben wird. Der (aus **1.5** Abb. 7) evidente Einfluß des Alters soll jetzt zur Demonstration auch noch formal zum Niveau  $\alpha = 5\%$  überprüft werden.

Für die  $n = 10$  Altersklassen der Daten aus **1.5.7 Tab. 2** erhält man aus

$$\begin{array}{lll} \bar{x} = 40,211 & \bar{y} = 130,392 \\ Swxx = 14\,215,06 & Swxy = 11\,094,90 & Swyy = 8\,942,26 \end{array}$$

die folgenden Schätzungen

$$\begin{array}{lll} \hat{a} = 99,001 & \hat{b} = 0,7805 & r_{xy} = 0,984 \\ \hat{\sigma}_Y = 5,9441 & \hat{\sigma}_{\hat{b}} = 0,0499. \end{array}$$

Der zweiseitige Test von  $H_0: b = 0$  gegen  $H: b \neq 0$  lehnt die Nullhypothese ab, weil

$$\begin{array}{lll} |t| = 15,666 \geq t_{8;2,5\%} = 2,306 & \text{bzw.} & P < 0,005\% \leq 5\% \quad \text{bzw.} \\ |r_{xy}| = 0,984 \geq r_{8;2,5\%} = 0,632. & & \end{array}$$

Folglich ist der (positive) Einfluß des Alters auf den Blutdruck mit einer Sicherheit von 95% erwiesen. Das zweiseitige 95%-Konfidenzintervall für  $b$  ergibt sich zu

$$\hat{b}_{un} = 0,666 \quad \hat{b}_{ob} = 0,895 \quad \text{mit} \quad \hat{d}_{2,5\%} = 0,115.$$

Die Darstellung der beobachteten Punkte zusammen mit der Regressionsgeraden  $y = \hat{f}(x)$  in *Abb. 4 (links)* und der zugehörige Residuenplot in *Abb. 4 (rechts)* geben keinen Anlaß, an dem linearen Regressionsmodell zu zweifeln, zumal sogar  $B_{xy} = 97\%$  der Streuung der  $y$ -Werte durch die Regressionsgerade „erklärt“ wird.

Zusätzlich sind in *Abb. 4 (links)* die zweiseitigen 95%-Konfidenzgrenzen  $\hat{f}_{un}(x)$  und  $\hat{f}_{ob}(x)$  für den prognostizierten (erwarteten) Blutdruck  $f(x)$  als Funktion des Alters  $x$  eingezeichnet. Man erkennt wieder deutlich, daß diese Konfidenzintervalle größer werden, je weiter  $x$  vom Mittelwert  $\bar{x}$  entfernt ist.

Konkret ergibt sich für  $x = 0$  das 95%-Konfidenzintervall für den Achsenabschnitt  $a = f(0)$  zu (vgl. in *Abb. 4 links* die Schnittpunkte von  $\hat{f}_{un}$  und  $\hat{f}_{ob}$  mit der  $y$ -Achse)

$$\hat{a}_{un} = \hat{f}_{un}(0) = 94,1 \quad \hat{a}_{ob} = \hat{f}_{ob}(0) = 103,9$$

Es gibt eine grobe *Faustregel* für den systolischen Blutdruck, die besagt

$$\text{Blutdruck} = 100 + \text{Alter},$$

und dies ist gleichbedeutend damit, daß sowohl  $a$  den Wert 100 hat (der im obigen Konfidenzbereich für  $a$  liegt) als auch, daß der Anstieg  $b$  den Wert 1 hat (der *nicht* im obigen

Konfidenzbereich für  $b$  liegt). Folglich kann diese Faustregel hier (bei Männern) nicht bestätigt werden.

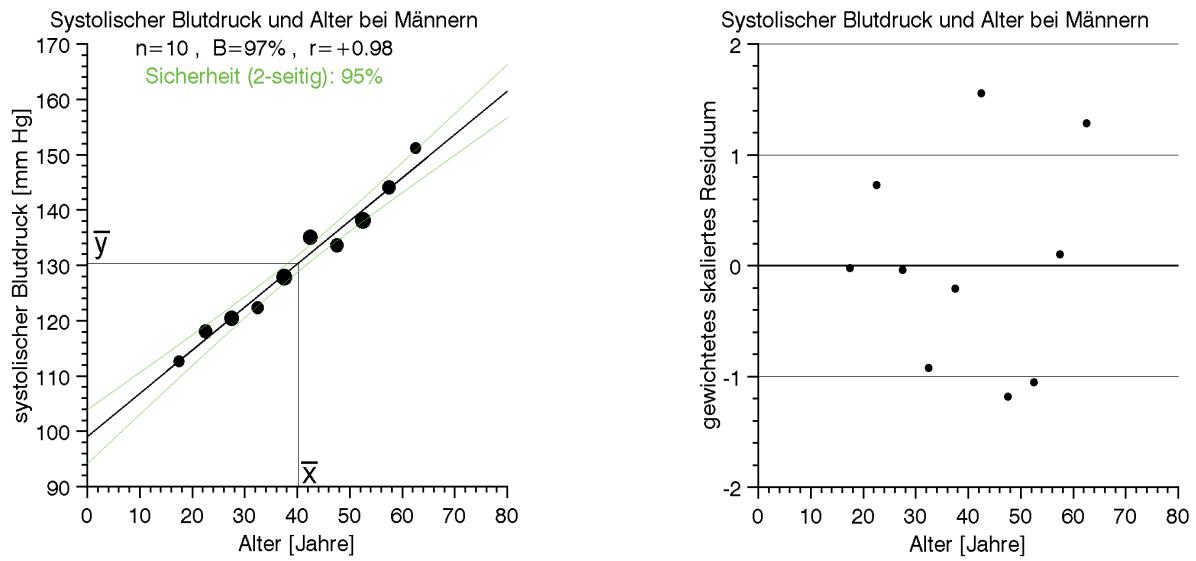


Abb. 4: Blutdruck und Alter

- links:* Punkte mit geschätzter Regressionsgeraden  $\hat{f}$  und den Funktionen  $\hat{f}_{un}$  und  $\hat{f}_{ob}$  für die zweiseitigen 95%-Konfidenzgrenzen (vgl. Text)
- rechts:* zugehöriger Residuenplot.

### 3.7 Anpassungstests: Chiadrat- und Likelihood-Quotienten-Test

Im Abschnitt 3.2 haben wir uns mit dem Testen einer *einzelnen* Wahrscheinlichkeit für ein interessierendes Ziel-Ereignis beschäftigt. Wir wollen jetzt die Wahrscheinlichkeit *mehrerer Ereignisse* gleichzeitig betrachten, Hypothesen über sie formulieren und diese dann mit einem sogenannten *Anpassungstest* überprüfen. Zur Einführung in die Problematik beginnen wir mit zwei Beispielen.

#### **Kreuzung von Monohybriden** (Kinder-Osius-Timm, 1982, Beispiel 7.25)

Man vermutet, daß die Blütenfarbe der japanischen Wunderblume (*Mirabilis jalapa*) von einem Gen mit den Allelen  $R, W$  und den dazugehörigen Phänotypen  $RR$  (rot),  $RW$  (rosa),  $WW$  (weiß) gesteuert wird. Zur Überprüfung des Erbganges wurden Monohybriden  $RW$  miteinander gekreuzt und die Aufteilung der Phänotypen aller  $n = 108$  Nachkommen ist in Tab. 1 angegeben.

Phänotyp	$RR$	$RW$	$WW$	insgesamt
Anzahl	30	47	31	108

Tabelle 1: Ergebnisse eines Kreuzungsexperiments (zu Details vgl. Text)

Nach den Mendelschen Gesetzen müssen die erwarteten Anzahlen für die Phänotypen  $RR, RW, WW$  im Verhältnis  $1 : 2 : 1$  stehen. Dies läßt sich als Nullhypothese über die Wahrscheinlichkeiten der einzelnen Phänotypen wie folgt formulieren:

$$H_0: \quad p_{RR} = \frac{1}{4}, \quad p_{RW} = \frac{1}{2}, \quad p_{WW} = \frac{1}{4} \quad (\text{Mendelsches Gesetz}). \quad (1)$$

Diese Nullhypothese soll unter Verwendung des obigen Kreuzungsexperiments überprüft werden, wobei mit  $\alpha = 10\%$  bewußt ein etwas höheres Risiko 1. Art gewählt werden soll, um das Fehlerrisiko 2. Art geringer zu halten.  $\square$

#### **Hardy-Weinberg Gleichgewicht** (Kinder-Osius-Timm, 1982, Beispiel 7.26)

Beim Haptoglobin-Blutgruppensystem werden die drei Blutgruppen-Phänotypen  $Hp\ 1-1$ ,  $Hp\ 1-2$  und  $Hp\ 2-2$  durch die entsprechende Kombination zweier Allele  $Hp^1$  und  $Hp^2$  bestimmt. Wir interessieren uns dafür, ob in einer bestimmten Population das *genetische Gleichgewicht* vorliegt. Bezeichnen  $p_1$  und  $p_2$  die Frequenzen beider Allele  $Hp^1$  und  $Hp^2$ , und sind  $p_{11}, p_{12}, p_{22}$  die Frequenzen der drei Phänotypen in der Population, so läßt sich das Gleichgewicht nach Hardy-Weinberg wie folgt als Nullhypothese formulieren:

$$H_0: \quad p_{11} = p_1^2, \quad p_{12} = 2p_1p_2, \quad p_{22} = p_2^2 \quad (\text{genetisches Gleichgewicht}). \quad (2)$$

Die Nullhypothese soll zum Niveau  $\alpha = 5\%$  überprüft werden. Hierzu wird aus der Population eine Stichprobe vom Umfang  $n = 500$  gezogen und die beobachteten Anzahlen für die drei Phänotypen sind in Tab. 2 angegeben.

Phänotyp	$Hp\ 1\text{-}1$	$Hp\ 1\text{-}2$	$Hp\ 2\text{-}2$	insgesamt
Anzahl	61	258	182	500

Tabelle 2: Haptoglobin-Blutgruppen einer Stichprobe (zu Details vgl. Text)

□

**Allgemeine Ausgangssituation (Modell):** Als Verallgemeinerung der Situation in obigen Beispielen betrachten wir einen stochastischen Vorgang (Zufallsexperiment), dessen möglichen Ergebnisse in  $K \geq 2$  Klassen (bzw. Ereignisse) eingeteilt sind (in obigen beiden Beispielen sind es jeweils  $K = 3$  Phänotypen). Hierbei soll jedes Ergebnis in genau einer Klasse liegen, d.h. die zugehörigen Ereignisse  $A_1, \dots, A_K$  sind disjunkt. Bezeichnet  $p_k = P(A_k)$  die Wahrscheinlichkeit für das zugehörige Ereignis  $A_k$  (d.h. das Ergebnis fällt in die Klasse  $k$ ), so ist die Summe aller Klassenwahrscheinlichkeiten gleich 1:

$$\sum_{k=1}^K p_k = p_1 + p_2 + \dots + p_K = 1. \quad (1)$$

Typischerweise sind die Wahrscheinlichkeiten  $p_k$  unbekannt und man will Hypothesen über sie testen. Hierzu werden  $n$  unabhängigen Wiederholungen (des Experiments) durchgeführt und die beobachteten absoluten Häufigkeiten der Klasse  $k$  mit  $b_k$  bezeichnet ( $b$  steht hier für beobachtet). Die relative Klassenhäufigkeit

$$\hat{p}_k = \frac{1}{n} b_k \quad \text{für } k = 1, \dots, K \quad (2)$$

ist dann eine Schätzung der unbekannten Wahrscheinlichkeiten  $p_k$  (vgl. 3.1.1).

Zum Überprüfen einer Nullhypothese über die Wahrscheinlichkeiten  $p_1, \dots, p_K$  unterscheiden wir zwischen *einfachen* und *zusammengesetzten* Hypothesen. Bei einer *einfachen* Nullhypothese werden die Wahrscheinlichkeiten vollständig durch vorgegebene „Sollwerte“ spezifiziert, wie z.B. im Mendelschen Gesetz (1). Dagegen spezifiziert eine *zusammengesetzte* Nullhypothese nur die Struktur der Wahrscheinlichkeiten  $p_k$  in dem Sinn, daß die Wahrscheinlichkeiten noch von unbekannten Parametern abhängen, z.B. beim Hardy-Weinberg-Gleichgewicht (2) von den unbekannten Allelfrequenzen  $p_1$  und  $p_2$ . Nullhypothesen mit unbekannten Parametern tauchen auch bei der Analyse von Kontingenztafeln auf (vgl. 3.8).

### 3.7.1 Einfache Nullhypotesen

Eine *einfache Nullhypothese*  $H_0$  spezifiziert die Klassenwahrscheinlichkeit  $p_k$  *vollständig* durch vorgegebene „Sollwerte“ (Referenzwerte)  $p_{0k}$  wie folgt:

$$\begin{aligned} H_0: p_1 &= p_{01}, \quad p_2 = p_{02}, \quad \dots, \quad p_K = p_{0K} && \text{bzw.} \\ H_0: p_k &= p_{0k} && \text{für alle Klassen } k = 1, \dots, K. \end{aligned} \quad (1)$$

Die *Hypothese H* ist die Negation der Nullhypothese

$$H: p_k \neq p_{0k} \quad \text{für mindestens eine Klasse } k. \quad (2)$$

Die *unter der Nullhypothese erwartete* Häufigkeit  $e_k$  der Klasse  $k$  ist gegeben durch ( $e$  steht hier für *erwartet*)

$$e_k = n \cdot p_{0k} \quad (\text{unter } H_0 \text{ erwartete Anzahl für Klasse } k). \quad (3)$$

Eine Maßzahl für die Abweichung der beobachteten von den unter der Nullhypothese erwarteten Anzahlen ist die *Chiadrat- oder Pearson-Statistik*

$$X^2 = \sum_{k=1}^K \frac{(b_k - e_k)^2}{e_k} = n \cdot \sum_{k=1}^K \frac{(\hat{p}_k - p_{0k})^2}{p_{0k}} \quad (\text{Chiadrat-Statistik}). \quad (4)$$

Jeder einzelne Summand von  $X^2$  beurteilt die Abweichung der *beobachteten* Anzahlen  $b_k$  von den *unter der Nullhypothese erwarteten* Anzahlen  $e_k$  bzw. die zugehörige Abweichung der Schätzung  $\hat{p}_k$  vom Sollwert  $p_{0k}$ . Je größer diese Abweichung ist, desto größer wird der Summand, und somit sprechen *große* Werte von  $X^2$  gegen die Nullhypothese. Unter der Nullhypothese hat die Teststatistik für *nicht zu kleinen Umfang n approximativ* eine sogenannte *Chiadrat-Verteilung* mit  $m = K - 1$  *Freiheitsgraden* (vgl. 3.7.4), wobei die Approximation für  $n \rightarrow \infty$  beliebig genau wird. Mit dem (oberen)  $\alpha$ -Quantil  $\chi_{m;\alpha}^2$  der  $\chi_m^2$ -Verteilung (vgl. Anhang T) ergibt sich dann (vgl. auch Abb. 1) der

*Chiadrat-Test zum Niveau  $\alpha$  (Fehlerrisiko 1. Art):*

$$\text{Ablehnung von } H_0 \Leftrightarrow X^2 \geq \chi_{m;\alpha}^2. \quad (5)$$

Beim Freiheitsgrad

$$m = K - 1 \quad (\text{Freiheitsgrad}) \quad (6)$$

wird die Anzahl  $K$  aller Klassen um 1 reduziert, weil sich die Häufigkeit  $b_K$  der *letzten* Klasse wegen  $n = \sum_k b_k$  bereits aus den Häufigkeiten  $b_1, \dots, b_{K-1}$  der ersten  $K-1$  Klassen ergibt, und somit nur  $K-1$  beobachtete Häufigkeiten „frei“ (d.h. voneinander unabhängig) variieren können.

Der sogenannte *P-Wert* (oder das *Signifikanzniveau*) des Testwerts  $X^2$  ist die *unter der Nullhypothese* berechnete Wahrscheinlichkeit dafür, daß die Chiquadrat-Statistik *mindestens* so groß ist wie der *beobachtete* Wert  $X^2$ . Unter Verwendung der Verteilungsfunktion  $\Phi_m$  der  $\chi^2_m$ -Verteilung läßt sich der *P-Wert* approximativ berechnen als (vgl. 3.7.4 sowie Abb. 1):

$$P\{\chi^2_m \geq X^2\} = 1 - \Phi_m(X^2) \quad (\text{P-Wert von } X^2). \quad (7)$$

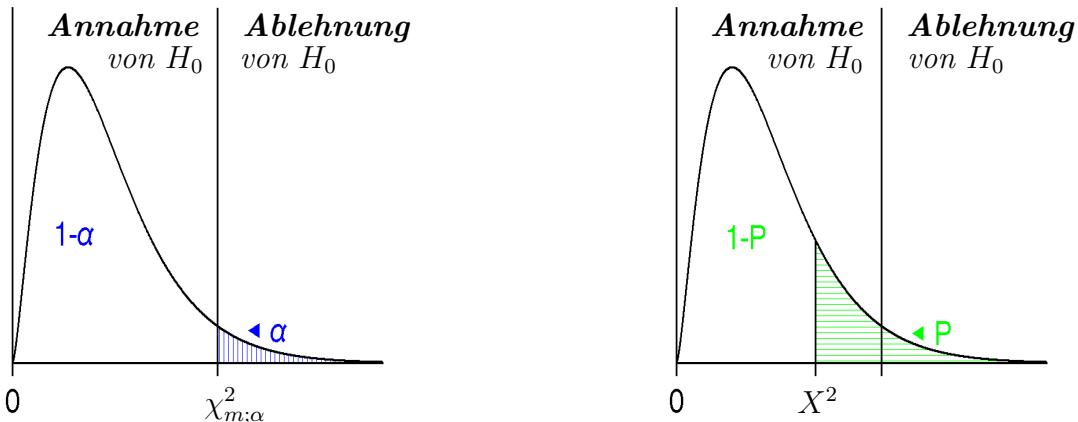


Abb. 1: Ablehnungsbereich und *P-Wert* beim Chiquadrat-Test  
 links: Dichte der (approximativ) Chiquadrat-verteilten Teststatistik  $X^2$  unter der Nullhypothese mit  $\alpha$ -Quantil für  $\alpha = 5\%$ .  
 rechts: Der zum Testwert  $X^2$  gehörige *P-Wert*  $P$  als markierte Fläche.

Die Test-Entscheidung läßt sich mit dem *P-Wert* äquivalent überprüfen, weil

$$\text{Testwert } X^2 \geq \chi^2_{m;\alpha} \Leftrightarrow \text{P-Wert } 1 - \Phi_m(X^2) \leq \alpha. \quad (8)$$

Der Chiquadrat-Test ist ein *asymptotischer* Test, d.h. das Fehlerrisiko 1. Art ist nur *approximativ* gleich dem vorgegebenen Niveau  $\alpha$ , wobei die Approximation für wachsenden Stichprobenumfang  $n$  immer besser wird. Bei der Anwendung des Tests soll daher der Umfang  $n$  und damit auch die erwarteten Anzahlen  $e_k$  nicht zu *klein* sein, und hierfür gibt es verschiedene Faustregeln:

*Konservative Faustregel:* Alle erwarteten Anzahlen  $e_k$  sind *mindestens*  $\geq 5$ .

*Liberale Faustregel:* Der Mittelwert  $\frac{n}{K}$  aller erwarteten Anzahlen ist

- für  $\alpha \geq 5\% : \text{mindestens } \geq 2,$
- für  $1\% \leq \alpha < 5\% : \text{mindestens } \geq 4.$

Neben  $X^2$  wird auch der folgende *Likelihood-Quotienten-Testwert* (kurz: *LQ-Testwert*) als Maß für die Abweichung der beobachteten Anzahlen  $b_k$  von den erwarteten Anzahlen  $e_k$  verwendet:

$$G^2 = 2 \sum_{k=1}^K b_k \cdot \ln \left( \frac{b_k}{e_k} \right) = 2n \cdot \sum_{k=1}^K \hat{p}_k \cdot \ln \left( \frac{\hat{p}_k}{p_{0k}} \right) \geq 0 \quad (\text{LQ-Testwert}), \quad (9)$$

wobei für  $b_k = 0$  der entsprechende Summand gleich Null gesetzt wird. Man beachte, daß zwar einzelne Summanden von  $G^2$  auch negativ sein können (wenn  $e_k > b_k$  ist), aber man zeigen kann, daß die Gesamtsumme  $G^2$  stets  $\geq 0$  ist.

Große Werte von  $G^2$  sprechen wieder *gegen* die Nullhypothese (was allerdings nicht *direkt* einsichtig ist) und unter der Nullhypothese hat die Teststatistik auch (approximativ) eine Chiadrat-Verteilung mit  $m = K - 1$  Freiheitsgraden. Deshalb ergibt sich analog zum Chiadrat-Test der

*Likelihood-Quotienten-Test zum Niveau (Fehlerrisiko 1. Art)  $\alpha$ :*

$$\begin{aligned} \text{Ablehnung von } H_0 &\Leftrightarrow \text{Testwert} & G^2 &\geq \chi_{m;\alpha}^2 \\ &\Leftrightarrow P\text{-Wert} & 1 - \Phi_m(G^2) &\leq \alpha. \end{aligned} \quad (10)$$

Auch der Likelihood-Quotient-Test hat (wie der Chiadrat-Test) nur *asymptotisch* das vorgegebene Fehlerrisiko  $\alpha$ .

Der Chiadrat- und der Likelihood-Quotienten-Test führen bei nicht zu kleinem Umfang  $n$  meist zu gleichen Test-Entscheidungen, und bei Gültigkeit der Nullhypothese unterscheiden sich die Testwerte  $X^2$  und  $G^2$  nur geringfügig. Daher wird meist der einfacher zu berechnende und besser zu interpretierende Testwert  $X^2$  verwendet.

**Kreuzung von Monohybriden (1. Fortsetzung):** Das Mendelsche Gesetz 3.7(1) ist eine *einfache* Nullhypothese mit folgenden Sollwerten für die Wahrscheinlichkeiten der Phänotypen

$$\begin{aligned} H_0: \quad p_{RR} &= p_{0RR}, & p_{RW} &= p_{0RW}, & p_{WW} &= p_{0WW} \\ \text{mit} \quad p_{0RR} &= \frac{1}{4}, & p_{0RW} &= \frac{1}{2}, & p_{0WW} &= \frac{1}{4}. \end{aligned}$$

Die Nullhypothese soll zum Niveau  $\alpha = 10\%$  geprüft werden. Die Berechnung der Teststatistiken  $X^2$  und  $G^2$  für die Daten von Tab. 1 ist in Tab. 3 zusammengestellt.

Phänotyp	beobachtet $b$	erwartet $e$	$X^2$ -Summand $(b - e)^2/e$	$G^2$ -Summand $2 b \cdot \ln(b/e)$
$RR$ (rot)	30	27	0,33	6,32
$RW$ (rosa)	47	54	0,91	-13,05
$WW$ (weiß)	31	27	0,59	8,57
Summe	108	108	$X^2 = 1,83$	$G^2 = 1,84$

Tabelle 3: Berechnung von  $X^2$  und  $G^2$  für die Daten aus Tab. 1.

Wegen  $m = 3 - 1 = 2$  und  $\chi_{2;\alpha}^2 = 4,605$  wird die Mendelsche Hypothese weder vom Chiadrat- noch vom Likelihood-Quotienten-Test abgelehnt (normalerweise führt man nur *einen* der beiden Tests durch anstatt – wie hier zur Demonstration – *beide*). Die zugehörigen (mit Computer berechneten)  $P$ -Werte sind deshalb auch beide *größer* als das Niveau  $\alpha = 10\%$ .

$$1 - \Phi_2(X^2) = 40,1\% \quad 1 - \Phi_2(G^2) = 39,9\%.$$

Da ein Fehler 2. Art vorliegen kann, ist die Mendelsche Hypothese durch das Experiment nicht mit *quantifizierbarer* Sicherheit bestätigt (sondern nur *unwiderlegt*). Das Fehlerrisiko 2. Art wird in 3.7.3 a näher untersucht.  $\square$

### 3.7.2 Nullhypotesen mit unbekannten Parametern

Wir lassen jetzt zu, daß die durch die Nullhypothese spezifizierten Sollwerte  $p_{0k}$  nicht durch konkrete Zahlenwerte gegeben sind, sondern über vorgegebene Funktionen  $g_k$  von einem oder mehreren *unbekannten Parametern*  $\theta_1, \theta_2, \dots, \theta_S$  abhängen können:

$$p_{0k} = g_k(\theta_1, \theta_2, \dots, \theta_S) \quad \text{für } k = 1, \dots, K. \quad (1)$$

Die S Parameter sollen hierbei *unabhängig voneinander* variieren können, d.h. jeder Parameter kann *einzel*n verändert werden, ohne daß sich gleichzeitig *andere* Parameter mit ändern. Die Nullhypothese lautet dann

$$H_0: p_k = p_{0k} \quad \text{mit} \quad p_{0k} = g_k(\theta_1, \theta_2, \dots, \theta_S) \quad \text{für alle } k = 1, \dots, K \quad (2)$$

für geeignete Werte der Parameter. Die *Alternative H* ist die *Negation* von  $H_0$ , d.h. für *jede* Wahl der Parameter gilt

$$H: p_k \neq p_k(\theta_1, \theta_2, \dots, \theta_S) \quad \text{für mindestens ein } k = 1, \dots, K.$$

**Hardy-Weinberg Gleichgewicht (Fortsetzung):** Die Nullhypothese

$$H_0: p_{11} = p_1^2, \quad p_{12} = 2p_1p_2, \quad p_{22} = p_2^2$$

hängt hier von den Frequenzen  $p_1$  und  $p_2$  beider Allele  $Hp^1$  und  $Hp^2$  ab. Wegen  $p_2 = 1 - p_1$  läßt sich die Nullhypothese auch ohne Verwendung von  $p_2$  formulieren:

$$H_0: p_{11} = p_1^2, \quad p_{12} = 2p_1(1 - p_1), \quad p_{22} = (1 - p_1)^2.$$

Hier haben wir es also mit nur *einem* Parameter  $\theta_1 = p_1$  zu tun (den wir aber suggestiv mit  $p_1$  statt  $\theta_1$  bezeichnen), und die zugehörigen Funktionen lauten

$$g_{11}(p_1) = p_1^2 \quad g_{12}(p_1) = 2p_1(1 - p_1), \quad g_{22}(p_1) = (1 - p_1)^2. \quad \square$$

Bei der Analyse von Kontingenztafeln (vgl. 3.8) spielen Nullhypotesen mit Parametern eine wichtige Rolle und dort finden sich auch weitere Anwendungen.

Um (analog zu *einfachen* Nullhypotesen in 3.7.1) auch hier *erwartete* Häufigkeiten berechnen können, muß man die unbekannten Parameter zuerst aus den beobachteten Daten *schätzen*. Wir wollen hier nicht darauf eingehen, wie man *ganz allgemein* (d.h. für beliebige Funktionen  $g_k$ ) solche Schätzungen erhält, sondern diese Schätzungen nur für konkrete Anwendungssituationen (und den dort verwendeten Funktionen  $g_k$ ) angeben, z.B. für das Hardy-Weinberg-Gleichgewicht oben sowie allgemein für die Analyse von Kontingenztafeln im Kapitel 3.8. Deshalb gehen wir hier erst einmal davon aus, daß wir die Schätzungen  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S$  für die Parameter bereits bestimmt haben. Durch Einsetzen dieser Schätzungen in die Funktionen  $g_k$  erhält man die zugehörigen *geschätzten* Klassen-Wahrscheinlichkeiten (unter  $H_0$ )

$$\hat{p}_{0k} = g_k(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S) \quad (\text{geschätzte Wahrscheinlichkeit unter } H_0). \quad (3)$$

Hieraus ergeben sich die (geschätzten) *erwarteten* Klassenhäufigkeiten

$$e_k = n \cdot \hat{p}_{0k} = n \cdot g_k(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S) \quad (\text{unter } H_0 \text{ erwartete Anzahl}). \quad (4)$$

Als Maßzahl für die Abweichung der beobachteten von den erwarteten Häufigkeiten berechnet man wieder die *Chiquaret-* oder *Likelihood-Quotienten-Statistik* aus 3.7.1 (vgl. Test-Box 6b). Beide Statistiken sind unter der Nullhypothese jeweils approximativ Chiquaret-verteilt, allerdings hier mit dem Freiheitsgrad

$$m = (K - 1) - S \quad (\text{Freiheitsgrad}). \quad (5)$$

Allgemein wird der Freiheitsgrad nach folgender Regel bestimmt:

$$\text{Freiheitsgrad} = \quad (6)$$

Anzahl voneinander unabhängiger beobachteter Anzahlen abzüglich  
Anzahl voneinander unabhängiger geschätzter Parameter.

Der Chiadrat- bzw. Likelihood-Quotienten-Test ergeben sich mit diesem Freiheitsgrad  $m$  wie in **3.7.1** zu

*Chiadrat-Test zum Niveau (Fehlerrisiko 1. Art)  $\alpha$ :*

$$\begin{aligned} \text{Ablehnung von } H_0 &\Leftrightarrow \text{Testwert } X^2 \geq \chi_{m;\alpha}^2 \\ &\Leftrightarrow P\text{-Wert } 1 - \Phi_m(X^2) \leq \alpha. \end{aligned} \quad (7)$$

*Likelihood-Quotienten-Test zum Niveau (Fehlerrisiko 1. Art)  $\alpha$ :*

$$\begin{aligned} \text{Ablehnung von } H_0 &\Leftrightarrow \text{Testwert } G^2 \geq \chi_{m;\alpha}^2 \\ &\Leftrightarrow P\text{-Wert } 1 - \Phi_m(G^2) \leq \alpha. \end{aligned} \quad (8)$$

**Hardy-Weinberg Gleichgewicht (2. Fortsetzung):** Als Schätzungen für die unbekannten Allelfrequenzen verwenden wir die relativen Häufigkeiten der Allele unter den insgesamt  $2n = 1000$  Allelen in der Stichprobe

$$\hat{p}_1 = \frac{1}{2n}(2b_{11} + b_{12}) = 0,38 \quad \text{und} \quad \hat{p}_2 = 1 - \hat{p}_1 = 0,62.$$

Unter Verwendung der relativen Häufigkeiten für die Phänotypen lässt sich die Schätzung auch darstellen als

$$\hat{p}_1 = \frac{1}{2}(2\hat{p}_{11} + \hat{p}_{12}) \quad \text{mit} \quad \hat{p}_{11} = \frac{1}{n}b_{11}, \quad \hat{p}_{12} = \frac{1}{n}b_{12}. \quad (9)$$

Die geschätzten Wahrscheinlichkeiten im Gleichgewichtsfall  $H_0$  sind dann

$$\hat{p}_{011} = \hat{p}_1^2 = 0,1444, \quad \hat{p}_{012} = 2\hat{p}_1\hat{p}_2 = 0,4712, \quad \hat{p}_{022} = \hat{p}_2^2 = 0,3844.$$

Die Nullhypothese soll zum Niveau  $\alpha = 5\%$  überprüft werden. Die Berechnung der Teststatistiken  $X^2$  und  $G^2$  für die Daten von Tab. 2 ist in Tab. 4 zusammengestellt.

Phänotyp	beobachtet	erwartet	$X^2$ -Summand	$G^2$ -Summand
	$b$	$e$	$(b - e)^2/e$	$2 b \cdot \ln(b/e)$
$Hp\ 1\text{-}1$	61	72,2	1,737	-20,565
$Hp\ 1\text{-}2$	258	235,6	2,130	40,865
$Hp\ 2\text{-}2$	181	192,2	0,653	-21,734
Summe	500	500,0	$X^2 = 4,520$	$G^2 = 4,566$

Tabelle 4: Berechnung von  $X^2$  und  $G^2$  für die Daten aus Tab. 2.

Die Anzahl der voneinander *unabhängigen* Parameter ist hier 1, da der vermeintlich „zweite“ Parameter  $p_2 = 1 - p_1$  durch den Parameter  $p_1$  bereits bestimmt ist. Also ist der Freiheitsgrad  $m = 3 - 1 - 1 = 1$ . Das 5%-Quantil  $\chi^2_{1;5\%} = 3,841$  wird von beiden Testwerten überschritten, d.h. die Nullhypothese (Gleichgewicht) wird sowohl vom Chiquaret- als auch vom Likelihood-Quotienten-Test *abgelehnt* (normalerweise führt man nur *einen* der beiden Tests durch anstatt – wie hier zur Demonstration – *beide*). Die zugehörigen (auf dem Computer berechneten)  $P$ -Werte sind deshalb beide *kleiner* als das Niveau  $\alpha = 5\%$ :

$$1 - \Phi_2(X^2) = 3,4\%, \quad 1 - \Phi_2(G^2) = 3,3\%.$$

Es kann hier nur ein Fehler 1. Art vorliegen, der durch  $\alpha = 5\%$  kontrolliert ist.  $\square$

### 3.7.3 Fehlerrisiko 2. Art und Versuchsplanung

Das *Fehlerrisiko 2. Art* und die Testschärfe des Chiquaret- und Likelihood-Quotienten-Tests in **3.7.1** bzw. **3.7.2** hängen von den unbekannten Klassenwahrscheinlichkeiten  $p_k$  ab und können unter Verwendung der Verteilungsfunktion  $\Phi_{m,\delta}$  der *nichtzentralen* Chiquaret-Verteilung mit Freiheitsgrad  $m$  und *Nichtzentralität*  $\delta$  wie folgt *approximiert* werden (vgl. auch Kinder-Osius-Timm, 1982, 7.4.2). Allerdings steht diese Verteilungsfunktion  $\Phi_{m,\delta}$  bisher *nicht* in Tabellenkalkulations-Programmen (wie z.B. *Excel*) direkt zur Verfügung, sondern lässt sich nur mit Hilfe entsprechender Zusatzmodule (vgl. die *Excel-Tipps* am Ende des Kapitels) oder spezieller Statistik-Software berechnen (auf die hier nicht eingegangen wird). Für eine *grobe* Bestimmung des Fehlerrisikos kann man auch Graphiken (z.B. Abb. 2) verwenden.

#### 3.7.3 a Chiquaret-Test für *einfache* Nullhypotesen

Für das Testproblem mit *einfacher Nullhypothese*

$H_0: p_k = p_{0k}$	für alle $k = 1, \dots, K$ ,	(1)
$H: p_k \neq p_{0k}$	für mindestens ein $k = 1, \dots, K$	

quantifizieren wir zunächst die *Abweichung* aller *wahren* Klassen-Wahrscheinlichkeiten  $p_k$  von den zugehörigen „Sollwerten“  $p_{0k}$  der Nullhypothese durch folgende Maßzahl

$$\Delta = \sum_{k=1}^K \frac{(p_k - p_{0k})^2}{p_{0k}} \quad (\text{Abweichung von } H_0). \quad (2)$$

Da der  $k$ -te Summand von  $\Delta$  genau dann gleich Null ist, wenn  $p_k = p_{0k}$  gilt, lassen sich die Hypothesen auch äquivalent formulieren durch

$$H_0: \Delta = 0 \quad \text{vs.} \quad H: \Delta > 0. \quad (3)$$

Multipliziert man die Abweichung  $\Delta$  mit dem Stichprobenumfang  $n$  so ergibt sich die sogenannte *Nichtzentralität*

$$\delta = n \cdot \Delta = n \sum_{k=1}^K \frac{(p_k - p_{0k})^2}{p_{0k}} \quad (\text{Nichtzentralität}). \quad (4)$$

Die Nichtzentralität ist völlig analog zur Chiquaret-Statistik aufgebaut: statt der *Schätzungen*  $\hat{p}_k$  in  $X^2$  stehen in  $\delta$  die *wahren Wahrscheinlichkeiten*  $p_k$ .

Beim *Chiquaret-Test* mit dem Niveau  $\alpha$  und dem Freiheitsgrad  $m$  ergeben sich die *Testschärfe*  $\gamma$  und das *Fehlerrisiko 2. Art*  $\beta$  bei *nicht zu kleinem* Stichprobenumfang  $n$  unter Verwendung der Verteilungsfunktion  $\Phi_{m,\delta}$  der *nichtzentralen* Chiquaret-Verteilung (mit Freiheitsgrad  $m$  und *Nichtzentralität*  $\delta$ ) zu

$$\gamma = \gamma(\delta) \approx 1 - \Phi_{m,\delta}(\chi^2_{m;\alpha}) \quad (\text{Testschärfe}), \quad (5)$$

$$\beta = \beta(\delta) \approx \Phi_{m,\delta}(\chi^2_{m;\alpha}) \quad (\text{Fehlerrisiko 2. Art}). \quad (6)$$

Ersetzt man in der Abweichung  $\Delta$  die *wahren* (unbekannten) Klassenwahrscheinlichkeiten  $p_k$  durch die *beobachteten* relativen Häufigkeiten  $\hat{p}_k$ , so erhält man die *beobachtete* Abweichung

$$\hat{\Delta} = \sum_{k=1}^K \frac{(\hat{p}_k - p_{0k})^2}{p_{0k}} = \frac{1}{n} X^2 \quad (\text{geschätzte Abweichung von } H_0) \quad (7)$$

als *Schätzung* von  $\Delta$ . Als zugehörige *Schätzung* der Nichtzentralität  $\delta$  ergibt sich

$$\hat{\delta} = n \cdot \hat{\Delta} = X^2 \quad (\text{geschätzte Nichtzentralität}). \quad (8)$$

Bei konkreten Anwendungen bestimmt man das Fehlerrisiko  $\beta$  typischerweise für *relevante* Abweichungen von der Nullhypothese, d.h. man gibt sich praktisch relevante *alternative* Wahrscheinlichkeiten  $p_k$  bzw. die zugehörigen Abweichungen  $d_k = p_k - p_{0k}$  vor und bestimmt  $\beta(\delta)$  für die zugehörige *Nichtzentralität*  $\delta > 0$ . Hierbei dürfen auch *einige* – aber natürlich nicht *alle* – Abweichungen gleich *Null* sein.

Das approximative Fehlerrisiko  $\beta(\delta) \approx \Phi_{m,\delta}(\chi^2_{m;\alpha})$  ist in Abb. 2 als Funktion der Nichtzentralität  $\delta$  für  $\alpha = 10\%$  und  $\alpha = 5\%$  und verschiedene Freiheitsgrade  $m$  dargestellt. Man erkennt daran, daß  $\beta(\delta)$  geringer wird, wenn  $\delta$  größer wird, d.h. bei wachsendem Umfang  $n$  oder wachsender Abweichung  $\Delta$ . Aus diesen Abbildungen kann man  $\beta(\delta)$  mit einer oft ausreichenden Genauigkeit ablesen.

Die *Berechnung* der Verteilungsfunktion  $\Phi_{m,\delta}$  (und damit des Fehlerrisikos bzw. der Testschärfe) auf dem Computer erfordert spezielle Programme oder entsprechende Zusatzmodule (Makros) für Tabellenkalkulationen (vgl. die *Excel-Tipps* am Ende des Kapitels).

Für die *Versuchsplanaung* ermittelt man den erforderlichen *Mindestumfang*  $n$ , bei dem für eine *vorgegebene relevante Abweichung*  $\Delta$  das Fehlerrisiko 2. Art (höchstens) den *vorgegebenen Wert*  $\beta$  erreicht, wie folgt. Zunächst bestimmt man aus Abb. 2 oder mit Computer dasjenige  $\delta = \delta(\beta)$ , welches dem vorgegebenen Wert  $\beta$  nach (6) entspricht. Dann ergibt sich der Mindestumfang  $n$  aus (4) zu

$$n \approx \frac{\delta(\beta)}{\Delta} \quad (\text{erforderlicher Mindestumfang}). \quad (9)$$

**Kreuzung von Monohybriden (2. Fortsetzung):** Wir wollen das Fehlerrisiko 2. Art (mit der Power-Box 6a) für den Fall bestimmen, daß die 3 Phänotypen *gleichverteilt* sind, d.h.  $p_{WW} = p_{RW} = p_{RR} = \frac{1}{3}$ . Aus der zugehörigen Abweichung

$$\Delta = \frac{(\frac{1}{3} - \frac{1}{4})^2}{\frac{1}{4}} + \frac{(\frac{1}{3} - \frac{1}{2})^2}{\frac{1}{2}} + \frac{(\frac{1}{3} - \frac{1}{4})^2}{\frac{1}{4}} = \frac{1}{9}$$

erhält man die Nichtzentralität  $\delta = n\Delta = 12$  und aus Abb. 2 liest man (für die *Gleichverteilung* der Phänotypen) das relativ geringe Fehlerrisiko von  $\beta(12) \approx 6,5\%$  ab (der mit Computer berechnete genauere Wert ist 6,62%).  $\square$

### 3.7.3 b Chiadrat-Test für Nullhypotesen mit Parametern

Für den *Chiadrat-Test der Nullhypothese mit unbekannten Parametern*

$$H_0: p_k = p_{0k} \quad \text{mit} \quad p_{0k} = g_k(\theta_1, \theta_2, \dots, \theta_S) \quad \text{für alle } k = 1, \dots, K. \quad (1)$$

ergeben sich das Fehlerrisiko 2. Art und die Testschärfe formal wie in 3.7.3 a (5) und (6), wobei allerdings die Sollwerte  $p_{0k}$  in  $\Delta$  durch „Schätzungen“  $\hat{p}_{0k} = g_k(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S)$

ersetzt werden:

$$\Delta = \sum_{k=1}^K \frac{(p_k - \hat{p}_{0k})^2}{\hat{p}_{0k}} \quad (\text{Abweichung von } H_0). \quad (2)$$

Die „geschätzten Parameter“  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S$  werden hierzu nach der gleichen Methode bestimmt wie bei einer konkreten Stichprobe, wobei man aber hier eine „ideale“ Stichprobe zu Grunde legt, deren relative Häufigkeiten  $\hat{p}_k$  exakt mit den wahren Klassen-Wahrscheinlichkeit  $p_k$  übereinstimmen, d.h. man verwendet die „idealen“ Beobachtungen  $b_k = n p_k$  bzw. die „idealen“ relativen Häufigkeiten  $\hat{p}_k = p_k$ .

Der Mindestumfang  $n$  bei einer Versuchplanung ergibt sich wie in **3.7.3 a** (9), aber mit der Abweichung  $\Delta$  aus (2).

**Hardy-Weinberg Gleichgewicht (3. Fortsetzung):** Bevor wir mit konkreten Zahlen arbeiten, wollen wir uns überlegen, wie man hier die Schätzung  $\hat{p}_1$  des Parameters  $p_1$  bestimmt. Für die „formalen Beobachtungen“

$$b_{11} = n p_{11}, \quad b_{12} = n p_{12}, \quad b_{22} = n p_{22}$$

ergeben sich die wahren Wahrscheinlichkeiten als zugehörige relative Häufigkeiten der Phänotypen

$$\hat{p}_{11} = p_{11}, \quad \hat{p}_{12} = p_{12}, \quad \hat{p}_{22} = p_{22}.$$

Nach **3.7.2** (9) erhält man die gesuchte Schätzung von  $p_1$  als

$$\hat{p}_1 = \frac{1}{2} (2p_{11} + p_{12}). \quad (3)$$

Wir wollen jetzt die Abweichung  $\Delta$  für die konkrete Alternative mit *gleichverteilten Phänotypen* bestimmen, d.h. für  $p_{11} = p_{12} = p_{22} = \frac{1}{3}$ . Nach (3) ist

$$\hat{p}_1 = \frac{1}{2} \left( 2 \cdot \frac{1}{3} + \frac{1}{3} \right) = \frac{1}{2},$$

und die zugehörigen geschätzten Wahrscheinlichkeiten *im Gleichgewichtsfall* sind

$$\hat{p}_{011} = \hat{p}_1^2 = \frac{1}{4}, \quad \hat{p}_{012} = 2\hat{p}_1\hat{p}_2 = \frac{1}{2}, \quad \hat{p}_{022} = \hat{p}_2^2 = \frac{1}{4}.$$

Also ist die Abweichung von  $H_0$  gegeben durch

$$\Delta = \frac{\left(\frac{1}{3} - \frac{1}{4}\right)^2}{\frac{1}{4}} + \frac{\left(\frac{1}{3} - \frac{1}{2}\right)^2}{\frac{1}{2}} + \frac{\left(\frac{1}{3} - \frac{1}{4}\right)^2}{\frac{1}{4}} = \frac{1}{9}.$$

Als Beispiel einer Versuchsplanung wollen wir (mit der Planungs-Box 6b) den *Mindestumfang  $n$*  bestimmen, für den der Chiadrat-Test mit  $\alpha = 5\%$  bei *gleichverteilten* Phänotypen ein Fehlerrisiko 2. Art von  $\beta = 10\%$  hat. Aus Abb. 2 (unten) entnimmt man die zugehörige Nichtzentralität  $\delta(10\%) \approx 10,5$  (der mit Computer berechnete genauere Wert ist 10,507). Der erforderliche Mindestumfang nach **3.7.3 a** (9) ist  $n \approx 94,5$  also aufgerundet  $n = 95$ .  $\square$

### 3.7.3 c Likelihood-Quotienten-Test

Die Ausführungen über das Fehlerrisiko 2. Art, die Testschärfe und den Mindestumfang beim *Chiadrat-Test* gelten für *kleine Abweichungen*  $\Delta$  approximativ auch für den *Likelihood-Quotienten-Test*.

<b>Test-Box 6a: Anpassungstest für eine einfache Nullhypothese</b>		
<i>K diskidente Ereignisse (Klassen):</i>	$A_1, \dots, A_K$	
<i>Klassen-Wahrscheinlichkeiten:</i>	$p_1 = P(A_1), \dots, p_K = P(A_K)$	
<i>Nullhypothese</i> $H_0: p_k = p_{0k}$ mit <i>festem „Sollwert“</i> $p_{0k}$ für alle $k = 1, \dots, K$		
<i>Stichprobenumfang:</i>	$n$	
<i>beobachtete Anzahlen:</i>	$b_1, \dots, b_K$	
<i>beobachtete relative Häufigkeiten</i>	$\hat{p}_1 = \frac{1}{n} b_1, \dots, \hat{p}_K = \frac{1}{n} b_K$	
<i>unter <math>H_0</math> erwartete Anzahlen:</i>	$e_1 = n p_{01}, \dots, e_K = n p_{0K}$	
<i>Chiadrat-Statistik:</i>	$X^2 = \sum_{k=1}^K \frac{(b_k - e_k)^2}{e_k} = n \cdot \sum_{k=1}^K \frac{(\hat{p}_k - p_{0k})^2}{p_{0k}}$	
<i>Likelihood-Quotienten-Statistik:</i>	$G^2 = 2 \sum_{k=1}^K b_k \cdot \ln \left( \frac{b_k}{e_k} \right) = 2n \cdot \sum_{k=1}^K \hat{p}_k \cdot \ln \left( \frac{\hat{p}_k}{p_{0k}} \right)$	
<i>Freiheitsgrad:</i>	$m = K - 1$	
<b>Anpassungstest:</b>	<i>Fehlerrisiko 1. Art</i> $\approx \alpha$	
<i>Anwendbar, wenn die erwarteten Anzahlen <math>e_k</math> nicht zu klein sind, z.B.</i>		
<i>alle <math>e_k \geq 5</math> (konservative Regel)</i> oder $\frac{n}{K} \geq 2$ bei $\alpha \geq 5\%$ (liberale)		
<i>oder</i> $\frac{n}{K} \geq 4$ bei $\alpha \geq 1\%$ Regeln)		
<i>Entscheidung: Ablehnung von <math>H_0</math>, falls gilt</i>		
<i>Chiadrat-Test</i>	<i>Likelihood-Quotienten-Test</i>	
<i>Testwert</i> $X^2 \geq \chi_{m;\alpha}^2$ bzw.	<i>Testwert</i> $G^2 \geq \chi_{m;\alpha}^2$ bzw.	
<i>P-Wert</i> $1 - \Phi_m(X^2) \leq \alpha$	<i>P-Wert</i> $1 - \Phi_m(G^2) \leq \alpha$	

**Erläuterung zur Test-Box 6 (a und b):** Das  $\alpha$ -Quantil  $\chi_{m;\alpha}^2$  ist im Anhang T tabelliert und kann für größeren Freiheitsgrad  $m$  nach 3.7.4 (2) approximiert werden. Eine exakte Berechnung des Quantils  $\chi_{m;\alpha}^2$  sowie der Verteilungsfunktion  $\Phi_m$  der  $\chi_m^2$ -Verteilung ist auf dem Computer z.B. in Excel möglich (siehe Excel-Tipps am Ende des Kapitels).

<b>Test-Box 6b: Anpassungstest für eine Nullhypothese mit <math>S</math> unbekannten Parametern <math>\theta_1, \dots, \theta_S</math></b>		
$K$ diskidente Ereignisse (Klassen):	$A_1, \dots, A_K$	
Klassen-Wahrscheinlichkeiten:	$p_1 = P(A_1), \dots, p_K = P(A_K)$	
Nullhypothese $H_0: p_k = p_{0k}$ mit $p_{0k} = g_k(\theta_1, \theta_2, \dots, \theta_S)$		für alle $k = 1, \dots, K$
Stichprobenumfang:	$n$	
beobachtete Anzahlen:	$b_1, \dots, b_K$	
beobachtete relative Häufigkeiten	$\hat{p}_1 = \frac{1}{n} b_1, \dots, \hat{p}_K = \frac{1}{n} b_K$	
geschätzte Parameter <sup>1</sup> :	$\hat{\theta}_1, \dots, \hat{\theta}_S$	
unter $H_0$ geschätzte Wahrscheinlichkeiten:	$\hat{p}_{0k} = g_k(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S)$	für alle $k$
unter $H_0$ erwartete Anzahlen:	$e_1 = n \hat{p}_{01}, \dots, e_K = n \hat{p}_{0K}$	für alle $k$
Chiadrat-Statistik	$X^2 = \sum_{k=1}^K \frac{(b_k - e_k)^2}{e_k} = n \cdot \sum_{k=1}^K \frac{(\hat{p}_k - \hat{p}_{0k})^2}{\hat{p}_{0k}}$	
Likelihood-Quotienten-Statistik:	$G^2 = 2 \sum_{k=1}^K b_k \cdot \ln \left( \frac{b_k}{e_k} \right) = 2n \cdot \sum_{k=1}^K \hat{p}_k \cdot \ln \left( \frac{\hat{p}_k}{\hat{p}_{0k}} \right)$	
Freiheitsgrad:	$m = k - 1 - S$	
<b>Anpassungstest:</b>	Fehlerrisiko 1. Art $\approx \alpha$	
Anwendbar, wenn die erwarteten Anzahlen $e_k$ nicht zu klein sind, z.B.		
alle $e_k \geq 5$ (konservative Regel) oder $\frac{n}{K} \geq 2$ bei $\alpha \geq 5\%$ (liberale)		
oder $\frac{n}{K} \geq 4$ bei $\alpha \geq 1\%$ Regeln)		
Entscheidung: Ablehnung von $H_0$ , falls gilt		
<i>Chiadrat-Test</i>	<i>Likelihood-Quotienten-Test</i>	
Testwert $X^2 \geq \chi^2_{m;\alpha}$ bzw.	Testwert $G^2 \geq \chi^2_{m;\alpha}$ bzw.	
P-Wert $1 - \Phi_m(X^2) \leq \alpha$	P-Wert $1 - \Phi_m(G^2) \leq \alpha$	

<sup>1</sup>Die Berechnung der Parameter-Schätzungen hängt von den konkreten Funktionen  $g_k(\theta_1, \theta_2, \dots, \theta_S)$  ab und wird hier nicht allgemein erläutert.

**Power- und Planungs-Box 6a: Für Chiadrat-Anpassungstest mit einfacher Nullhypothese**

$K$  diskidente Ereignisse (Klassen):  $A_1, \dots, A_K$   
 Klassen-Wahrscheinlichkeiten:  $p_1 = P(A_1), \dots, p_K = P(A_K)$

Nullhypothese  $H_0: p_k = p_{0k}$  mit festem „Sollwert“  $p_{0k}$  für alle  $k = 1, \dots, K$

Abweichung von  $H_0$ :

$$\Delta = \sum_{k=1}^K \frac{(p_k - p_{0k})^2}{p_{0k}}$$

Freiheitsgrad:  $m = K - 1$

Fehlerrisiko 1. Art (Testniveau):  $\alpha$

Chiadrat-Quantil:  $\chi^2_{m;\alpha}$

**Berechnung des Fehlerrisikos 2. Art und der Testschärfe**

Stichprobenumfang:  $n$

Nichtzentralität:  $\delta = n \cdot \Delta$

Verteilungsfunktion:  $\Phi_{m,\delta}$  (der nichtzentralen Chiadrat-Verteilung)

Fehlerrisiko 2. Art:  $\beta(\delta) \approx \Phi_{m,\delta}(\chi^2_{m;\alpha})$  (oder grob aus Abb. 2)

Testschärfe:  $\gamma(\delta) \approx 1 - \Phi_{m,\delta}(\chi^2_{m;\alpha})$

**Mindestumfang zu vorgegebenem Fehlerrisiko 2. Art  $\beta$**

Nichtzentralität:  $\delta(\beta)$  bestimmen aus  $\beta \approx \Phi_{m,\delta(\beta)}(\chi^2_{m;\alpha})$   
 oder grob aus Abb. 2

erforderlicher Mindestumfang:  $n \approx \frac{\delta(\beta)}{\Delta}$

**Power- und Planungs-Box 6b:** Für Chiadrat-Anpassungstest bei einer Nullhypothese mit S unbekannten Parametern  $\theta_1, \dots, \theta_S$

K diskidente Ereignisse (Klassen):  $A_1, \dots, A_K$   
 Klassen-Wahrscheinlichkeiten:  $p_1 = P(A_1), \dots, p_K = P(A_K)$

Nullhypothese  $H_0: p_k = p_{0k}$  mit  $p_{0k} = g_k(\theta_1, \theta_2, \dots, \theta_S)$  für alle  $k = 1, \dots, K$

Schätzungen<sup>1</sup> aus „idealer“ Stichprobe (d.h. mit  $\hat{p}_k = p_k$  für alle k):

Parameter:  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S$   
 Wahrscheinlichkeiten:  $\hat{p}_{0k} = g_k(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_S)$  für alle k

Abweichung von  $H_0$ :  $\Delta = \sum_{k=1}^K \frac{(p_k - \hat{p}_{0k})^2}{\hat{p}_{0k}}$

Freiheitsgrad:  $m = K - 1 - S$

Fehlerrisiko 1. Art (Testniveau):  $\alpha$

Chiadrat-Quantil:  $\chi^2_{m;\alpha}$

### Berechnung des Fehlerrisikos 2. Art und der Testschärfe

Stichprobenumfang:  $n$

Nichtzentralität:  $\delta = n \cdot \Delta$

Verteilungsfunktion:  $\Phi_{m,\delta}$  (der nichtzentralen Chiadrat-Verteilung)

Fehlerrisiko 2. Art:  $\beta(\delta) \approx \Phi_{m,\delta}(\chi^2_{m;\alpha})$  (oder grob aus Abb. 2)

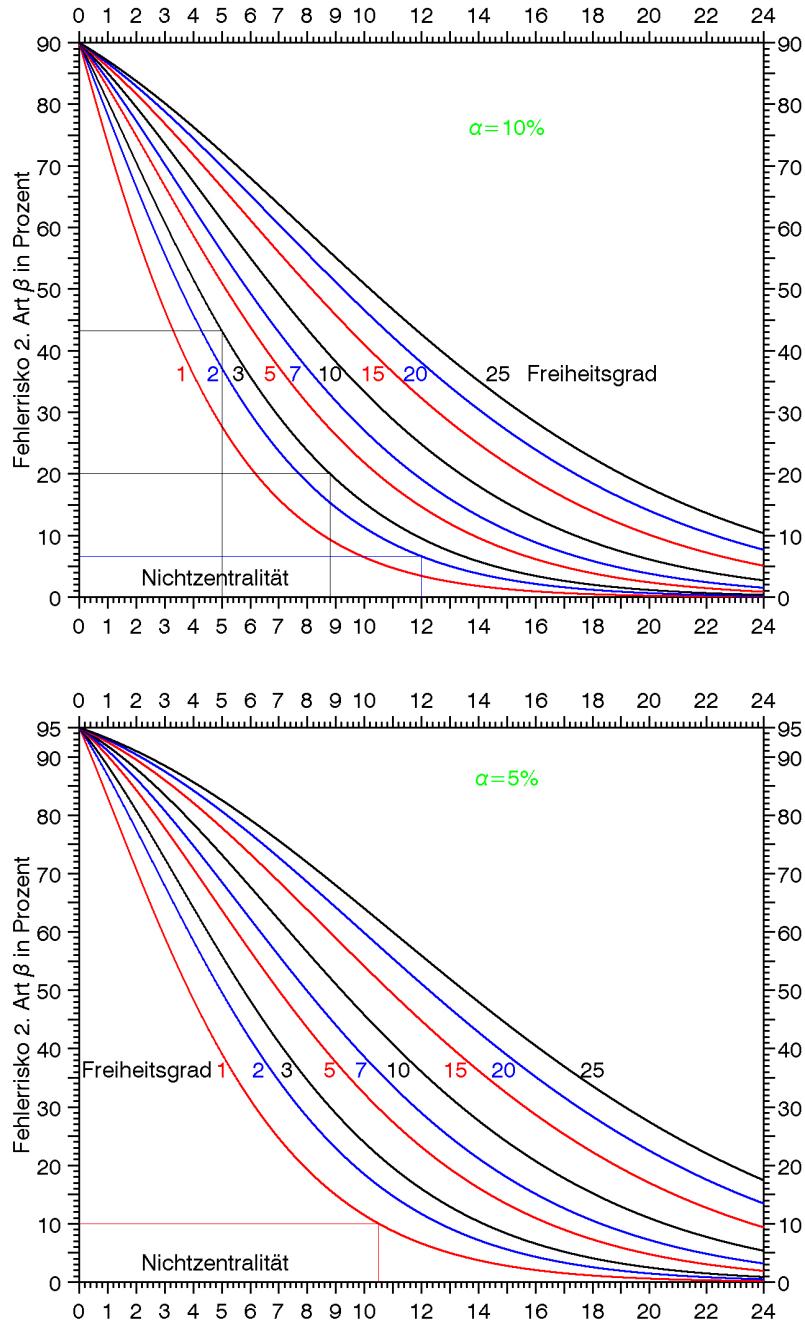
Testschärfe:  $\gamma(\delta) \approx 1 - \Phi_{m,\delta}(\chi^2_{m;\alpha})$

### Mindestumfang zu vorgegebenem Fehlerrisiko 2. Art $\beta$

Nichtzentralität:  $\delta(\beta)$  bestimmen aus  $\beta \approx \Phi_{m,\delta(\beta)}(\chi^2_{m;\alpha})$   
 oder grob aus Abb. 2

erforderlicher Mindestumfang:  $n \approx \frac{\delta(\beta)}{\Delta}$

<sup>1</sup>Die Berechnung der Parameter-Schätzungen hängt von den konkreten Funktionen  $g_k(\theta_1, \theta_2, \dots, \theta_S)$  ab und wird hier nicht allgemein erläutert.



**Abb. 2:** Approximatives Fehlerrisiko 2. Art  $\beta(\delta) \approx \Phi_{m,\delta}(\chi^2_{m;\alpha})$  als Funktion der Nichtzentralität  $\delta$  für  $\alpha = 10\%$  und  $\alpha = 5\%$  und verschiedene Freiheitsgrade. Die eingezeichneten Ablesebeispiele (oben) werden in den Beispielen im Text erläutert. Für einen nicht eingezeichneten Freiheitsgrad  $m$  liest man für die beiden zu  $m$  benachbarten Freiheitsgrade ab und „interpoliert“ zwischen den abgelesenen Werten entsprechend.

### 3.7.4 Die Chiadrat-Verteilung

#### Die zentrale Chiadrat-Verteilung

Die (zentrale)  $\chi_m^2$ -Verteilung mit *Freiheitsgrad*  $m$  nimmt nur *positive* Werte an, d.h. ihr Träger ist das Intervall  $(0, \infty)$  aller *positiven* reellen Zahlen. Ihre Dichtefunktion ist für  $m = 1$  und  $m = 2$  eine fallende Kurve, und für  $m > 2$  eine schiefe Glockenkurve mit einem Maximum für  $x = m - 2$  (vgl. Abb. 3).

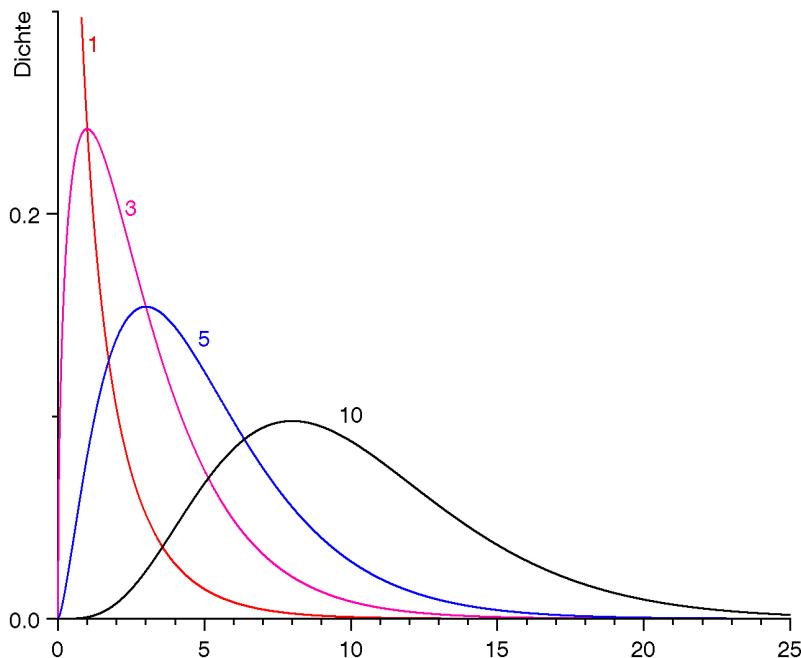
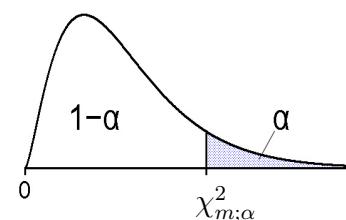


Abb. 3: Dichten der  $\chi_m^2$ -Verteilung für die Freiheitsgrade  $m = 1, 3, 5$  und  $10$ .

Die Verteilungsfunktion  $\Phi_m$  der  $\chi_m^2$ -Verteilung lässt sich mit geeigneten Computer-Programmen (z.B. der Tabellenkalkulationen *Excel*) berechnen.

Das  $\alpha$ -Quantil  $\chi_{m;\alpha}^2$  der  $\chi_m^2$ -Verteilung ist gegeben durch  
(vgl. Abb. rechts)

$$\Phi_m(\chi_{m;\alpha}^2) = 1 - \alpha. \quad (1)$$



Das  $\alpha$ -Quantil lässt sich für großen Freiheitsgrad  $m$  auch unter Verwendung des  $\alpha$ -Quantils  $z_\alpha$  von  $N(0, 1)$  nach *Wilson-Hilferty* wie folgt approximieren

$$\chi_{m,\alpha}^2 \approx m [z_\alpha \sqrt{c} + 1 - c]^3 \quad \text{mit} \quad c = \frac{2}{9m}. \quad (2)$$

Diese Approximation verbessert sich mit wachsendem Freiheitsgrad  $m$ . Sie sollte erst dort benutzt werden, wo die Tabellen im Anhang **T** aufhören, also frühestens

$$\begin{array}{lll} \text{für } m \geq 30 & \text{im Bereich} & 1\% \leq \alpha \leq 99\%, \\ \text{für } m \geq 50 & \text{im Bereich} & 0,1\% \leq \alpha \leq 99,9\%. \end{array}$$

**Für Fortgeschrittene:** Die Dichte  $\varphi_m$  der  $\chi^2_m$ -Verteilung ist gegeben durch

$$\varphi_m(x) = a_m \cdot x^{\frac{m}{2}-1} \cdot e^{-\frac{x}{2}} \quad \text{für } x > 0,$$

wobei die Konstante  $a_m$  durch die *Eulersche Gammafunktion*  $\Gamma$  definiert ist

$$a_m = \left[ 2^{\frac{m}{2}} \cdot \Gamma\left(\frac{m}{2}\right) \right]^{-1}. \quad \square$$

### Die *nichtzentrale* Chiadrat-Verteilung

Die *nichtzentrale*  $\chi^2_m(\delta)$ -Verteilung mit *Freiheitsgrad*  $m$  und *Nichtzentralität*  $\delta \geq 0$  nimmt ebenfalls nur *positive* Werte an. Ihre Dichtefunktion ist für  $m = 5$  und verschiedene Werte der Nichtzentralität  $\delta$  in Abb. 4 dargestellt. Man erkennt daran, daß sich die Dichte für wachsende Nichtzentralität nach rechts verschiebt.

Auf weitere Einzelheiten zur  $\chi^2_m(\delta)$ -Verteilung gehen wir nicht ein.

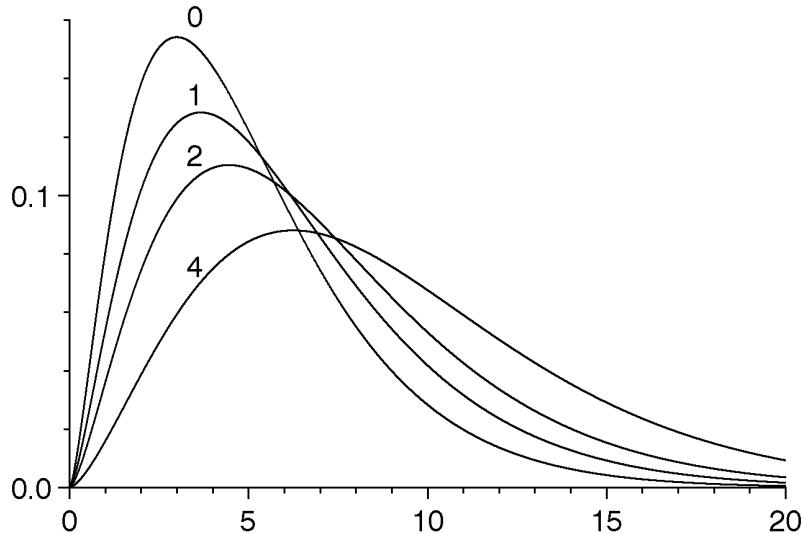


Abb. 4: Dichten der  $\chi^2_m(\delta)$ -Verteilung für  $m = 5$  und Nichtzentralitäten  $\delta = 0, 1, 2$  und 4.

### Tipps für *Microsoft Excel*

**Funktionen:** Das  $\alpha$ -Quantil und die Verteilungsfunktion der *zentralen* Chiquadrat-Verteilung  $\chi_m^2$  mit  $m$  Freiheitsgraden lässt sich mit diesen *Statistik-Funktionen* berechnen:

$$\begin{aligned}\chi_{m;\alpha}^2 &= \text{CHIINV}(\alpha; m) \\ 1 - \Phi_m(x) &= \text{CHIVERT}(x; m) \quad \text{bzw.} \\ \Phi_m(x) &= 1 - \text{CHIVERT}(x; m).\end{aligned}$$

Die für das Fehlerrisiko 2. Art erforderliche Verteilungsfunktion  $\Phi_{m,\delta}$  der *nichtzentralen* Chiquadrat-Verteilung ist in *Excel nicht* implementiert und steht erst nach Installation einer entsprechenden *Erweiterung (Add-In)* zur Verfügung. Ein für diese Zwecke entwickeltes *Add-In (GO Statistik Funktionen)* findet man auf der *Veranstaltungs-Homepage*

[http://www.math.uni-bremen.de/~osius/download/lehre/Statistik\\_NW/](http://www.math.uni-bremen.de/~osius/download/lehre/Statistik_NW/)

und kann es dort (zusammen mit einer Anleitung) herunterladen.

### 3.8 Zweidimensionale Kontingenztafeln

Wir wollen jetzt spezielle *Anpassungstests* betrachten, bei denen die Nullhypothese unbekannte Parameter enthält. Hierbei geht es um die inhaltliche Frage, ob es einen *Zusammenhang zwischen* zwei qualitativen Merkmalen  $X$  und  $Y$  gibt oder nicht. Das Merkmal  $X$  habe  $I \geq 2$  verschiedene Ausprägungen – die wir der Einfachheit halber hier durch Zahlen  $i = 1, \dots, I$  codieren – und die  $J \geq 2$  verschiedenen Ausprägungen von  $Y$  werden entsprechend durch  $j = 1, \dots, J$  codiert. Insgesamt ergeben sich  $K = I \cdot J$  mögliche Ausprägungen für das Paar  $(X, Y)$ .

Für eine Stichprobe vom Umfang  $n$  lassen sich die beobachteten Anzahlen aller Kombinationen beider Merkmale in einer sogenannten  $I \times J$ -Kontingenztafel (oder Kreuztabelle, vgl. auch **1.4.1**) zusammenfassen:

Merkmal $X$	Merkmal $Y$					Zeilensumme
	1	...	$j$	...	$J$	
1	$b_{11}$	...	$b_{1j}$	...	$b_{1J}$	$b_{1+}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$b_{i1}$	...	$b_{ij}$	...	$b_{iJ}$	$b_{i+}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$b_{I1}$	...	$b_{Ij}$	...	$b_{IJ}$	$b_{I+}$
Spaltensumme	$b_{+1}$	...	$b_{+j}$	...	$b_{+J}$	$n = b_{++}$

Tabelle 1: Kontingenztafel der beobachteten Anzahlen

Bei den folgenden Analysen unterscheiden wir nach der Art der Datenerhebung:

- Ziehung aus der *Grundgesamtheit*: es wird *eine* Stichprobe vom Umfang  $n$  aus der *Grundgesamtheit* gezogen, d.h. es liegen  $n$  unabhängige Wiederholungen des Zufallspaars  $(X, Y)$  vor.
- Ziehung *bedingt auf*  $X$ : für jede Stufe  $i = 1, \dots, I$  des Merkmals  $X$  werden  $n_i$  unabhängige Wiederholungen des Merkmals  $Y$  *bedingt auf*  $X = i$  (d.h. aus der *Teilgesamtheit* mit  $X = i$ ) gezogen. Hier sind also die *Zeilensummen* der Kontingenztafel vorher fixiert durch  $b_{i+} = n_i$ . Das Pluszeichen „+“ als *Index* deutet hier immer an, daß über diesen Index *summiert* wurde.
- Ziehung *bedingt auf*  $Y$ : analog der Ziehung *bedingt auf*  $X$  (man vertausche dort  $X$  mit  $Y$ ).

Obwohl die erforderlichen Berechnungen für die statistische Analyse bei drei Daten erhebungen weitgehend übereinstimmen, wollen wir hier die Fälle getrennt untersuchen. Die Ziehung bedingt auf  $Y$  wird allerdings nicht extra behandelt, weil sie nach *Vertauschung* beider Variablen  $X$  und  $Y$  der Ziehung bedingt auf  $X$  entspricht.

### 3.8.1 Vergleich der Verteilung eines Merkmals in verschiedenen Gruppen

**AB0-Blutgruppe und Geschlecht:** Anhand der *StatLab-Auswahl 1985* soll überprüft werden, ob es geschlechtspezifische Unterschiede in der Verteilung der AB0-Blutgruppen gibt (Hypothese) oder nicht (Nullhypothese). Bezeichnen

$$p_{0|w}, \ p_{A|w}, \ p_{B|w}, \ p_{AB|w} \quad \text{bzw.} \quad p_{0|m}, \ p_{A|m}, \ p_{B|m}, \ p_{AB|m}$$

die Frequenzen (Wahrscheinlichkeiten) der vier Blutgruppen beim weiblichen ( $w$ ) bzw. männlichen ( $m$ ) Geschlecht, so lautet die *Nullhypothese*

$$\begin{aligned} H_0: \quad p_{0|w} &= p_{0|m} = p_0, & p_{A|w} &= p_{A|m} = p_A, \\ p_{B|w} &= p_{B|m} = p_B, & p_{AB|w} &= p_{AB|m} = p_{AB}. \end{aligned}$$

Dabei sind  $p_0, \dots, p_{AB}$  die *geschlechtsunspezifischen* Frequenzen der Blutgruppen und stellen hier die *unbekannten Parameter* dar. Die *alternative Hypothese* ist wieder das Gegenteil der Nullhypothese. Bei den Neugeborenen der *StatLab-Auswahl 1985* ergibt sich die Verteilung der Blutgruppen aus *Tabelle 2*, die im folgenden analysiert wird.

Geschlecht	0 (1)	A (2)	B (3)	AB (4)	Summe
weiblich (1)	23	20	4	3	50
männlich (2)	18	18	11	3	50
insgesamt	41	38	15	6	100

*Tabelle 2: AB0-Blutgruppen nach Geschlecht für die StatLab-Auswahl 1985.*

□

**Ziel und Modell:** Allgemeiner soll nun überprüft werden, ob eine *diskrete* Zufallsvariable (Merkmal)  $Y$  in den *verschiedenen*  $X$ -Gruppen *dieselbe* Verteilung hat oder nicht. Die Verteilung von  $Y$  in der Gruppe  $X = i$  ist hierbei gegeben durch die (bedingten) Wahrscheinlichkeiten (vgl. *Tabelle 3*)

$$p_{j|i} = P\{Y = j \mid X = i\} \quad \text{für } j = 1, \dots, J. \quad (1)$$

**Datenerhebung (bedingt auf  $X$ ):** Aus jeder  $X$ -Gruppe  $i$  wird eine Stichprobe mit  $n_i$  unabhängigen Wiederholungen des Merkmals  $Y$  gezogen. Bezeichnet  $b_{ij}$  die *Anzahl*, mit der das Ereignis  $\{Y = j\}$  in der  $X$ -Gruppe  $i$  eingetreten ist, so lässt sich die gesamte Stichprobe vom Umfang  $n = n_1 + \dots + n_I$  durch *Tabelle 1* zusammenfassen, wobei die Zeilensummen *vorgegeben* sind durch  $b_{i+} = n_i$ .

Gruppe $X$	Merkmal $Y$					Zeilensumme
	1	...	$j$	...	$J$	
1	$p_{1 1}$	...	$p_{j 1}$	...	$p_{J 1}$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$p_{1 i}$	...	$p_{j i}$	...	$p_{J i}$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$p_{1 I}$	...	$p_{j I}$	...	$p_{J I}$	1

Tabelle 3: Wahrscheinlichkeiten für  $Y$  in den  $X$ -Gruppen

**Hypothesen:** Die zu prüfende Nullhypothese lautet

$H_0$ : Die Verteilung von $Y$ ist in allen $X$ -Gruppen gleich	bzw.	(2)
$H_0$ : $p_{j 1} = p_{j 2} = \dots = p_{j I}$	für alle $j$	bzw.
$H_0$ : $p_{j i} = p_j$	für alle $i, j$	

wobei  $p_j = P\{Y = j\}$  die – unter  $H_0$  nicht mehr von der Gruppe  $i$  abhängende – Wahrscheinlichkeit für das Ereignis  $Y = j$  ist. Die Wahrscheinlichkeiten  $p_1, \dots, p_J$  sind hier *unbekannte Parameter* mit  $p_1 + \dots + p_J = 1$ , von denen aber nur die ersten  $S = J - 1$  voneinander unabhängig variieren können, weil  $p_J = 1 - (p_1 + \dots + p_{J-1})$ .

Die *alternative Hypothese* ist die Negation der Nullhypothese

$H$ : Die Verteilung von $Y$ ist nicht in allen $X$ -Gruppen gleich	bzw.	(3)
$H$ : $p_{j i} \neq p_j$	für mindestens eine Kombination $i, j$ .	

**Teststatistik und Tests:** Die Beurteilung der Nullhypothese beruht (wie in 3.7.3) auf einem Vergleich von beobachteten mit den unter der Nullhypothese erwarteten Häufigkeiten. Hierzu werden die unbekannten Parameter  $p_j = P\{Y = j\}$  der Nullhypothese zunächst geschätzt durch die relative Häufigkeit des Ereignisses  $\{Y = j\}$  in der *gesamten* Stichprobe, d.h. durch

$$\hat{p}_j = \frac{1}{n} b_{+j} \quad \text{relative Häufigkeit von } Y = j \text{ in der Gesamtstichprobe.} \quad (4)$$

Hieraus ergeben sich die unter  $H_0$  erwarteten Häufigkeiten

$$e_{ij} = n_i \hat{p}_j = \frac{1}{n} b_{i+} b_{+j} \quad (\text{unter } H_0 \text{ erwartete Anzahl}). \quad (5)$$

Die *beobachteten* Häufigkeiten werden mit diesen *erwarteten* Häufigkeiten wieder verglichen mit dem Chiadrat- bzw. Likelihood-Quotienten-Testwert

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(b_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^I \sum_{j=1}^J \frac{n_i(\hat{p}_{j|i} - \hat{p}_j)^2}{\hat{p}_j} \quad (6)$$

$$G^2 = \sum_{i=1}^I \sum_{j=1}^J 2b_{ij} \cdot \ln \left( \frac{b_{ij}}{e_{ij}} \right) = \sum_{i=1}^I \sum_{j=1}^J 2n_i \hat{p}_{j|i} \cdot \ln \left( \frac{\hat{p}_{j|i}}{\hat{p}_j} \right) \quad \text{mit} \quad (7)$$

wobei  $\hat{p}_{j|i}$  (in der jeweils zweiten Darstellung) die relative Häufigkeit für  $Y = j$  in der  $X$ -Gruppe  $i$  darstellt, d.h.

$$\hat{p}_{j|i} = \frac{1}{n_i} b_{ij} \quad \text{relative Häufigkeit von } Y = j \text{ in der } X\text{-Gruppe } i. \quad (8)$$

Unter der Nullhypothese sind beide Teststatistiken  $X^2$  und  $G^2$  bei *großem* Stichprobenumfang  $n$  approximativ  $\chi_m^2$ -verteilt, wobei die Approximation für  $n \rightarrow \infty$  beliebig genau wird. Der Freiheitsgrad  $m$  ergibt sich wie folgt mit der Regel 3.7.2 (6). Zunächst können pro  $X$ -Gruppe  $i$  von den  $J$  beobachteten Anzahlen  $b_{i1}, \dots, b_{iJ}$  (vgl. Zeile  $i$  in Tabelle 1) nur die ersten  $J-1$  voneinander *unabhängig* variieren, weil die Zeilensumme  $b_{i1} + \dots + b_{iJ} = n_i$  vorgegeben ist. Insgesamt liegen daher  $I(J-1)$  voneinander *unabhängige* beobachtete Anzahlen vor. Da  $S = J-1$  voneinander *unabhängige* Parameter geschätzt wurden, erhält man den Freiheitsgrad  $m = I(J-1) - S$  hier als

$$m = (I-1)(J-1) \quad (\text{Freiheitsgrad}). \quad (9)$$

Die Anpassungstests ergeben sich dann wie in 3.7.1 (vgl. Test-Box 7):

*Chiadrat-Test zum Niveau (Fehlerrisiko 1. Art)  $\alpha$ :*

$$\begin{aligned} \text{Ablehnung von } H_0 &\Leftrightarrow \text{Testwert } X^2 \geq \chi_{m;\alpha}^2 \\ &\Leftrightarrow P\text{-Wert } 1 - \Phi_m(X^2) \leq \alpha. \end{aligned} \quad (10)$$

*Likelihood-Quotienten-Test zum Niveau (Fehlerrisiko 1. Art)  $\alpha$ :*

$$\begin{aligned} \text{Ablehnung von } H_0 &\Leftrightarrow \text{Testwert } G^2 \geq \chi_{m;\alpha}^2 \\ &\Leftrightarrow P\text{-Wert } 1 - \Phi_m(G^2) \leq \alpha. \end{aligned} \quad (11)$$

**Test Box 7 Zweidimensionale Kontingenztafel für diskrete Merkmale  $X$  und  $Y$**

Beobachtete Kontingenztafel

Merkmal $X$ $X$	Merkmal $Y$					Zeilensumme
	1	...	$j$	...	$J$	
1	$b_{11}$	...	$b_{1j}$	...	$b_{1J}$	$b_{1+}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$b_{i1}$	...	$b_{ij}$	...	$b_{iJ}$	$b_{i+}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$b_{I1}$	...	$b_{Ij}$	...	$b_{IJ}$	$b_{I+}$
Spaltensumme	$b_{+1}$	...	$b_{+j}$	...	$b_{+J}$	$n = b_{++}$

Nullhypothese  $H_0$ : Die Verteilung von  $Y$  ist in allen  $X$ -Gruppen gleich bzw.  $X$  und  $Y$  sind unabhängig.

unter  $H_0$  erwartete Anzahlen:  $e_{ij} = \frac{1}{n} b_{i+} b_{+j}$

Chiquadrat-Statistik:

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(b_{ij} - e_{ij})^2}{e_{ij}}$$

Likelihood-Quotienten-Statistik:

$$G^2 = \sum_{i=1}^I \sum_{j=1}^J 2 b_{ij} \cdot \ln \left( \frac{b_{ij}}{e_{ij}} \right)$$

Freiheitsgrad:  $m = (I - 1)(J - 1)$

**Anpassungstest:** Fehlerrisiko 1. Art  $\approx \alpha$

Anwendbar, wenn die erwarteten Anzahlen  $e_{ij}$  nicht zu klein sind, z.B.

alle  $e_{ij} \geq 5$  (konervative Regel) oder  $\frac{n}{I \cdot J} \geq 2$  bei  $\alpha \geq 5\%$  (liberale  
oder  $\frac{n}{I \cdot J} \geq 4$  bei  $\alpha \geq 1\%$  Regeln)

Entscheidung: Ablehnung von  $H_0$ , falls gilt

Chiquadrat-Test	Likelihood-Quotienten-Test
Testwert $X^2 \geq \chi^2_{m;\alpha}$ bzw. $P$ -Wert $1 - \Phi_m(X^2) \leq \alpha$	Testwert $G^2 \geq \chi^2_{m;\alpha}$ bzw. $P$ -Wert $1 - \Phi_m(G^2) \leq \alpha$

Erläuterung: Zum Quantil  $\chi^2_{m;\alpha}$  vgl. Anhang T. Die Verteilungsfunktion  $\Phi_m$  der  $\chi^2_m$ -Verteilung kann per Computer berechnet werden (vgl. Excel-Tipps in 3.7.4).

Bei einer *Ablehnung* der Nullhypothese kann man sich einen ersten Eindruck von der *Abhangigkeit* zwischen  $X$  und  $Y$  verschaffen, indem man die beobachtete Verteilung von  $Y$  in den verschiedenen  $X$ -Gruppen  $i$  bestimmt – d.h. die Schatzungen  $\hat{p}_{j|i}$  aus (8) berechnet – und diese miteinander vergleicht (vgl. hierzu auch Abschnitt 1.4.1).

**Fur Fortgeschrittene:** Fur  $I = J = 2$  reduziert sich *Tabelle 1* zu einer  $2 \times 2$ -Tafel, wie wir sie im Kapitel 3.3 behandelt haben. Der obige *Chiquaret-Test* entspricht genau dem *zweiseitigen Test* aus 3.3.2 (9).  $\square$

**AB0-Blutgruppe und Geschlecht des Kindes (1. Fortsetzung):** Damit das Fehlerrisiko 2. Art  $\beta$  nicht zu gro wird, soll hier mit  $\alpha = 10\%$  ein etwas hoheres Fehlerrisiko 1. Art hingenommen werden. Die zur Berechnung von  $X^2$  und  $G^2$  erforderlichen Groen sind in *Tabelle 4* zusammengestellt. Fur den Freiheitsgrad  $m = (2 - 1)(4 - 1) = 3$  ergibt sich das Quantil  $\chi^2_{3;10\%} = 6,251$ , welches sowohl groer als  $X^2 = 3,98$  als auch als  $G^2 = 4,11$  ist (man beachte, da  $X^2$  und  $G^2$  nur wenig voneinander abweichen). Folglich lehnen sowohl der Chiquaret- als auch der Likelihood-Quotienten-Test die Nullhypothese *nicht* ab. Die zugehorigen  $P$ -Werte sind deshalb auch beide groer als das Niveau  $\alpha = 10\%$ :

$$1 - \Phi_2(X^2) = 26,4\% \quad 1 - \Phi_2(G^2) = 25,0\%.$$

Eine unterschiedliche Blutgruppenverteilung fur beide Geschlechter kann mit diesen Daten *nicht* nachgewiesen werden. Da ein Fehler 2. Art vorliegen kann, ist die Gleichheit der Blutgruppenverteilung fur beide Geschlechter nicht mit *quantifizierbarer* Sicherheit bestigt (das Fehlerrisiko 2. Art wird weiter unten genauer analysiert).  $\square$

## Beobachtete relative Häufigkeiten

Geschlecht	0	A	B	AB	Summe
weiblich	46%	40%	8%	6%	100%
	$\hat{p}_{0 w}$	$\hat{p}_{A w}$	$\hat{p}_{B w}$	$\hat{p}_{AB w}$	
männlich	36%	36%	22%	6%	100%
	$\hat{p}_{0 m}$	$\hat{p}_{A m}$	$\hat{p}_{B m}$	$\hat{p}_{AB m}$	
insgesamt	41%	38%	15%	6%	100%
	$\hat{p}_0$	$\hat{p}_A$	$\hat{p}_B$	$\hat{p}_{AB}$	

## Beobachtete und erwartete Anzahlen

Geschlecht	0	A	B	AB	Summe
weiblich	23	20	4	3	50
	20,5	19,0	7,5	3,0	
(erwartet)	18	18	11	3	50
	20,5	19,0	7,5	3,0	
insgesamt	41	38	15	6	100

Tabelle 4a: AB-Blutgruppen nach Geschlecht für die StatLab-Auswahl 1985

Summanden  $(b - e)^2/e$  von  $X^2$ 

Geschlecht	0	A	B	AB	Summe
weiblich	0,30	0,05	1,63	0,00	1,99
männlich	0,30	0,05	1,63	0,00	1,99
insgesamt	0,61	0,11	3,27	0,00	$X^2 = 3,98$

Summanden  $2 b \ln(b/e)$  von  $G^2$ 

Geschlecht	0	A	B	AB	Summe
weiblich	5,29	2,05	-5,03	0,00	2,32
männlich	-4,68	-1,95	8,43	0,00	1,80
insgesamt	0,61	0,11	3,40	0,00	$G^2 = 4,11$

Tabelle 4b: AB0-Blutgruppen nach Geschlecht für die StatLab-Auswahl 1985:  
Berechnung der Summanden für  $X^2$  und  $G^2$ .

**Fehlerrisiko 2. Art und Testschärfe:** Das Fehlerrisiko 2. Art und die Testschärfe des *Chiquaret-Tests* lassen sich (analog 3.7.3 b) mit der Verteilungsfunktion der *nichtzentralen Chiquaret-Verteilung* bestimmen. Die hierfür erforderliche Abweichung  $\Delta$  von der Nullhypothese hängt sowohl von den *wahren* Wahrscheinlichkeiten  $p_{j|i}$  (vgl. Tabelle 3) als auch von den *vorgegebenen* Zeilensummen  $n_i$  ab. Genauer hängt  $\Delta$  nur von Teilstichproben-Anteilen am Gesamtumfang  $n$  ab, d.h. von

$$r_i = \frac{1}{n} n_i \quad (\text{Anteil der } i\text{-ten Teilstichprobe am Gesamtumfang } n). \quad (12)$$

Die „Schätzung“  $\bar{p}_j$  des Parameters  $p_j$  für die „ideale“ Stichprobe (vgl. 3.7.3 b) mit den „idealnen“ Beobachtungen  $b_{ij} = n_i p_{j|i}$  ergibt sich aus (4) zu

$$\bar{p}_j = \frac{1}{n} \sum_i n_i p_{j|i} = \sum_i r_i p_{j|i}. \quad (13)$$

Folglich ist  $\bar{p}_j$  hier das *gewichtete* Mittel der gruppenspezifischen Wahrscheinlichkeiten  $p_{j|1}, \dots, p_{j|I}$ . Die zugehörige Abweichung von der Nullhypothese ist dann

$$\Delta = \sum_{i=1}^I \sum_{j=1}^J \frac{r_i(p_{j|i} - \bar{p}_j)^2}{\bar{p}_j} \quad (\text{Abweichung von der Nullhypothese}). \quad (14)$$

Mit diesem  $\Delta$  ergibt sich die Nichtzentralität  $\delta$  sowie das zugehörige Fehlerrisiko 2. Art  $\beta(d)$  und die Testschärfe  $\gamma(d)$  des *Chiquaret-Tests* wie in der *Power-Box 6b* und die Rechenschritte sind in der *Power- und Planungs-Box 7a* zusammengestellt. Bei nicht zu großem  $\Delta$  ergeben sich das Fehlerrisiko 2. Art sowie die Testschärfe beim *Likelihood-Quotienten-Tests* approximativ wie beim Chiquaret-Test.

**Versuchsplanung:** Im Rahmen einer Versuchsplanung sind die Stichprobenanteile  $n_1, \dots, n_I$  aus konkreten Vorgaben zu bestimmen. Dabei wird man typischerweise ein *balanciertes Design* bevorzugen, d.h. *gleiche* Teilstichprobenumfänge wählen:

$$n_1 = n_2 = \dots = n_I \quad (\text{balanciertes Design}). \quad (15)$$

Im balancierten Fall stimmen alle Anteile  $r_i = \frac{1}{I}$  überein, aber die folgenden Ausführungen gelten auch für *nicht-balancierte* Designs, bei denen man sich die Anteile  $r_i$  beliebig vorgeben kann. Wie in der *Planungs-Box 6b* bestimmt man aus dem vorgegebenen Fehlerrisiko 2. Art  $\beta$  zusammen mit  $\alpha$  und dem Freiheitsgrad  $m$  zunächst die zugehörige Nichtzentralität  $\delta(\beta)$ . Aus der Abweichung  $\Delta$  (die sich aus den Vorgaben nach (14) ergibt) erhält man den erforderlichen Mindestumfang

$$n \approx \frac{\delta(\beta)}{\Delta} \quad (\text{erforderlicher Mindestumfang}). \quad (16)$$

Die resultierenden Teilstichproben-Umfänge  $n_i = r_i \cdot n$  sind *ganzzahlig aufzurunden* (vgl. *Power- und Planungs-Box 7a*).

**Power- und Planungs-Box 7a:****Vergleich der Verteilung des Merkmals  $Y$  in verschiedenen  $Y$ -Gruppen**

Wahrscheinlichkeiten für  $Y$  in den  $X$ -Gruppen

Gruppe $X$	Merkmal $Y$						Zeilensumme
	1	...	$j$	...	$J$		
1	$p_{1 1}$	...	$p_{j 1}$	...	$p_{J 1}$		1
:	:	:	:	:	:		:
$i$	$p_{1 i}$	...	$p_{j i}$	...	$p_{J i}$		1
:	:	:	:	:	:		:
$I$	$p_{1 I}$	...	$p_{j I}$	...	$p_{J I}$		1

Nullhypothese  $H_0$ : Die Verteilung von  $Y$  ist in allen  $X$ -Gruppen gleich.

Stichprobenumfang:  $n$

$X$ -Gruppen:  $n_i$  (Umfang),  $r_i = \frac{1}{n} n_i$  (Stichproben-Anteil)

gewichtete Mittel:  $\bar{p}_j = \sum_i r_i p_{j|i}$

Abweichung von  $H_0$ :  $\Delta = \sum_{i=1}^I \sum_{j=1}^J \frac{r_i(p_{j|i} - \bar{p}_j)^2}{\bar{p}_j}$

Freiheitsgrad:  $m = (I-1)(J-1)$

Fehlerrisiko 1. Art (Testniveau):  $\alpha$

Chiadrat-Quantil:  $\chi^2_{m;\alpha}$

**Berechnung des Fehlerrisikos 2. Art und der Testschärfe**

Nichtszentralität:  $\delta = n \cdot \Delta$

Verteilungsfunktion:  $\Phi_{m,\delta}$  (der nichtzentralen Chiadrat-Verteilung)

Fehlerrisiko 2. Art:  $\beta(\delta) \approx \Phi_{m,\delta}(\chi^2_{m;\alpha})$  (oder grob aus 3.7 Abb. 2)

Testschärfe:  $\gamma(\delta) \approx 1 - \Phi_{m,\delta}(\chi^2_{m;\alpha})$

<b>Mindestumfang zu vorgegebenem Fehlerrisiko 2. Art</b>			
Vorgaben:	$r_i$	(Stichproben-Anteile)	
	typisch:	$r_1 = \dots = r_I = \frac{1}{I}$	(balanciertes Design)
	$p_{j i}$	(Wahrscheinlichkeiten)	oder $\Delta$ (Abweichung)
Nichtzentralität:	$\delta(\beta)$	bestimmen aus	$\beta \approx \Phi_{m,\delta(\beta)}(\chi^2_{m;\alpha})$
		oder grob aus	<b>3.7 Abb. 2</b>
erforderlicher Mindestumfang	$n \approx \frac{\delta(\beta)}{\Delta}$		
		Gruppenumfänge:	$n_i = r_i \cdot n$ (ganzzahlig aufzurunden)

**AB0-Blutgruppe und Geschlecht des Kindes (2. Fortsetzung):** Da beim vorliegenden Testergebnis mit der StatLab-Auswahl Daten ein Fehler 2. Art möglich ist, wollen wir das Fehlerrisiko  $\beta$  exemplarisch berechnen. Hierbei gehen wir davon aus, daß für die beiden Blutgruppen 0 und  $B$  (also die mit den größten Abweichungen zwischen beiden Geschlechtern) die *beobachteten* Blutgruppenhäufigkeiten den *wahren* Frequenzen entsprechen. Für die beiden anderen Blutgruppen  $A$  und  $AB$  dagegen setzen wir etwas stärkere Unterschiede als beobachtet an und gehen dabei konkret von der Blutgruppenverteilung in *Tabelle 4c* aus. Man beachte, daß dort die Blutgruppen-Verteilung für beide Geschlechter zusammen der beobachteten Verteilung (vgl. letzte Zeile in *obere Tabelle 4a*) entspricht. Die Stichprobenanteile sind hier balanciert  $r_1 = r_2 = \frac{1}{2}$  und die zugehörigen Mittelwerte  $\bar{p}_j$  sind in der letzten Zeile von *Tabelle 4c* angegeben. Die Abweichung von der Nullhypothese berechnet man zu  $\Delta = 0,050$  und somit ist die Nichtzentralität  $\delta = n \cdot \Delta = 5$ . Aus **3.7 Abb. 2** liest man für  $m = 3$  das relativ hohe Fehlerrisiko  $\beta(5) \approx 43\%$ . ab (der mit Computer berechnete genauere Wert ist 43,19%).

Will man im Rahmen einer Versuchsplanung mit balanciertem Design für die Verteilung aus *Tabelle 4c* bzw. für die Abweichung  $\Delta = 0,05$  ein Fehlerrisiko 2. Art von nur  $\beta = 20\%$  erreichen, so ergibt sich aus **3.7 Abb. 2** zunächst das zugehörige  $\delta(20\%) \approx 8,8$  (mit Computer erhält man den genaueren Wert 8,798). Der erforderliche Mindestumfang ist dann  $n = 8,8/0,05 = 176$  und die Teilstichprobenumfänge sind  $n_1 = n_2 = \frac{1}{2}n = 88$ , also je 88 Neugeborene pro Geschlecht.  $\square$

Geschlecht	0	A	B	AB	Summe
weiblich	46%	42%	8%	4%	100%
	$p_{0 w}$	$p_{A w}$	$p_{B w}$	$p_{AB w}$	
männlich	36%	34%	22%	8%	100%
	$p_{0 m}$	$p_{A m}$	$p_{B m}$	$p_{AB m}$	
insgesamt (Mittelwert)	41%	38%	15%	6%	100%
	$\bar{p}_0$	$\bar{p}_A$	$\bar{p}_B$	$\bar{p}_{AB}$	

Tabelle 4c: Hypothetische Verteilung der AB0-Blutgruppen nach Geschlecht für die Berechnung des Fehlerrisikos 2. Art und die Versuchsplanung.

### 3.8.2 Test auf Unabhängigkeit zweier Merkmale

**Diagnose bei Appendektomien und Operationstag** (Kinder-Osius-Timm, 1982, Beispiele 5.5 und 7.3.8): Bei den Appendektomien (Entfernung des Blinddarms) eines großen Krankenhauses interessierte unter anderem, ob der Operationstag von der Diagnose *unabhängig* ist oder nicht. Die beobachteten Anzahlen sind in Tabelle 6a angegeben und werden im folgenden analysiert.  $\square$

**Ziel und Modell:** Allgemeiner soll jetzt überprüft werden, ob zwei *diskrete* Zufallsvariablen (Merkmale)  $X$  und  $Y$  voneinander *unabhängig* sind (Nullhypothese) oder nicht (Hypothese). Bezeichnen wir die Wahrscheinlichkeiten für die insgesamt  $K = I \cdot J$  verschiedenen Ausprägungen für das Paar  $(X, Y)$  mit

$$p_{ij} = P\{X = i \text{ und } Y = j\} \quad \text{für } i = 1, \dots, I \text{ und } j = 1, \dots, J, \quad (1)$$

so ist die Verteilung von  $X$  bzw.  $Y$  gegeben durch die Wahrscheinlichkeiten

$$p_{i+} = P\{X = i\} = \sum_j p_{ij} \quad \text{für } i = 1, \dots, I, \quad (2)$$

$$p_{+j} = P\{Y = j\} = \sum_i p_{ij} \quad \text{für } j = 1, \dots, J. \quad (3)$$

Diese Wahrscheinlichkeiten sind in Tabelle 5 schematisch dargestellt. Für jede Ausprägung  $i$  des Merkmals  $X$  ergibt sich die Verteilung von  $Y$  in der durch  $X = i$  definierten *Teilpopulation* aus den *bedingten* Wahrscheinlichkeiten (vgl. auch Tabelle 3)

$$p_{j|i} = P\{Y = j \mid X = i\} = \frac{p_{ij}}{p_{i+}}. \quad (4)$$

Merkmal $X$	Merkmal $Y$					Zeilensumme
	1	...	$j$	...	$J$	
1	$p_{11}$	...	$p_{1j}$	...	$p_{1J}$	$p_{1+}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$p_{i1}$	...	$p_{ij}$	...	$p_{iJ}$	$p_{i+}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$p_{I1}$	...	$p_{Ij}$	...	$p_{IJ}$	$p_{I+}$
Spaltensumme	$p_{+1}$	...	$p_{+j}$	...	$p_{+J}$	1

Tabelle 5: Wahrscheinlichkeiten für das Paar  $(X, Y)$ 

**Beobachtete Daten:** Ausgangspunkt ist eine Stichprobe mit  $n$  unabhängigen Wiederholungen des Paares  $(X, Y)$ . Bezeichnet  $b_{ij}$  die Anzahl in der Stichprobe, mit der das kombinierte Ereignis  $\{X = i \text{ und } Y = j\}$  eingetreten ist, so lässt sich die Stichprobe wieder durch Tabelle 1 zusammenfassen.

**Unabhängigkeitshypothese:** Die zu prüfende Unabhängigkeit beider Merkmale  $X$  und  $Y$  entspricht einem Testproblem mit der Nullhypothese

$$H_0: X \text{ und } Y \text{ sind unabhängig} \quad \text{bzw.} \quad (5)$$

$$\begin{aligned} H_0: P\{X = i \text{ und } Y = j\} &= P\{X = i\} \cdot P\{Y = j\} && \text{für alle } i, j \text{ bzw.} \\ H_0: p_{ij} &= p_{i+} p_{+j} && \text{für alle } i, j \end{aligned}$$

und der Negation als Hypothese

$$H: X \text{ und } Y \text{ sind nicht unabhängig} \quad \text{bzw.} \quad (6)$$

$$H: p_{ij} \neq p_{i+} p_{+j} \quad \text{für mindestens eine Kombination } i, j.$$

Die Nullhypothese lässt sich auch äquivalent dadurch formulieren, daß die Verteilung von  $Y$  in allen  $X$ -Gruppen gleich ist, d.h. die auf  $X = i$  bedingte Verteilung von  $Y$  stimmt für alle  $X$ -Gruppen  $i = 1, \dots, I$  überein (was der Nullhypothese aus 3.8.1 entspricht), d.h.

$$H_0: p_{j|i} = p_j \quad \text{für alle } i, j. \quad (7)$$

**Teststatistik und Tests:** Die Überprüfung der Nullhypothese kann hier als Spezialfall von 3.7.2 durchgeführt werden. Die Nullhypothese spezifiziert insgesamt  $K = I \cdot J$  Wahrscheinlichkeiten  $p_{ij}$  unter Verwendung der unbekannten Wahrscheinlichkeiten  $p_{i+}$  und  $p_{+j}$ , die hier den Parametern  $\theta_1, \dots, \theta_S$  aus 3.7.3 entsprechen. Da die Summe der Wahrscheinlichkeiten  $p_{i+}$  bzw.  $p_{+j}$  jeweils gleich 1 ist, handelt es sich hier um  $S = (I - 1) + (J - 1)$

unabhängige Parameter. Wegen  $K = I \cdot J$  ergibt sich der Freiheitsgrad aus **3.7.2** (5) hier zu

$$m = (I - 1)(J - 1) \quad (\text{Freiheitsgrad}). \quad (8)$$

Die unbekannten Wahrscheinlichkeiten (Parameter)  $p_{i+}$  und  $p_{+j}$  werden durch die entsprechenden relativen Häufigkeiten in der Stichprobe geschätzt

$$\hat{p}_{i+} = \frac{1}{n} b_{i+} \quad \text{und} \quad \hat{p}_{+j} = \frac{1}{n} b_{+j}. \quad (9)$$

Hieraus ergeben sich die unter der Nullhypothese *erwarteten Häufigkeiten* zu

$$e_{ij} = n \hat{p}_{i+} \hat{p}_{+j} = \frac{1}{n} b_{i+} b_{+j} \quad (\text{erwartete Anzahl}). \quad (10)$$

Die *beobachteten* Häufigkeiten werden mit diesen *erwarteten* Häufigkeiten verglichen unter Verwendung des *Chiadrat-* bzw. *Likelihood-Quotienten-Testwerts*

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(b_{ij} - e_{ij})^2}{e_{ij}} = n \cdot \sum_{i=1}^I \sum_{j=1}^J \frac{(\hat{p}_{ij} - \hat{p}_{i+}\hat{p}_{+j})^2}{\hat{p}_{i+}\hat{p}_{+j}} \quad (11)$$

$$G^2 = \sum_{i=1}^I \sum_{j=1}^J 2 b_{ij} \cdot \ln \left( \frac{b_{ij}}{e_{ij}} \right) = n \cdot \sum_{i=1}^I \sum_{j=1}^J 2 n_i \hat{p}_{ij} \cdot \ln \left( \frac{\hat{p}_{ij}}{\hat{p}_{i+}\hat{p}_{+j}} \right) \quad (12)$$

mit

$$\hat{p}_{ij} = \frac{1}{n} b_{ij} \quad \text{relative Häufigkeit für } \{X = i \text{ und } Y = j\}. \quad (13)$$

Unter der Nullhypothese sind beide Teststatistiken  $X^2$  und  $G^2$  bei *großem* Stichprobenumfang  $n$  wieder approximativ  $\chi_m^2$ -verteilt, wobei die Approximation für  $n \rightarrow \infty$  beliebig genau wird. Hieraus ergeben sich die Anpassungstests

*Chiadrat-Test zum Niveau (Fehlerrisiko 1. Art)  $\alpha$ :*

$$\begin{aligned} \text{Ablehnung von } H_0 &\Leftrightarrow \text{Testwert } X^2 \geq \chi_{m;\alpha}^2 \\ &\Leftrightarrow P\text{-Wert } 1 - \Phi_m(X^2) \leq \alpha. \end{aligned} \quad (14)$$

*Likelihood-Quotienten-Test zum Niveau (Fehlerrisiko 1. Art)  $\alpha$ :*

$$\begin{aligned} \text{Ablehnung von } H_0 &\Leftrightarrow \text{Testwert } G^2 \geq \chi_{m;\alpha}^2 \\ &\Leftrightarrow P\text{-Wert } 1 - \Phi_m(G^2) \leq \alpha. \end{aligned} \quad (15)$$

Obwohl sich die Datenerhebung für die beobachtete Tafel hier von der *bedingten* Daten-erhebung in **3.8.1** unterscheidet (dort waren alle Umfänge der  $X$ -Gruppen *vorgegeben*), ergeben sich hier sowohl dieselben erwarteten Häufigkeiten  $e_{ij}$  als auch derselbe Freiheitsgrad  $m$ . Folglich ergeben sich auch dieselben Tests wie in *Test-Box 7*, die deshalb in dieser Situation ebenfalls anwendbar ist. Lediglich das *Fehlerrisiko 2* und die Versuchsplanung (vgl. unten) unterscheiden sich hier gegenüber **3.8.1**.

Bei einer *Ablehnung* der Nullhypothese kann man sich – wie in **3.8.1** – einen ersten Eindruck von der *Abhängigkeit* zwischen  $X$  und  $Y$  verschaffen, wenn man die beobachtete Verteilung von  $Y$  in den verschiedenen  $X$ -Gruppen bestimmt – d.h. die Schätzungen  $\hat{p}_{j|i} = \frac{1}{n_i} b_{ij}$  berechnet – und diese für verschiedene  $i$  miteinander vergleicht.

Diagnose $X$	Operationstag $Y$							Summe
	Mo (1)	Di (2)	Mi (3)	Do (4)	Fr (5)	Sa (6)	So (7)	
perforiert (1)	13 15,83	10 17,94	16 9,66	12 17,43	15 15,40	11 4,94	9 4,79	86
phlegmonös (2)	35 32,96	27 37,34	25 20,11	21 36,28	24 32,05	24 10,28	23 9,98	179
akut (3)	49 53,76	64 60,92	39 32,80	58 59,19	52 52,28	11 16,77	19 16,28	292
chronisch (4)	35 29,83	36 33,80	10 18,20	46 32,84	25 29,01	8 9,30	2 9,03	162
subakut (5)	59 60,58	82 68,63	35 36,96	77 66,69	60 58,91	11 18,90	5 18,34	329
sonstiges (6)	27 25,04	28 28,37	8 15,28	26 27,57	36 24,35	3 7,81	8 7,58	136
Summe	218	247	133	240	212	68	66	1184

Tabelle 6a: beobachtete und (jeweils darunter) erwartete Anzahlen

Diagnose $X$	Operationstag $Y$							Summe
	Mo (1)	Di (2)	Mi (3)	Do (4)	Fr (5)	Sa (6)	So (7)	
perforiert	15%	12%	19%	14%	17%	13%	10%	100%
phlegmonös	20%	15%	14%	12%	13%	13%	13%	100%
akut	17%	22%	13%	20%	18%	4%	7%	100%
chronisch	22%	22%	6%	28%	15%	5%	1%	100%
subakut	18%	25%	11%	23%	18%	3%	2%	100%
sonstiges	20%	21%	6%	19%	26%	2%	6%	100%
insgesamt	18%	21%	11%	20%	18%	6%	6%	100%

Tabelle 6b: beobachtete Verteilung der Operationstage pro Diagnosegruppe

**Diagnose bei Appendektomien und Operationstag (Fortsetzung):** Die beobachteten und erwarteten Anzahlen sind in Tabelle 6a angegeben. Der Test auf Unabhängigkeit soll zum Niveau  $\alpha = 5\%$  erfolgen, und für den Freiheitsgrad  $m = (6 - 1)(7 - 1) = 30$  ist das Quantil  $\chi^2_{30;5\%} = 43,773$ . Hier ergibt  $X^2 = 119,029$  und somit lehnt der Chiquadrat-Test die Nullhypothese ab. Der Likelihood-Quotienten-Test kommt wegen  $G^2 = 116,595$  zum gleichen Resultat. Die zugehörigen  $P$ -Werte sind hier extrem gering:  $P < 0,00005\%$ . Da nur ein Fehler 1. Art möglich ist, ist die Abhängigkeit von Operationstag und Diagnose mit der Irrtumswahrscheinlichkeit von  $\alpha = 5\%$  abgesichert. Ein erster Eindruck von dieser Abhängigkeit ergibt sich durch Betrachtung der beobachteten Verteilung des Operationstages  $Y$  in den verschiedenen Diagnosegruppen (vgl. Tabelle 6b). Auffällig ist hier die geringere Operationshäufigkeit am Wochenende in den letzten vier Diagnosegruppen im Vergleich mit den ersten beiden Diagnosen.

**Fehlerrisiko 2. Art und Testschärfe:** Das Fehlerrisiko 2. Art und die Testschärfe des Chiquadrat-Tests lassen sich wie in 3.7.3 b unter Verwendung der Verteilungsfunktion der nichtzentralen Chiquadrat-Verteilung bestimmen. Ausgangspunkt sind hierbei die wahren Wahrscheinlichkeiten  $p_{ij}$ , zu denen wir die zugehörige Abweichung  $\Delta$  von der Nullhypothese bestimmen müssen. Die „Schätzungen“ der Parameter  $p_{i+}$  bzw.  $p_{+j}$  für die „ideale“ Stichprobe (vgl. 3.7.3 b) mit den „idealen“ Beobachtungen  $b_{ij} = n p_{ij}$  ergeben sich aus (9) zu  $\hat{p}_{i+} = p_{i+}$  bzw.  $\hat{p}_{+j} = p_{+j}$ . Damit lässt sich die Abweichung  $\Delta$  darstellen als

$$\Delta = \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - p_{i+}p_{+j})^2}{p_{i+}p_{+j}} \quad (\text{Abweichung von der Nullhypothese}). \quad (16)$$

Mit diesem  $\Delta$  ergibt sich die Nichtzentralität  $\delta$  sowie das zugehörige Fehlerrisiko 2. Art  $\beta(\delta)$  und die Testschärfe  $\gamma(\delta)$  des *Chiquaret-Tests* wie in der *Power-Box 6b* und die Rechenschritte sind in *Power- und Planungs-Box 7b* zusammengestellt. Bei *nicht zu großem*  $\Delta$  ergeben sich das Fehlerrisiko 2. Art sowie die Testschärfe beim *Likelihood-Quotienten-Test* approximativ wie beim Chiquaret-Test.

**Versuchsplanung:** Bei der Versuchsplanung geht es wieder darum, den erforderlichen Mindestumfang  $n$  zu bestimmen, so daß beim *Chiquaret-Test* zum Niveau  $\alpha$  für *vorgegebene* Wahrscheinlichkeiten  $p_{ij}$  (bzw. für die zugehörige Abweichung  $\Delta$  aus (16)) das Fehlerrisiko 2. Art (höchstens) gleich der Vorgabe  $\beta$  ist. Wie in der *Planungs-Box 6b* bestimmt man aus dem vorgegebenen  $\beta$  zusammen mit  $\alpha$  und dem Freiheitsgrad  $m$  zunächst die zugehörige Nichtzentralität  $\delta(\beta)$ . Aus der Abweichung  $\Delta$  ergibt sich der erforderliche (noch aufzurundende) Mindestumfang  $n$  wie in der *Planungs-Box 6*. Die erforderlichen Schritte sind in der *Power- und Planungs-Box 7b* zusammengestellt.

**Power- und Planungs-Box 7b:****Test auf Unabhängigkeit der Merkmale  $X$  und  $Y$** 

Wahrscheinlichkeiten für das Paar  $(X, Y)$

Merkmal $X$	Merkmal $Y$					Zeilensumme
	1	...	$j$	...	$J$	
1	$p_{11}$	...	$p_{1j}$	...	$p_{1J}$	$p_{1+}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$p_{i1}$	...	$p_{ij}$	...	$p_{iJ}$	$p_{i+}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$p_{I1}$	...	$p_{Ij}$	...	$p_{IJ}$	$p_{I+}$
Spaltensumme	$p_{+1}$	...	$p_{+j}$	...	$p_{+J}$	1

Nullhypothese

$H_0: X$  und  $Y$  sind unabhängig.

Abweichung von  $H_0$ :

$$\Delta = \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - p_{i+}p_{+j})^2}{p_{i+}p_{+j}}$$

Freiheitsgrad:

$$m = (I-1)(J-1)$$

Fehlerrisiko 1. Art:

$$\alpha$$

Chiquadrat-Quantil:

$$\chi^2_{m;\alpha}$$

**Berechnung des Fehlerrisikos 2. Art und der Testschärfe**

Stichprobenumfang:

$$n$$

Nichtzentralität:

$$\delta = n \cdot \Delta$$

Verteilungsfunktion:

$$\Phi_{m,\delta}$$

(der nichtzentralen Chiquadrat-Verteilung)

Fehlerrisiko 2. Art:

$$\beta(d) \approx \Phi_{m,\delta}(\chi^2_{m;\alpha})$$

(oder grob aus **3.7 Abb. 2**)

Testschärfe:

$$\gamma(\delta) \approx 1 - \Phi_{m,\delta}(\chi^2_{m;\alpha})$$

**Mindestumfang zu vorgegebenem Fehlerrisiko 2. Art  $\beta$** 

Nichtzentralität:

$$\delta(\beta)$$

bestimmen aus

$$\beta \approx \Phi_{m,\delta(\beta)}(\chi^2_{m;\alpha})$$

oder grob aus

**3.7 Abb. 2**

erforderlicher Mindestumfang:

$$n \approx \frac{\delta(\beta)}{\Delta}$$

## A Literatur-Auswahl

Von der äußerst umfangreichen Literatur zur Statistik und deren Anwendungen in den Biowissenschaften (Biometrie) sind hier nur einige wenige Lehrbücher aufgeführt und kurz dokumentiert. Auf die Aufgabe von Zeitschriftenartikeln wird hier verzichtet, weil sie im laufenden Text (z.B. bei Anwendungs-Beispielen) zitiert werden.

Weitere ausführliche Literaturhinweise finden man in den angegebenen Büchern.

### **Lehrbücher**

**Batschelet, E.:** *Mathematik für Biologen*. Springer, Berlin/Heidelberg/New York 1980.

Grundlegendes Werk zur *Mathematik* inclusive *Wahrscheinlichkeitsrechnung*, aber *ohne Statistik*.

**Hodges, J.L.; Krech, D.; Crutchfield, R.S.:** *StatLab - An Empirical Introduction to Statistics*. McCraw-Hill, New York 1975.

Alternativer empirischer Einstieg in die Statistik.

**Kinder, Hans-Peter; Osius, Gerhard; Timm, Jürgen:** *Statistik für Biologen und Mediziner*. Vieweg & Sohn, Braunschweig/Wiesbaden 1982. (vergriffen!) Empfehlenswert als Begleitlektüre und Ergänzung.

**Kesel, Antonia; Junge, Monika M.; Nachtigal, Werner:** *Einführung in die angewandte Statistik für Biowissenschaftler*. Birkhäuser Verlag, Basel 1999.

Empfehlenswert als Begleitlektüre und Ergänzung, mit vielen Beispielen.

**Köhler, Wolfgang; Schachtel, Gabriel; Voleske, Peter:** *Biometrie - Einführung in die Statistik für Biologen und Agrarwissenschaftler*. Springer, Berlin/Heidelberg 1996.

Empfehlenswert als Begleitlektüre und Ergänzung, mit vielen Beispielen.

**Lorenz, Rolf. J.:** *Grundbegriffe der Biometrie*. 3. Auflage,

G. Fischer Verlag, Stuttgart 1992.

Empfehlenswert als Begleitlektüre und Ergänzung, mit vielen Beispielen.

**Sachs, Lothar:** *Angewandte Statistik: Anwendung statistischer Methoden* 9. Auflage, Springer, Berlin 1999.

Ein Standard-Nachschlagewerk für Anwender.

Als unterhaltsames und lehrreiches Sachbuch über Statistik wird empfohlen:

**Beck-Bornholdt, H.-P. und Dubben, H.-H.:** *Der Hund der Eier legt*

- *Erkennen von Fehlinformationen durch Querdenken*.

rororo Sachbuch, Rowohl Taschenbuch Verlag 1998.

## B Statistik-Funktionen von *Microsoft Excel*

Die *wichtigsten* Statistik-Funktionen aus den *Tipps für Microsoft Excel* im Skript sind hier tabellarisch zusammengestellt. Für Details konsultiere man die angegebene Seite oder die *Microsoft Excel-Hilfe*.

<b>Funktionen in Excel</b>	<b>ggf. Beschreibung (Kommentar)</b>	<b>Seite</b>
HÄUFIGKEIT	Matrix-Formel: Achtung bei Eingabe	5
MITTELWERT		13
MEDIAN		13
STABW	Standardabweichung	13
VARIANZ		13
STEIGUNG	$\hat{b}$ der <i>ungewichteten</i> Regressionsgeraden	30
ACHSENABSCHNITT	$\hat{a}$ der <i>ungewichteten</i> Regressionsgeraden	30
BESTIMMTHEITSMASS	der <i>ungewichteten</i> Regressionsgeraden	30
KORREL	Korrelationskoeffizient ( <i>ungewichtet</i> )	30
SUMMENPRODUKT	(für <i>gewichtete</i> Regressionsgeraden etc.)	34,67
BINOMVERT( $x; n; p; 0$ )	$P\{X = x\}$ für <i>Binomial(n, p)-verteiltes X</i>	92
BINOMVERT( $x; n; p; 1$ )	$P\{X \leq x\}$ für <i>Binomial(n, p)-verteiltes X</i>	92
FAKULTÄT( $n$ )	$n!$	94
KOMBINATIONEN( $n, k$ )	$\binom{n}{k}$	94
POISSON( $x; \mu; 0$ )	$P\{X = x\}$ für <i>Poisson(<math>\mu</math>)-verteiltes X</i>	122
POISSON( $x; \mu; 1$ )	$P\{X \leq x\}$ für <i>Poisson(<math>\mu</math>)-verteiltes X</i>	122
STANDNORMVERT	Verteilungsfunktion $\Phi$ der Stand.-Normalvertlg.	107
STANDNORMINV	Inverse der Funktion $\Phi$ (für $\alpha$ -Quantil $z_\alpha$ )	139
NORMVERT( $x; \mu; \sigma; 1$ )	$P\{X \leq x\}$ für <i>Normal(<math>\mu, \sigma^2</math>)-verteiltes X</i>	107
TVERT( $x; m; 1$ )	$P\{X \geq x\}$ für <i><math>t_m</math>-verteiltes X (nur für <math>x \geq 0</math>)</i>	179
TINV( $2\alpha; m$ )	$\alpha$ -quantil $t_{m;\alpha}$ der $t_m$ -Verteilung	179
CHIVERT( $x; m$ )	$P\{X \geq x\}$ für $\chi_m^2$ -verteiltes X	241
CHIINV( $\alpha; m$ )	$\alpha$ -Quantil $\chi_{m;\alpha}^2$ der $\chi_m^2$ -Verteilung	241
FVERT( $x; m; n$ )	$P\{X \geq x\}$ für $F_{mn}$ -verteiltes X	-
FINV( $\alpha; m; n$ )	$\alpha$ -Quantil $F_{mn;\alpha}$ der $F_{mn}$ -Verteilung	139

## C StatLab-Daten

Die Daten wurden im Rahmen einer medizinischen Studie von Neugeborenen („Kaiser Foundation Health Plan“) in Oakland (Kalifornien) unter Leitung von J. Yerushalmy in den Jahren 1961-1972) erhoben. Für jedes Kind wurden zahlreiche Daten zu zwei Zeitpunkten registriert, bei der Geburt (1961-63) und 10 Jahre später (Kontrolle). Folgende ausgewählte Daten sind hier aufgeführt.

### Kind (Zeitpunkt: Geburt)

- B:** Blutgruppe im AB0-System,
- Rh:** Resus-Faktor,
- Grö:** Grösse (in Zoll),
- Gw:** Gewicht (in Pounds),
- M:** Monat der Geburt (1=Jan, . . . , 12=Dez.),
- W:** Wochentag der Geburt(1=Sonntag, . . . , 7=Samstag),
- H:** Uhrzeit der Geburt (in Stunden).

### Kind (Zeitpunkt: Kontrolle)

- Grö:** Grösse (in Zoll),
- Gw:** Gewicht (in Pounds),

### Mutter(Zeitpunkt Geburt)

- B:** Blutgruppe im AB0-System,
- Rh:** Rhesus-Faktor,
- Al:** Alter (in vollendeten Lebensjahren),
- Gw:** Gewicht (in Pounds),
- Ra:** Rauchen (in Zigaretten/Tag, A=aufgehört, N=nie geraucht).

### Mutter (Zeitpunkt: Kontrolle)

- Grö:** Grösse (in Zoll),
- Gw:** Gewicht (in Pounds),

### Vater (Zeitpunkt: Geburt)

- A1:** Alter (in vollendeten Lebensjahren),
- Ra:** Rauchen (in Zigaretten(Tag, A=aufgehört, N=nie geraucht)).

### Vater: (Zeitpunkt: Kontrolle)

- Grö:** Grösse (in Zoll),
- Gw:** Gewicht (in Points).

Für Beispiel-Analysen werden jeweils *verschiedene* Teil-Datensätze, bestehend aus je 50 *zufällig* ausgewählten Mädchen und 50 Jungen verwendet, und zwar:

im Skript (Kurs): StatLab-Auswahl 1985,  
in den Übungen: StatLab-Auswahl Ü[Jahr]

**Umrechnung:** 1 Zoll = 2,54 cm; 1 Pound = 0,45359237 kg.

### Quelle:

J.L. Hodges, D. Krech, R.S. Crutchfield (1975). StatLab: an empirical introduction to statistics. McGraw-Hill, New York.

StatLab-Auswahl 1985													50 Mädchen							
Nr.	Kind (weiblich)								Mutter					Vater						
	Geburt				Kontrolle				Geburt				Kontrolle		Geburt	Kontrolle				
	B	Rh	Grö	Gw	M	W	H	Grö	Gw	B	Rh	AI	Gw	Ra	Grö	Gw	AI	Ra	Grö	Gw
1	0	+	19,8	6,4	7	7	5	51,6	78	A	+	32	130	20	61,6	153	33	N	66,0	181
2	0	-	19,5	6,3	10	3	16	56,3	68	B	+	25	98	N	65,6	98	35	20	70,1	166
3	B	+	19,5	7,1	2	1	5	49,8	63	0	+	32	122	20	60,6	163	38	20	68,0	185
4	B	+	20,0	7,3	10	7	5	54,3	68	A	+	30	128	A	66,6	117	33	5	73,5	208
5	A	-	19,0	7,1	10	4	1	50,4	56	A	-	34	138	20	66,0	175	42	30	67,8	141
6	A	+	22,0	10,6	3	2	10	56,3	77	A	+	43	132	N	65,6	141	39	N	72,0	185
7	0	-	21,8	7,9	9	1	2	55,8	70	0	-	19	115	A	63,3	150	20	N	69,0	180
8	A	-	20,0	6,4	5	7	13	48,9	59	A	+	17	130	20	62,8	159	23	11	65,0	130
9	0	+	21,5	9,6	4	5	19	48,7	57	0	+	29	128	N	63,1	148	31	N	69,0	162
10	A	+	20,5	7,3	8	5	3	54,7	80	A	+	25	145	2	67,0	200	26	6	69,0	175
11	0	+	20,5	7,3	6	3	19	57,0	82	A	+	26	172	N	68,7	159	39	20	74,0	180
12	A	+	20,0	7,1	6	2	16	51,6	66	A	+	35	125	N	60,5	135	43	N	62,3	109
13	0	+	20,0	6,3	5	3	14	52,4	66	A	+	20	102	A	64,0	120	21	20	70,5	150
14	A	-	14,8	2,3	11	1	16	50,4	54	A	+	24	155	20	67,9	152	28	20	71,8	176
15	A	+	20,5	7,3	6	7	14	51,1	61	0	+	24	110	N	58,7	112	25	N	68,0	172
16	0	+	19,5	6,4	4	5	7	51,1	59	0	+	29	94	20	62,8	104	31	20	72,0	195
17	B	+	20,5	6,6	3	7	17	55,3	81	0	-	33	133	A	66,0	145	38	A	72,0	210
18	AB	+	20,5	6,6	6	3	19	53,8	74	AB	+	21	135	6	63,5	145	26	5	70,5	177
19	A	+	20,5	7,7	7	6	16	49,1	46	0	+	27	109	A	60,2	120	25	10	68,0	130
20	AB	+	19,5	6,9	4	3	1	51,4	60	AB	+	22	118	N	63,1	159	23	30	68,0	185
21	0	+	20,0	8,8	6	1	10	56,2	73	A	+	32	126	N	66,7	126	35	N	70,0	160
22	AB	-	20,0	7,6	9	5	16	52,9	75	B	+	23	185	15	68,6	247	25	20	68,0	161
23	A	+	20,5	7,1	10	3	10	53,9	68	A	+	19	135	N	65,6	151	22	N	69,0	160
24	A	+	20,5	6,4	4	2	18	52,2	84	A	+	18	102	4	62,3	120	20	20	72,0	150
25	0	+	18,5	5,6	10	5	14	50,3	58	B	+	33	130	N	63,8	141	34	N	68,0	180
26	0	+	20,5	7,3	6	2	2	55,1	73	0	+	22	108	10	62,1	120	25	10	71,5	177
27	0	+	21,0	7,4	7	3	15	55,3	73	0	+	30	140	N	63,6	164	33	20	67,0	185
28	0	+	20,0	7,0	9	6	20	57,2	92	0	+	36	124	N	62,3	146	42	N	65,0	150
29	0	+	21,0	7,1	8	3	13	51,1	64	B	+	27	154	15	66,8	181	37	20	70,0	175
30	A	+	21,0	9,8	10	1	20	52,8	64	A	-	22	145	N	66,0	140	23	20	68,0	170
31	0	+	19,5	6,5	7	6	5	54,2	78	B	+	40	162	N	63,3	160	43	N	64,0	150
32	A	-	21,0	6,8	7	3	23	57,9	114	A	-	27	134	20	64,6	154	34	N	71,0	189
33	0	+	22,5	9,3	12	3	8	53,4	106	B	+	36	167	A	67,1	215	40	24	67,0	135
34	A	+	19,5	5,8	8	2	20	55,9	77	A	+	30	138	A	66,7	159	30	N	65,0	170
35	0	+	20,5	7,6	2	6	2	52,6	68	0	+	36	135	N	64,3	144	41	A	68,5	160
36	A	+	19,5	5,6	1	6	13	51,3	61	B	+	25	110	N	64,4	125	24	20	72,3	147
37	0	+	20,5	7,3	1	4	18	53,3	59	0	+	24	128	10	67,0	122	24	24	68,1	178
38	0	+	19,0	5,3	3	1	9	51,7	56	0	+	20	97	N	59,6	138	21	5	64,9	169
39	A	+	19,5	6,9	5	6	19	51,7	74	A	+	23	185	10	65,1	184	31	A	71,0	190
40	A	+	21,0	6,9	9	1	3	53,8	60	0	+	35	116	50	65,8	107	40	50	70,0	170
41	0	+	19,5	6,1	8	4	6	56,4	74	0	+	23	112	15	64,0	119	27	A	69,0	155
42	A	+	18,0	6,8	6	1	17	54,4	72	A	+	39	145	20	64,7	159	37	N	71,0	160
43	0	+	20,0	7,1	10	7	2	54,1	73	0	+	22	155	N	64,0	168	24	20	70,0	180
44	A	+	20,5	8,7	7	6	23	47,4	62	A	+	34	144	A	65,0	144	39	30	68,5	185
45	B	+	21,5	9,1	3	4	10	57,5	108	B	+	36	119	5	62,5	146	37	10	72,0	175
46	0	+	19,0	5,9	11	4	8	51,9	72	0	+	24	120	20	65,2	139	25	A	74,0	175
47	0	-	20,0	7,1	4	4	9	52,4	81	0	-	22	160	15	65,9	174	26	20	71,0	220
48	A	+	21,0	8,2	8	2	8	51,5	68	0	+	25	120	10	66,5	122	27	20	70,6	202
49	0	-	20,0	7,4	3	6	3	52,9	57	0	+	28	202	N	67,5	175	39	N	72,0	155
50	A	+	21,5	7,5	9	7	13	60,0	100	0	+	28	130	1	65,5	137	29	A	72,6	193
Nr.	B	Rh	Grö	Gw	M	W	H	Grö	Gw	B	Rh	AI	Gw	Ra	Grö	Gw	AI	Ra	Grö	Gw

StatLab-Auswahl 1985												50 Jungen								
Nr	Kind (männlich)								Mutter						Vater					
	Geburt				Kontrolle				Geburt				Kontrolle		Geburt	Kontrolle				
	B	Rh	Grö	Gw	M	W	H	Grö	Gw	B	Rh	AI	Gw	Ra	Grö	Gw	AI	Ra	Grö	Gw
1	B	+	20,0	6,5	7	1	3	53,2	70	B	+	24	120	N	61,1	108	30	30	66,5	170
2	A	+	19,0	6,1	4	7	14	53,8	71	B	+	22	107	3	61,8	103	23	20	75,0	215
3	B	+	19,5	6,4	6	4	3	48,9	52	0	+	28	97	N	58,8	101	30	N	66,0	135
4	A	+	21,0	8,2	11	2	7	53,8	68	AB	+	22	124	10	65,3	136	22	N	71,8	170
5	0	+	19,0	7,4	9	4	24	53,1	72	A	+	25	146	N	61,7	170	32	6	68,7	168
6	A	+	19,8	6,4	12	7	1	52,9	64	A	+	31	150	15	64,4	178	38	20	67,5	190
7	0	-	21,0	10,3	5	6	14	52,6	71	0	-	29	145	N	66,5	170	29	N	68,0	129
8	0	+	21,0	7,9	7	1	18	53,1	68	0	+	25	150	N	62,8	188	26	12	70,5	198
9	0	+	20,8	7,4	8	5	10	52,1	69	A	+	26	133	N	66,6	172	26	10	73,0	170
10	0	+	21,0	8,1	9	4	11	55,0	69	A	+	30	142	1	66,9	141	34	1	71,0	165
11	A	-	21,0	7,7	12	5	10	52,4	66	A	+	17	140	N	64,0	212	21	A	65,5	151
12	A	+	18,0	5,6	5	2	4	51,6	57	A	+	24	114	N	66,0	129	25	N	65,9	145
13	A	+	20,0	7,1	9	4	22	53,1	76	0	+	36	140	A	59,4	135	34	20	67,0	145
14	A	+	22,5	8,2	9	2	18	54,8	67	0	+	21	126	N	66,4	132	24	20	72,0	160
15	0	+	23,0	9,9	6	2	12	59,2	114	0	+	28	132	N	64,8	158	34	10	72,0	182
16	0	+	20,5	7,0	10	1	20	56,4	96	A	+	21	120	N	62,0	125	29	N	69,9	151
17	B	+	20,5	6,9	6	5	20	50,8	59	AB	+	18	108	2	61,9	119	19	A	68,0	155
18	0	+	20,0	7,2	9	7	24	50,4	56	0	+	26	130	N	62,3	153	30	10	62,5	145
19	0	-	21,5	8,4	8	2	21	55,1	74	0	-	19	95	N	60,3	109	24	20	74,0	170
20	B	+	21,0	7,2	8	7	21	53,3	69	0	+	34	128	20	65,8	144	38	A	71,0	175
21	A	+	20,0	6,4	12	4	4	55,2	75	A	+	25	93	7	60,6	104	32	3	75,0	190
22	0	+	20,0	8,2	4	7	2	54,3	65	A	+	24	92	N	62,1	110	28	30	73,0	183
23	0	+	20,0	7,7	8	4	18	50,3	57	A	+	27	125	N	63,9	135	34	N	72,0	200
24	0	+	21,0	7,9	4	3	18	52,0	72	0	+	28	165	5	63,0	180	38	N	65,0	175
25	AB	+	20,0	7,6	3	4	19	55,2	87	B	+	34	125	N	61,3	144	38	20	67,0	175
26	B	-	20,5	7,9	9	2	8	53,1	62	0	+	26	155	17	67,6	178	32	N	71,0	140
27	B	-	19,5	6,5	10	4	16	50,7	66	B	-	28	130	N	63,6	139	29	N	72,0	200
28	B	+	22,0	8,8	10	2	10	53,8	75	0	+	28	210	N	65,0	195	39	20	75,0	195
29	0	+	21,0	9,1	11	6	19	56,8	74	0	-	24	125	A	67,3	121	26	20	73,5	175
30	A	+	20,8	7,6	12	1	3	54,2	64	0	-	22	133	N	66,0	130	28	35	70,1	162
31	AB	+	20,0	7,1	9	7	24	49,3	50	AB	+	29	113	N	64,8	113	36	A	69,0	160
32	B	-	21,0	8,8	10	7	13	52,3	64	0	-	26	127	N	62,0	150	27	20	71,6	205
33	A	+	22,0	8,6	2	4	12	56,1	67	A	+	20	120	N	63,1	134	24	10	76,0	188
34	A	+	21,0	7,6	6	4	8	50,9	64	A	+	21	134	N	64,1	149	25	4	66,0	145
35	B	+	20,0	7,4	12	6	21	51,1	60	B	+	25	112	A	64,0	125	28	10	70,0	165
36	B	+	21,0	7,9	12	1	13	50,4	58	B	+	39	134	N	61,5	127	49	N	68,0	189
37	A	+	20,5	6,5	5	7	11	51,4	60	A	+	26	115	N	62,6	107	28	A	69,5	180
38	A	+	20,5	7,8	11	4	6	55,5	66	A	+	19	135	A	67,1	143	23	N	71,0	165
39	0	+	19,0	5,8	8	2	3	50,2	58	A	-	25	108	20	60,7	136	28	20	68,0	190
40	B	+	20,5	7,6	6	6	14	51,2	54	B	-	28	124	A	64,5	191	30	25	64,0	120
41	0	+	21,0	7,1	4	5	7	53,3	61	0	+	25	140	20	65,5	137	27	N	72,0	185
42	0	+	20,0	7,2	6	2	13	50,6	60	0	+	25	155	20	61,9	166	33	15	69,0	160
43	A	+	20,5	7,9	8	3	22	52,2	64	A	+	23	120	N	63,3	139	26	N	69,0	170
44	A	+	20,5	7,1	11	3	13	53,1	66	A	+	29	122	20	61,7	161	30	20	71,6	164
45	0	+	20,5	7,9	4	4	3	52,9	60	0	+	22	165	20	68,0	204	25	15	68,0	160
46	A	+	22,0	8,7	10	3	18	60,0	116	0	+	30	162	N	69,6	167	31	N	74,6	231
47	A	+	19,5	7,9	5	1	20	54,4	81	0	+	31	127	N	63,6	140	33	20	68,0	170
48	0	-	22,5	9,9	8	4	15	54,0	72	A	-	30	140	N	67,0	154	30	N	67,0	160
49	A	+	20,5	7,9	1	6	4	54,3	77	A	-	25	144	N	69,1	174	25	15	72,0	200
50	AB	-	20,0	7,9	9	2	5	51,6	70	AB	+	32	128	N	65,9	138	60	N	60,8	160
Nr	B	Rh	Grö	Gw	M	W	H	Grö	Gw	B	1	AI	Gw	Ra	Grö	Gw	AI	Ra	Grö	Gw

# T Statistische Tabellen

## Gerhard Osius

1. Verteilungsfunktion  $\Phi(x)$  der Normalverteilung für

*negative* Argumente

*positive* Argumente

2. Quantile  $z_\alpha$  der  $N(0, 1)$ -Verteilung für  $\alpha = 0,1\%, \dots, 10\%$

Quantile  $t_{FG;\alpha}$  der  $t$ -Verteilung  $t_{FG}$  für  $\alpha = 0,1\%, \dots, 10\%$

Quantile  $z_\alpha$  der  $N(0, 1)$ -Verteilung für  $\alpha = 15\%, \dots, 45\%$

Quantile  $t_{FG;\alpha}$  der  $t$ -Verteilung  $t_{FG}$  für  $\alpha = 15\%, \dots, 45\%$

3. Quantile  $\chi^2_{FG;\alpha}$  der *Chiquadrat*-Verteilung  $\chi^2_{FG}$  für

$FG = 1, \dots, 50$  und  $\alpha = 0,1\%, \dots, 10\%$

$\alpha = 20\%, \dots, 80\%$

$\alpha = 90\%, \dots, 99,9\%$

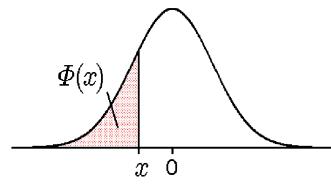
4.  $\alpha$ -Quantile  $F_{m,n;\alpha}$  der  $F$ -Verteilung mit den Freiheitsgraden

$m, n = 1, \dots, 15, 30, 60, 120, 240$  und

$\alpha = 10\%, 5\%, 2,5\%, 1\%, 0,5\%$

**Verteilungsfunktion  $\Phi(x)$  der Normalverteilung  $N(0, 1)$  in Prozent für negative Argumente  $x < 0$**

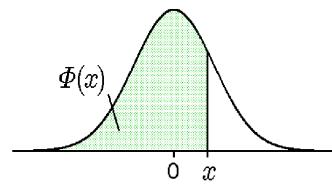
Ablesebeispiel:  $\Phi(-1,00) = 15,87\%$ ,  
 $\Phi(-1,64) = 5,050\%$ .



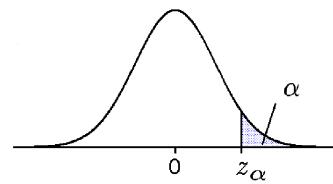
$x$	-,00	-,01	-,02	-,03	-,04	-,05	-,06	-,07	-,08	-,09	$x$
-0,0	50,00	49,60	49,20	48,80	48,40	48,01	47,61	47,21	46,81	46,41	-0,0
-0,1	46,02	45,62	45,22	44,83	44,43	44,04	43,64	43,25	42,86	42,47	-0,1
-0,2	42,07	41,68	41,29	40,90	40,52	40,13	39,74	39,36	38,97	38,59	-0,2
-0,3	38,21	37,83	37,45	37,07	36,69	36,32	35,94	35,57	35,20	34,83	-0,3
-0,4	34,46	34,09	33,72	33,36	33,00	32,64	32,28	31,92	31,56	31,21	-0,4
-0,5	30,85	30,50	30,15	29,81	29,46	29,12	28,77	28,43	28,10	27,76	-0,5
-0,6	27,43	27,09	26,76	26,43	26,11	25,78	25,46	25,14	24,83	24,51	-0,6
-0,7	24,20	23,89	23,58	23,27	22,96	22,66	22,36	22,06	21,77	21,48	-0,7
-0,8	21,19	20,90	20,61	20,33	20,05	19,77	19,49	19,22	18,94	18,67	-0,8
-0,9	18,41	18,14	17,88	17,62	17,36	17,11	16,85	16,60	16,35	16,11	-0,9
-1,0	15,87	15,62	15,39	15,15	14,92	14,69	14,46	14,23	14,01	13,79	-1,0
-1,1	13,57	13,35	13,14	12,92	12,71	12,51	12,30	12,10	11,90	11,70	-1,1
-1,2	11,51	11,31	11,12	10,93	10,75	10,56	10,38	10,20	10,03	9,853	-1,2
-1,3	9,680	9,510	9,342	9,176	9,012	8,851	8,692	8,534	8,379	8,226	-1,3
-1,4	8,076	7,927	7,780	7,636	7,493	7,353	7,215	7,078	6,944	6,811	-1,4
-1,5	6,681	6,552	6,426	6,301	6,178	6,057	5,938	5,821	5,705	5,592	-1,5
-1,6	5,480	5,370	5,262	5,155	5,050	4,947	4,846	4,746	4,648	4,551	-1,6
-1,7	4,457	4,363	4,272	4,182	4,093	4,006	3,920	3,836	3,754	3,673	-1,7
-1,8	3,593	3,515	3,438	3,362	3,288	3,216	3,144	3,074	3,005	2,938	-1,8
-1,9	2,872	2,807	2,743	2,680	2,619	2,559	2,500	2,442	2,385	2,330	-1,9
-2,0	2,275	2,222	2,169	2,118	2,068	2,018	1,970	1,923	1,876	1,831	-2,0
-2,1	1,786	1,743	1,700	1,659	1,618	1,578	1,539	1,500	1,463	1,426	-2,1
-2,2	1,390	1,355	1,321	1,287	1,255	1,222	1,191	1,160	1,130	1,101	-2,2
-2,3	1,072	1,044	1,017	0,990	0,964	0,939	0,914	0,889	0,866	0,842	-2,3
-2,4	0,820	0,798	0,776	0,755	0,734	0,714	0,695	0,676	0,657	0,639	-2,4
-2,5	0,621	0,604	0,587	0,570	0,554	0,539	0,523	0,508	0,494	0,480	-2,5
-2,6	0,466	0,453	0,440	0,427	0,415	0,402	0,391	0,379	0,368	0,357	-2,6
-2,7	0,347	0,336	0,326	0,317	0,307	0,298	0,289	0,280	0,272	0,264	-2,7
-2,8	0,256	0,248	0,240	0,233	0,226	0,219	0,212	0,205	0,199	0,193	-2,8
-2,9	0,187	0,181	0,175	0,169	0,164	0,159	0,154	0,149	0,144	0,139	-2,9
-3,0	0,135	0,131	0,126	0,122	0,118	0,114	0,111	0,107	0,104	0,100	-3,0
-3,1	0,097	0,094	0,090	0,087	0,084	0,082	0,079	0,076	0,074	0,071	-3,1
-3,2	0,069	0,066	0,064	0,062	0,060	0,058	0,056	0,054	0,052	0,050	-3,2
-3,3	0,048	0,047	0,045	0,043	0,042	0,040	0,039	0,038	0,036	0,035	-3,3
-3,4	0,034	0,032	0,031	0,030	0,029	0,028	0,027	0,026	0,025	0,024	-3,4
$x$	-,00	-,01	-,02	-,03	-,04	-,05	-,06	-,07	-,08	-,09	$x$

**Verteilungsfunktion  $\Phi(x)$  der Normalverteilung  $N(0, 1)$  in Prozent für positive Argumente  $x > 0$**

Ablesebeispiel:  $\Phi(+1,00) = 84,13\%$ ,  
 $\Phi(+1,64) = 94,95\%$ .

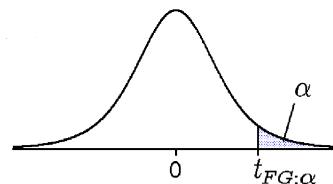


$x$	+,00	+,01	+,02	+,03	+,04	+,05	+,06	+,07	+,08	+,09	$x$
+0,0	50,00	50,40	50,80	51,20	51,60	51,99	52,39	52,79	53,19	53,59	+0,0
+0,1	53,98	54,38	54,78	55,17	55,57	55,96	56,36	56,75	57,14	57,53	+0,1
+0,2	57,93	58,32	58,71	59,10	59,48	59,87	60,26	60,64	61,03	61,41	+0,2
+0,3	61,79	62,17	62,55	62,93	63,31	63,68	64,06	64,43	64,80	65,17	+0,3
+0,4	65,54	65,91	66,28	66,64	67,00	67,36	67,72	68,08	68,44	68,79	+0,4
+0,5	69,15	69,50	69,85	70,19	70,54	70,88	71,23	71,57	71,90	72,24	+0,5
+0,6	72,57	72,91	73,24	73,57	73,89	74,22	74,54	74,86	75,17	75,49	+0,6
+0,7	75,80	76,11	76,42	76,73	77,04	77,34	77,64	77,94	78,23	78,52	+0,7
+0,8	78,81	79,10	79,39	79,67	79,95	80,23	80,51	80,78	81,06	81,33	+0,8
+0,9	81,59	81,86	82,12	82,38	82,64	82,89	83,15	83,40	83,65	83,89	+0,9
+1,0	84,13	84,38	84,61	84,85	85,08	85,31	85,54	85,77	85,99	86,21	+1,0
+1,1	86,43	86,65	86,86	87,08	87,29	87,49	87,70	87,90	88,10	88,30	+1,1
+1,2	88,49	88,69	88,88	89,07	89,25	89,44	89,62	89,80	89,97	90,15	+1,2
+1,3	90,32	90,49	90,66	90,82	90,99	91,15	91,31	91,47	91,62	91,77	+1,3
+1,4	91,92	92,07	92,22	92,36	92,51	92,65	92,79	92,92	93,06	93,19	+1,4
+1,5	93,32	93,45	93,57	93,70	93,82	93,94	94,06	94,18	94,29	94,41	+1,5
+1,6	94,52	94,63	94,74	94,84	94,95	95,05	95,15	95,25	95,35	95,45	+1,6
+1,7	95,54	95,64	95,73	95,82	95,91	95,99	96,08	96,16	96,25	96,33	+1,7
+1,8	96,41	96,49	96,56	96,64	96,71	96,78	96,86	96,93	96,99	97,06	+1,8
+1,9	97,13	97,19	97,26	97,32	97,38	97,44	97,50	97,56	97,61	97,67	+1,9
+2,0	97,72	97,78	97,83	97,88	97,93	97,98	98,03	98,08	98,12	98,17	+2,0
+2,1	98,21	98,26	98,30	98,34	98,38	98,42	98,46	98,50	98,54	98,57	+2,1
+2,2	98,61	98,64	98,68	98,71	98,75	98,78	98,81	98,84	98,87	98,90	+2,2
+2,3	98,93	98,96	98,98	99,01	99,04	99,06	99,09	99,11	99,13	99,16	+2,3
+2,4	99,18	99,20	99,22	99,25	99,27	99,29	99,31	99,32	99,34	99,36	+2,4
+2,5	99,38	99,40	99,41	99,43	99,45	99,46	99,48	99,49	99,51	99,52	+2,5
+2,6	99,53	99,55	99,56	99,57	99,59	99,60	99,61	99,62	99,63	99,64	+2,6
+2,7	99,65	99,66	99,67	99,68	99,69	99,70	99,71	99,72	99,73	99,74	+2,7
+2,8	99,74	99,75	99,76	99,77	99,77	99,78	99,79	99,79	99,80	99,81	+2,8
+2,9	99,81	99,82	99,82	99,83	99,84	99,84	99,85	99,85	99,86	99,86	+2,9
+3,0	99,87	99,87	99,87	99,88	99,88	99,89	99,89	99,89	99,90	99,90	+3,0
+3,1	99,90	99,91	99,91	99,91	99,92	99,92	99,92	99,92	99,93	99,93	+3,1
+3,2	99,93	99,93	99,94	99,94	99,94	99,94	99,94	99,95	99,95	99,95	+3,2
+3,3	99,95	99,95	99,95	99,96	99,96	99,96	99,96	99,96	99,96	99,97	+3,3
+3,4	99,97	99,97	99,97	99,97	99,97	99,97	99,97	99,97	99,97	99,98	+3,4
$x$	+,00	+,01	+,02	+,03	+,04	+,05	+,06	+,07	+,08	+,09	$x$

Quantile  $z_\alpha$  der  $N(0, 1)$ -Verteilung für  $\alpha = 0,1\%, \dots, 10\%$ 

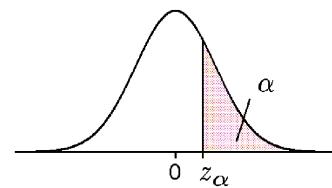
Beispiel: Für  $\alpha = 5\%$  ist  $z_{5\%} = 1,645$ .

$\alpha$	0,1%	0,25%	0,5%	1%	2,5%	5%	10%	$\alpha$
$z_\alpha$	3,090	2,807	2,576	2,326	1,960	1,645	1,282	$z_\alpha$

Quantile  $t_{FG;\alpha}$  der  $t$ -Verteilung  $t_{FG}$  für  $\alpha = 0,1\%, \dots, 10\%$ 

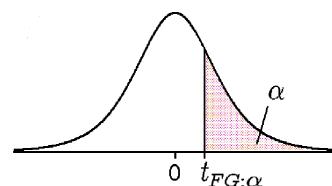
Beispiel: Für  $FG = 10, \alpha = 5\%$  ist  $t_{10;5\%} = 1,812$ .

$FG$	0,1%	0,25%	0,5%	1%	2,5%	5%	10%	$FG$
1	318,29	127,32	63,656	31,821	12,706	6,314	3,078	1
2	22,328	14,089	9,925	6,965	4,303	2,920	1,886	2
3	10,214	7,453	5,841	4,541	3,182	2,353	1,638	3
4	7,173	5,598	4,604	3,747	2,776	2,132	1,533	4
5	5,894	4,773	4,032	3,365	2,571	2,015	1,476	5
6	5,208	4,317	3,707	3,143	2,447	1,943	1,440	6
7	4,785	4,029	3,499	2,998	2,365	1,895	1,415	7
8	4,501	3,833	3,355	2,896	2,306	1,860	1,397	8
9	4,297	3,690	3,250	2,821	2,262	1,833	1,383	9
10	4,144	3,581	3,169	2,764	2,228	1,812	1,372	10
11	4,025	3,497	3,106	2,718	2,201	1,796	1,363	11
12	3,930	3,428	3,055	2,681	2,179	1,782	1,356	12
13	3,852	3,372	3,012	2,650	2,160	1,771	1,350	13
14	3,787	3,326	2,977	2,624	2,145	1,761	1,345	14
15	3,733	3,286	2,947	2,602	2,131	1,753	1,341	15
16	3,686	3,252	2,921	2,583	2,120	1,746	1,337	16
17	3,646	3,222	2,898	2,567	2,110	1,740	1,333	17
18	3,610	3,197	2,878	2,552	2,101	1,734	1,330	18
19	3,579	3,174	2,861	2,539	2,093	1,729	1,328	19
20	3,552	3,153	2,845	2,528	2,086	1,725	1,325	20
21	3,527	3,135	2,831	2,518	2,080	1,721	1,323	21
22	3,505	3,119	2,819	2,508	2,074	1,717	1,321	22
23	3,485	3,104	2,807	2,500	2,069	1,714	1,319	23
24	3,467	3,091	2,797	2,492	2,064	1,711	1,318	24
25	3,450	3,078	2,787	2,485	2,060	1,708	1,316	25
26	3,435	3,067	2,779	2,479	2,056	1,706	1,315	26
27	3,421	3,057	2,771	2,473	2,052	1,703	1,314	27
28	3,408	3,047	2,763	2,467	2,048	1,701	1,313	28
29	3,396	3,038	2,756	2,462	2,045	1,699	1,311	29
30	3,385	3,030	2,750	2,457	2,042	1,697	1,310	30
40	3,307	2,971	2,704	2,423	2,021	1,684	1,303	40
50	3,261	2,937	2,678	2,403	2,009	1,676	1,299	50
60	3,232	2,915	2,660	2,390	2,000	1,671	1,296	60
80	3,195	2,887	2,639	2,374	1,990	1,664	1,292	80
100	3,174	2,871	2,626	2,364	1,984	1,660	1,290	100
200	3,131	2,838	2,601	2,345	1,972	1,653	1,286	200
300	3,118	2,828	2,592	2,339	1,968	1,650	1,284	300
400	3,111	2,823	2,588	2,336	1,966	1,649	1,284	400
500	3,107	2,820	2,586	2,334	1,965	1,648	1,283	500
$\infty$	3,090	2,807	2,576	2,326	1,960	1,645	1,282	$\infty$
$FG$	0,1%	0,25%	0,5%	1%	2,5%	5%	10%	$FG$

Quantile  $z_\alpha$  der  $N(0, 1)$ -Verteilung für  $\alpha = 15\%, \dots, 45\%$ 

Beispiel: Für  $\alpha = 25\%$  ist  $z_{25\%} = 0,674$ .

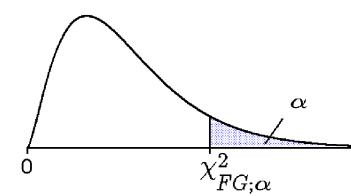
$\alpha$	15%	20%	25%	30%	35%	40%	45%	$\alpha$
$z_\alpha$	1,036	0,842	0,674	0,524	0,385	0,253	0,126	$z_\alpha$

Quantile  $t_{FG;\alpha}$  der  $t$ -Verteilung  $t_{FG}$  für  $\alpha = 15\%, \dots, 45\%$ 

Beispiel: Für  $FG = 10$ ,  $\alpha = 25\%$  ist  $t_{10;25\%} = 0,700$ .

$FG$	15,0%	20,00%	25,0%	30%	35,0%	40%	45%	$FG$
1	1,963	1,376	1,000	0,727	0,510	0,325	0,158	1
2	1,386	1,061	0,816	0,617	0,445	0,289	0,142	2
3	1,250	0,978	0,765	0,584	0,424	0,277	0,137	3
4	1,190	0,941	0,741	0,569	0,414	0,271	0,134	4
5	1,156	0,920	0,727	0,559	0,408	0,267	0,132	5
6	1,134	0,906	0,718	0,553	0,404	0,265	0,131	6
7	1,119	0,896	0,711	0,549	0,402	0,263	0,130	7
8	1,108	0,889	0,706	0,546	0,399	0,262	0,130	8
9	1,100	0,883	0,703	0,543	0,398	0,261	0,129	9
10	1,093	0,879	0,700	0,542	0,397	0,260	0,129	10
11	1,088	0,876	0,697	0,540	0,396	0,260	0,129	11
12	1,083	0,873	0,695	0,539	0,395	0,259	0,128	12
13	1,079	0,870	0,694	0,538	0,394	0,259	0,128	13
14	1,076	0,868	0,692	0,537	0,393	0,258	0,128	14
15	1,074	0,866	0,691	0,536	0,393	0,258	0,128	15
16	1,071	0,865	0,690	0,535	0,392	0,258	0,128	16
17	1,069	0,863	0,689	0,534	0,392	0,257	0,128	17
18	1,067	0,862	0,688	0,534	0,392	0,257	0,127	18
19	1,066	0,861	0,688	0,533	0,391	0,257	0,127	19
20	1,064	0,860	0,687	0,533	0,391	0,257	0,127	20
21	1,063	0,859	0,686	0,532	0,391	0,257	0,127	21
22	1,061	0,858	0,686	0,532	0,390	0,256	0,127	22
23	1,060	0,858	0,685	0,532	0,390	0,256	0,127	23
24	1,059	0,857	0,685	0,531	0,390	0,256	0,127	24
25	1,058	0,856	0,684	0,531	0,390	0,256	0,127	25
26	1,058	0,856	0,684	0,531	0,390	0,256	0,127	26
27	1,057	0,855	0,684	0,531	0,389	0,256	0,127	27
28	1,056	0,855	0,683	0,530	0,389	0,256	0,127	28
29	1,055	0,854	0,683	0,530	0,389	0,256	0,127	29
30	1,055	0,854	0,683	0,530	0,389	0,256	0,127	30
40	1,050	0,851	0,681	0,529	0,388	0,255	0,126	40
50	1,047	0,849	0,679	0,528	0,388	0,255	0,126	50
60	1,045	0,848	0,679	0,527	0,387	0,254	0,126	60
80	1,043	0,846	0,678	0,526	0,387	0,254	0,126	80
100	1,042	0,845	0,677	0,526	0,386	0,254	0,126	100
200	1,039	0,843	0,676	0,525	0,386	0,254	0,126	200
300	1,038	0,843	0,675	0,525	0,386	0,254	0,126	300
400	1,038	0,843	0,675	0,525	0,386	0,254	0,126	400
500	1,038	0,842	0,675	0,525	0,386	0,253	0,126	500
$\infty$	1,036	0,842	0,674	0,524	0,385	0,253	0,126	$\infty$
$FG$	15,0%	20,00%	25,0%	30%	35,0%	40%	45%	$FG$

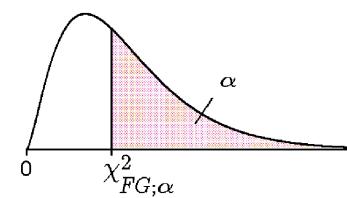
Quantile  $\chi_{FG;\alpha}^2$  der Chiquadrat-Verteilung  $\chi_{FG}^2$  für  
 $FG = 1, \dots, 50$  (Zeilen) und  $\alpha = 0,1\%, \dots, 10\%$  (Spalten)



Beispiel: Für  $FG = 10$ ,  $\alpha = 5\%$  ist  $\chi_{10;5\%}^2 = 18,307$ .

<b><math>FG</math></b>	<b>0,1%</b>	<b>0,25%</b>	<b>0,5%</b>	<b>1%</b>	<b>2,5%</b>	<b>5%</b>	<b>10%</b>	<b><math>FG</math></b>
1	10,827	9,140	7,879	6,635	5,024	3,841	2,706	1
2	13,815	11,983	10,597	9,210	7,378	5,991	4,605	2
3	16,266	14,320	12,838	11,345	9,348	7,815	6,251	3
4	18,466	16,424	14,860	13,277	11,143	9,488	7,779	4
5	20,515	18,385	16,750	15,086	12,832	11,070	9,236	5
6	22,457	20,249	18,548	16,812	14,449	12,592	10,645	6
7	24,321	22,040	20,278	18,475	16,013	14,067	12,017	7
8	26,124	23,774	21,955	20,090	17,535	15,507	13,362	8
9	27,877	25,463	23,589	21,666	19,023	16,919	14,684	9
10	29,588	27,112	25,188	23,209	20,483	18,307	15,987	10
11	31,264	28,729	26,757	24,725	21,920	19,675	17,275	11
12	32,909	30,318	28,300	26,217	23,337	21,026	18,549	12
13	34,527	31,883	29,819	27,688	24,736	22,362	19,812	13
14	36,124	33,426	31,319	29,141	26,119	23,685	21,064	14
15	37,698	34,949	32,801	30,578	27,488	24,996	22,307	15
16	39,252	36,456	34,267	32,000	28,845	26,296	23,542	16
17	40,791	37,946	35,718	33,409	30,191	27,587	24,769	17
18	42,312	39,422	37,156	34,805	31,526	28,869	25,989	18
19	43,819	40,885	38,582	36,191	32,852	30,144	27,204	19
20	45,314	42,336	39,997	37,566	34,170	31,410	28,412	20
21	46,796	43,775	41,401	38,932	35,479	32,671	29,615	21
22	48,268	45,204	42,796	40,289	36,781	33,924	30,813	22
23	49,728	46,623	44,181	41,638	38,076	35,172	32,007	23
24	51,179	48,034	45,558	42,980	39,364	36,415	33,196	24
25	52,619	49,435	46,928	44,314	40,646	37,652	34,382	25
26	54,051	50,829	48,290	45,642	41,923	38,885	35,563	26
27	55,475	52,215	49,645	46,963	43,195	40,113	36,741	27
28	56,892	53,594	50,994	48,278	44,461	41,337	37,916	28
29	58,301	54,966	52,335	49,588	45,722	42,557	39,087	29
30	59,702	56,332	53,672	50,892	46,979	43,773	40,256	30
31	61,098	57,692	55,002	52,191	48,232	44,985	41,422	31
32	62,487	59,046	56,328	53,486	49,480	46,194	42,585	32
33	63,869	60,395	57,648	54,775	50,725	47,400	43,745	33
34	65,247	61,738	58,964	56,061	51,966	48,602	44,903	34
35	66,619	63,076	60,275	57,342	53,203	49,802	46,059	35
36	67,985	64,410	61,581	58,619	54,437	50,998	47,212	36
37	69,348	65,738	62,883	59,893	55,668	52,192	48,363	37
38	70,704	67,063	64,181	61,162	56,895	53,384	49,513	38
39	72,055	68,383	65,475	62,428	58,120	54,572	50,660	39
40	73,403	69,699	66,766	63,691	59,342	55,758	51,805	40
41	74,744	71,011	68,053	64,950	60,561	56,942	52,949	41
42	76,084	72,319	69,336	66,206	61,777	58,124	54,090	42
43	77,418	73,624	70,616	67,459	62,990	59,304	55,230	43
44	78,749	74,925	71,892	68,710	64,201	60,481	56,369	44
45	80,078	76,223	73,166	69,957	65,410	61,656	57,505	45
46	81,400	77,517	74,437	71,201	66,616	62,830	58,641	46
47	82,720	78,808	75,704	72,443	67,821	64,001	59,774	47
48	84,037	80,096	76,969	73,683	69,023	65,171	60,907	48
49	85,350	81,382	78,231	74,919	70,222	66,339	62,038	49
50	86,660	82,664	79,490	76,154	71,420	67,505	63,167	50

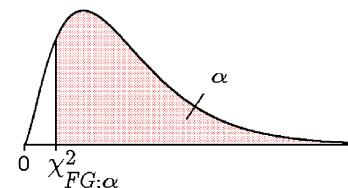
Quantile  $\chi^2_{FG;\alpha}$  der Chiquadrat-Verteilung  $\chi^2_{FG}$  für  
 $FG = 1, \dots, 50$  (Zeilen) und  $\alpha = 20\%, \dots, 80\%$  (Spalten)



Beispiel: Für  $FG = 10$ ,  $\alpha = 50\%$  ist  $\chi^2_{10;50\%} = 9,342$ .

$FG$	20%	30%	40%	50%	60%	70%	80%	$FG$
1	1,642	1,074	0,708	0,455	0,275	0,148	0,064	1
2	3,219	2,408	1,833	1,386	1,022	0,713	0,446	2
3	4,642	3,665	2,946	2,366	1,869	1,424	1,005	3
4	5,989	4,878	4,045	3,357	2,753	2,195	1,649	4
5	7,289	6,064	5,132	4,351	3,656	3,000	2,343	5
6	8,558	7,231	6,211	5,348	4,570	3,828	3,070	6
7	9,803	8,383	7,283	6,346	5,493	4,671	3,822	7
8	11,030	9,524	8,351	7,344	6,423	5,527	4,594	8
9	12,242	10,656	9,414	8,343	7,357	6,393	5,380	9
10	13,442	11,781	10,473	9,342	8,295	7,267	6,179	10
11	14,631	12,899	11,530	10,341	9,237	8,148	6,989	11
12	15,812	14,011	12,584	11,340	10,182	9,034	7,807	12
13	16,985	15,119	13,636	12,340	11,129	9,926	8,634	13
14	18,151	16,222	14,685	13,339	12,078	10,821	9,467	14
15	19,311	17,322	15,733	14,339	13,030	11,721	10,307	15
16	20,465	18,418	16,780	15,338	13,983	12,624	11,152	16
17	21,615	19,511	17,824	16,338	14,937	13,531	12,002	17
18	22,760	20,601	18,868	17,338	15,893	14,440	12,857	18
19	23,900	21,689	19,910	18,338	16,850	15,352	13,716	19
20	25,038	22,775	20,951	19,337	17,809	16,266	14,578	20
21	26,171	23,858	21,992	20,337	18,768	17,182	15,445	21
22	27,301	24,939	23,031	21,337	19,729	18,101	16,314	22
23	28,429	26,018	24,069	22,337	20,690	19,021	17,187	23
24	29,553	27,096	25,106	23,337	21,652	19,943	18,062	24
25	30,675	28,172	26,143	24,337	22,616	20,867	18,940	25
26	31,795	29,246	27,179	25,336	23,579	21,792	19,820	26
27	32,912	30,319	28,214	26,336	24,544	22,719	20,703	27
28	34,027	31,391	29,249	27,336	25,509	23,647	21,588	28
29	35,139	32,461	30,283	28,336	26,475	24,577	22,475	29
30	36,250	33,530	31,316	29,336	27,442	25,508	23,364	30
31	37,359	34,598	32,349	30,336	28,409	26,440	24,255	31
32	38,466	35,665	33,381	31,336	29,376	27,373	25,148	32
33	39,572	36,731	34,413	32,336	30,344	28,307	26,042	33
34	40,676	37,795	35,444	33,336	31,313	29,242	26,938	34
35	41,778	38,859	36,475	34,336	32,282	30,178	27,836	35
36	42,879	39,922	37,505	35,336	33,252	31,115	28,735	36
37	43,978	40,984	38,535	36,336	34,222	32,053	29,635	37
38	45,076	42,045	39,564	37,335	35,192	32,992	30,537	38
39	46,173	43,105	40,593	38,335	36,163	33,932	31,441	39
40	47,269	44,165	41,622	39,335	37,134	34,872	32,345	40
41	48,363	45,224	42,651	40,335	38,105	35,813	33,251	41
42	49,456	46,282	43,679	41,335	39,077	36,755	34,157	42
43	50,548	47,339	44,706	42,335	40,050	37,698	35,065	43
44	51,639	48,396	45,734	43,335	41,022	38,641	35,974	44
45	52,729	49,452	46,761	44,335	41,995	39,585	36,884	45
46	53,818	50,507	47,787	45,335	42,968	40,529	37,795	46
47	54,906	51,562	48,814	46,335	43,942	41,474	38,708	47
48	55,993	52,616	49,840	47,335	44,915	42,420	39,621	48
49	57,079	53,670	50,866	48,335	45,889	43,366	40,534	49
50	58,164	54,723	51,892	49,335	46,864	44,313	41,449	50

Quantile  $\chi^2_{FG;\alpha}$  der Chiquadrat-Verteilung  $\chi^2_{FG}$  für  
 $FG = 1, \dots, 50$  (Zeilen) und  $\alpha = 90\%, \dots, 99,9\%$  (Spalten)

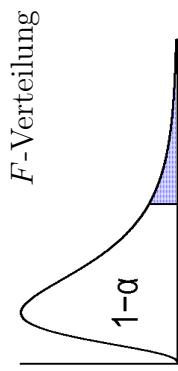


Beispiel: Für  $FG = 10$ ,  $\alpha = 95\%$  ist  $\chi^2_{10;95\%} = 3,940$ .

$FG$	90%	95%	97,5%	99%	99,5%	99,75%	99,9%	$FG$
1	0,016	0,004	0,001	0,000	0,000	0,000	0,000	1
2	0,211	0,103	0,051	0,020	0,010	0,005	0,002	2
3	0,584	0,352	0,216	0,115	0,072	0,045	0,024	3
4	1,064	0,711	0,484	0,297	0,207	0,145	0,091	4
5	1,610	1,145	0,831	0,554	0,412	0,307	0,210	5
6	2,204	1,635	1,237	0,872	0,676	0,527	0,381	6
7	2,833	2,167	1,690	1,239	0,989	0,794	0,599	7
8	3,490	2,733	2,180	1,647	1,344	1,104	0,857	8
9	4,168	3,325	2,700	2,088	1,735	1,450	1,152	9
10	4,865	3,940	3,247	2,558	2,156	1,827	1,479	10
11	5,578	4,575	3,816	3,053	2,603	2,232	1,834	11
12	6,304	5,226	4,404	3,571	3,074	2,661	2,214	12
13	7,041	5,892	5,009	4,107	3,565	3,112	2,617	13
14	7,790	6,571	5,629	4,660	4,075	3,582	3,041	14
15	8,547	7,261	6,262	5,229	4,601	4,070	3,483	15
16	9,312	7,962	6,908	5,812	5,142	4,573	3,942	16
17	10,085	8,672	7,564	6,408	5,697	5,092	4,416	17
18	10,865	9,390	8,231	7,015	6,265	5,623	4,905	18
19	11,651	10,117	8,907	7,633	6,844	6,167	5,407	19
20	12,443	10,851	9,591	8,260	7,434	6,723	5,921	20
21	13,240	11,591	10,283	8,897	8,034	7,289	6,447	21
22	14,041	12,338	10,982	9,542	8,643	7,865	6,983	22
23	14,848	13,091	11,689	10,196	9,260	8,450	7,529	23
24	15,659	13,848	12,401	10,856	9,886	9,044	8,085	24
25	16,473	14,611	13,120	11,524	10,520	9,646	8,649	25
26	17,292	15,379	13,844	12,198	11,160	10,256	9,222	26
27	18,114	16,151	14,573	12,878	11,808	10,873	9,803	27
28	18,939	16,928	15,308	13,565	12,461	11,497	10,391	28
29	19,768	17,708	16,047	14,256	13,121	12,128	10,986	29
30	20,599	18,493	16,791	14,953	13,787	12,765	11,588	30
31	21,434	19,281	17,539	15,655	14,458	13,407	12,196	31
32	22,271	20,072	18,291	16,362	15,134	14,055	12,810	32
33	23,110	20,867	19,047	17,073	15,815	14,709	13,431	33
34	23,952	21,664	19,806	17,789	16,501	15,368	14,057	34
35	24,797	22,465	20,569	18,509	17,192	16,032	14,688	35
36	25,643	23,269	21,336	19,233	17,887	16,700	15,324	36
37	26,492	24,075	22,106	19,960	18,586	17,373	15,965	37
38	27,343	24,884	22,878	20,691	19,289	18,050	16,611	38
39	28,196	25,695	23,654	21,426	19,996	18,732	17,261	39
40	29,051	26,509	24,433	22,164	20,707	19,417	17,917	40
41	29,907	27,326	25,215	22,906	21,421	20,106	18,576	41
42	30,765	28,144	25,999	23,650	22,138	20,799	19,238	42
43	31,625	28,965	26,785	24,398	22,860	21,496	19,905	43
44	32,487	29,787	27,575	25,148	23,584	22,196	20,576	44
45	33,350	30,612	28,366	25,901	24,311	22,899	21,251	45
46	34,215	31,439	29,160	26,657	25,041	23,606	21,929	46
47	35,081	32,268	29,956	27,416	25,775	24,316	22,610	47
48	35,949	33,098	30,754	28,177	26,511	25,029	23,294	48
49	36,818	33,930	31,555	28,941	27,249	25,745	23,983	49
50	37,689	34,764	32,357	29,707	27,991	26,464	24,674	50

$\alpha$ -Quantil  $F_{m,n;\alpha}$  der  $F_{m,n}$ -Verteilung mit den Freiheitsgraden m (Zähler), n (Nenner)

$\alpha = 10\%$       Beispiele:       $F_{5,14;10\%} = 2,307$        $F_{14,5;10\%} = 3,247$



$n$	$m$													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,47	60,71	60,90	61,07
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,401	9,408	9,415	9,420
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,222	5,216	5,210	5,205
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,907	3,896	3,886	3,878
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,282	3,268	3,257	3,247
6	3,776	3,463	3,289	3,181	3,108	3,055	3,014	2,983	2,958	2,937	2,920	2,905	2,892	2,881
7	3,589	3,257	3,074	2,961	2,883	2,827	2,785	2,752	2,725	2,703	2,684	2,668	2,654	2,643
8	3,458	3,113	2,924	2,806	2,726	2,668	2,624	2,589	2,561	2,538	2,519	2,502	2,488	2,475
9	3,360	3,006	2,813	2,693	2,611	2,551	2,505	2,469	2,440	2,416	2,396	2,379	2,364	2,351
10	3,285	2,924	2,728	2,605	2,522	2,461	2,414	2,377	2,347	2,323	2,302	2,284	2,269	2,255
11	3,225	2,860	2,660	2,536	2,451	2,389	2,342	2,304	2,274	2,248	2,227	2,209	2,193	2,179
12	3,177	2,807	2,606	2,480	2,394	2,331	2,283	2,245	2,214	2,188	2,166	2,147	2,131	2,117
13	3,136	2,763	2,560	2,434	2,347	2,283	2,234	2,195	2,164	2,138	2,116	2,097	2,080	2,066
14	3,102	2,726	2,522	2,395	2,307	2,243	2,193	2,154	2,122	2,095	2,073	2,054	2,037	2,022
15	3,073	2,695	2,490	2,361	2,273	2,208	2,158	2,119	2,086	2,059	2,037	2,017	2,000	1,985
16	3,048	2,668	2,462	2,333	2,244	2,178	2,128	2,088	2,055	2,028	2,005	1,985	1,968	1,953
17	3,026	2,645	2,437	2,308	2,218	2,152	2,102	2,061	2,028	2,001	1,978	1,958	1,940	1,925
18	3,007	2,624	2,416	2,286	2,196	2,130	2,079	2,038	2,005	1,977	1,954	1,933	1,916	1,897
19	2,990	2,606	2,397	2,266	2,176	2,109	2,058	2,017	1,984	1,956	1,932	1,912	1,894	1,878
20	2,975	2,589	2,380	2,249	2,158	2,091	2,040	1,999	1,965	1,937	1,913	1,892	1,875	1,859
21	2,961	2,575	2,365	2,233	2,142	2,075	2,023	1,982	1,948	1,920	1,896	1,875	1,857	1,841
22	2,949	2,561	2,351	2,219	2,128	2,060	2,008	1,967	1,933	1,904	1,880	1,859	1,841	1,825
23	2,937	2,549	2,339	2,207	2,115	2,047	1,995	1,953	1,919	1,890	1,866	1,845	1,827	1,811
24	2,927	2,538	2,327	2,195	2,103	2,035	1,983	1,941	1,906	1,877	1,853	1,832	1,814	1,797
25	2,918	2,528	2,317	2,184	2,092	2,024	1,971	1,929	1,895	1,866	1,841	1,820	1,795	1,771
26	2,909	2,519	2,307	2,174	2,082	2,014	1,961	1,919	1,884	1,855	1,830	1,809	1,790	1,774
27	2,901	2,511	2,299	2,165	2,073	2,005	1,952	1,909	1,874	1,845	1,820	1,799	1,780	1,764
28	2,894	2,503	2,291	2,157	2,064	1,996	1,943	1,900	1,865	1,836	1,811	1,790	1,771	1,754
29	2,887	2,495	2,283	2,149	2,057	1,988	1,935	1,892	1,857	1,827	1,802	1,781	1,762	1,745
30	2,881	2,489	2,276	2,142	2,049	1,980	1,927	1,884	1,849	1,819	1,794	1,773	1,754	1,731
60	2,791	2,393	2,177	2,041	1,946	1,875	1,819	1,775	1,738	1,707	1,680	1,657	1,637	1,603
120	2,748	2,347	2,130	1,992	1,896	1,824	1,767	1,722	1,684	1,652	1,625	1,601	1,580	1,554
240	2,727	2,325	2,107	1,968	1,871	1,799	1,742	1,696	1,658	1,625	1,598	1,573	1,552	1,533

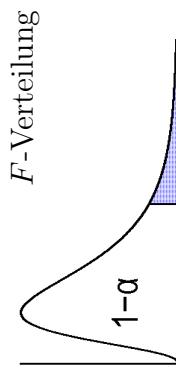
$\alpha$ -Quantil  $F_{m,n;\alpha}$  der  $F_{m,n}$ -Verteilung mit den Freiheitsgraden m (Zähler), n (Nenner)

$\alpha = 5\%$

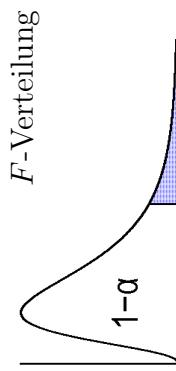
Beispiele:

$F_{5,14;5\%} = 2,958$

$F_{14,5;5\%} = 4,636$



$n$	$m$																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	30	60	120	240
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0	243,9	244,7	245,4	245,9	250,1	252,2	253,3	253,8
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,46	19,48	19,49	19,49	19,49	19,49
3	10,13	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785	8,745	8,729	8,715	8,703	8,617	8,572	8,549	8,538	8,538
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936	5,912	5,891	5,873	5,858	5,746	5,688	5,658	5,643
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,704	4,678	4,655	4,636	4,619	4,496	4,431	4,398	4,382
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,027	4,000	3,976	3,956	3,938	3,808	3,740	3,705	3,687
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,603	3,575	3,550	3,529	3,511	3,376	3,304	3,267	3,249
8	5,318	4,459	4,066	3,838	3,688	3,581	3,500	3,438	3,388	3,347	3,313	3,284	3,259	3,237	3,218	3,079	3,005	2,967	2,947
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,102	3,048	3,025	3,006	2,864	2,787	2,748	2,727	2,727
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,943	2,887	2,865	2,845	2,700	2,621	2,580	2,559	2,559
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854	2,818	2,788	2,761	2,739	2,719	2,570	2,490	2,448	2,426
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753	2,717	2,687	2,660	2,637	2,617	2,466	2,384	2,341	2,319
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671	2,635	2,604	2,577	2,554	2,533	2,380	2,297	2,252	2,230
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602	2,565	2,534	2,507	2,484	2,463	2,308	2,223	2,178	2,155
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544	2,507	2,475	2,448	2,424	2,403	2,247	2,160	2,114	2,090
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494	2,456	2,425	2,397	2,373	2,352	2,194	2,106	2,059	2,035
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450	2,413	2,381	2,353	2,329	2,308	2,148	2,058	2,011	1,986
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412	2,374	2,342	2,314	2,290	2,269	2,107	2,017	1,968	1,943
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378	2,340	2,308	2,280	2,256	2,234	2,071	1,980	1,930	1,905
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348	2,310	2,278	2,250	2,225	2,203	2,039	1,946	1,896	1,870
21	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,366	2,321	2,283	2,250	2,222	2,197	2,176	2,010	1,916	1,866	1,839
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342	2,297	2,259	2,226	2,198	2,173	2,151	1,984	1,889	1,838	1,811
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320	2,275	2,236	2,204	2,175	2,150	2,128	1,961	1,865	1,813	1,785
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255	2,216	2,183	2,155	2,130	2,108	1,939	1,842	1,790	1,762
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236	2,198	2,165	2,136	2,111	2,089	1,919	1,822	1,768	1,740
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265	2,220	2,181	2,148	2,119	2,094	2,072	1,901	1,803	1,749	1,720
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250	2,204	2,166	2,132	2,103	2,078	2,056	1,884	1,785	1,731	1,702
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190	2,151	2,118	2,089	2,064	2,041	1,869	1,769	1,714	1,685
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223	2,177	2,138	2,104	2,075	2,050	2,027	1,854	1,754	1,698	1,669
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165	2,126	2,092	2,063	2,037	2,015	1,841	1,740	1,683	1,654
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993	1,952	1,917	1,887	1,860	1,836	1,649	1,534	1,467	1,430
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959	1,910	1,869	1,834	1,803	1,775	1,750	1,554	1,429	1,352	1,307
240	3,881	3,033	2,642	2,409	2,252	2,136	2,048	1,977	1,919	1,870	1,829	1,793	1,761	1,733	1,708	1,507	1,375	1,290	1,237

$\alpha$ -Quantil  $F_{m,n;\alpha}$  der  $F_{m,n}$ -Verteilung mit den Freiheitsgraden m (Zähler), n (Nenner) $\alpha = 2,5\%$       Beispiele:       $F_{5,14;2,5\%} = 3,663$        $F_{14,5;2,5\%} = 6,456$ 

n	m																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	30	60	120	240
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,6	963,3	968,6	973,0	976,7	979,8	982,5	984,9	1001,4	1009,8	1014,0	1016,1
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,42	39,43	39,43	39,46	39,48	39,48	39,49	39,49
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,37	14,34	14,30	14,28	14,25	14,08	13,99	13,95	13,92
4	12,22	10,65	9,979	9,604	9,364	9,197	9,074	8,980	8,905	8,844	8,794	8,751	8,715	8,684	8,657	8,461	8,360	8,309	8,283
5	10,01	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,681	6,619	6,568	6,525	6,488	6,456	6,428	6,227	6,123	6,069	6,042
6	8,813	7,260	6,599	6,227	5,988	5,820	5,695	5,600	5,523	5,461	5,410	5,366	5,329	5,297	5,269	5,065	4,959	4,904	4,877
7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,823	4,761	4,709	4,666	4,628	4,568	4,568	4,362	4,254	4,199	4,171
8	7,571	6,059	5,416	5,053	4,817	4,652	4,529	4,433	4,357	4,295	4,243	4,200	4,162	4,130	4,101	3,894	3,784	3,728	3,699
9	7,209	5,715	5,078	4,718	4,484	4,320	4,197	4,102	4,026	3,964	3,912	3,868	3,831	3,798	3,769	3,560	3,449	3,392	3,363
10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,779	3,717	3,665	3,621	3,583	3,550	3,522	3,311	3,198	3,140	3,110
11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,588	3,526	3,474	3,430	3,392	3,359	3,330	3,118	3,004	2,944	2,914
12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,436	3,374	3,321	3,277	3,239	3,206	3,177	2,963	2,848	2,787	2,756
13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,312	3,250	3,197	3,153	3,115	3,082	3,053	2,837	2,720	2,659	2,628
14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,209	3,147	3,095	3,050	3,012	2,979	2,949	2,732	2,614	2,552	2,520
15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,123	3,060	3,008	2,963	2,925	2,891	2,862	2,644	2,524	2,461	2,429
16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	3,049	2,986	2,934	2,889	2,851	2,817	2,788	2,568	2,447	2,383	2,350
17	6,042	4,619	4,011	3,665	3,438	3,277	3,156	3,061	2,985	2,922	2,870	2,825	2,786	2,753	2,723	2,502	2,380	2,315	2,282
18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,929	2,866	2,814	2,769	2,730	2,696	2,667	2,445	2,321	2,256	2,222
19	5,922	4,508	3,903	3,559	3,333	3,172	3,051	2,956	2,880	2,817	2,765	2,720	2,681	2,647	2,617	2,394	2,270	2,203	2,169
20	5,871	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,837	2,774	2,721	2,676	2,637	2,603	2,573	2,349	2,223	2,156	2,121
21	5,827	4,420	3,819	3,475	3,250	3,090	2,969	2,874	2,798	2,735	2,682	2,637	2,598	2,564	2,534	2,308	2,182	2,114	2,079
22	5,786	4,383	3,783	3,440	3,215	3,055	2,934	2,839	2,763	2,700	2,647	2,602	2,563	2,528	2,498	2,272	2,145	2,076	2,040
23	5,750	4,349	3,750	3,408	3,183	3,023	2,902	2,808	2,731	2,668	2,615	2,570	2,531	2,497	2,466	2,239	2,111	2,041	2,005
24	5,717	4,319	3,721	3,379	3,155	2,995	2,874	2,779	2,703	2,640	2,586	2,541	2,502	2,468	2,437	2,209	2,080	2,010	1,973
25	5,686	4,291	3,694	3,353	3,129	2,969	2,848	2,753	2,677	2,613	2,560	2,515	2,476	2,441	2,411	2,182	2,052	1,981	1,944
26	5,659	4,265	3,670	3,329	3,105	2,945	2,824	2,729	2,653	2,590	2,536	2,491	2,452	2,417	2,387	2,157	2,026	1,954	1,917
27	5,633	4,242	3,647	3,307	3,083	2,923	2,802	2,707	2,631	2,568	2,514	2,469	2,429	2,395	2,364	2,133	2,002	1,930	1,892
28	5,610	4,221	3,626	3,286	3,063	2,903	2,782	2,687	2,611	2,547	2,494	2,448	2,409	2,374	2,344	2,112	1,980	1,907	1,869
29	5,588	4,201	3,607	3,267	3,044	2,884	2,763	2,669	2,592	2,529	2,475	2,430	2,390	2,355	2,325	2,092	1,959	1,886	1,847
30	5,568	4,182	3,589	3,250	3,026	2,867	2,746	2,651	2,575	2,511	2,458	2,412	2,372	2,338	2,307	2,074	1,940	1,866	1,827
60	5,286	3,925	3,343	3,008	2,786	2,627	2,507	2,412	2,334	2,270	2,216	2,169	2,129	2,093	2,061	1,815	1,667	1,581	1,534
120	5,152	3,805	3,227	2,894	2,674	2,515	2,395	2,299	2,222	2,157	2,102	2,055	2,014	1,977	1,945	1,690	1,530	1,433	1,376
240	5,088	3,746	3,171	2,839	2,620	2,461	2,245	2,167	2,102	2,047	1,999	1,958	1,921	1,888	1,628	1,460	1,354	1,289	1,224

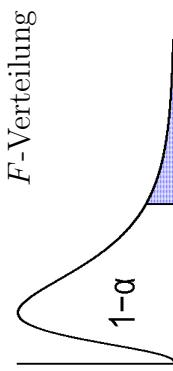
$\alpha$ -Quantil  $F_{m,n;\alpha}$  der  $F_{m,n}$ -Verteilung mit den Freiheitsgraden m (Zähler), n (Nenner)

$\alpha = 1\%$

Beispiele:

$F_{5,14;1\%} = 4,695$

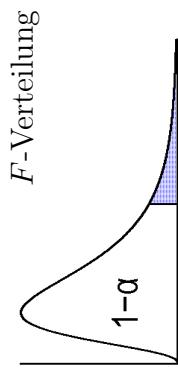
$F_{14,5;1\%} = 9,770$



$n$	$m$													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	4052,2	4999,3	5403,5	5624,3	5764,0	5859,0	5928,3	5981,0	6022,4	6055,9	6083,4	6125,8	6143,0	6157,0
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,41	99,42	99,43	99,47
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05	26,98	26,87
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37	14,31	14,25
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,963	9,888	9,825	9,770
6	13,75	10,925	9,780	9,148	8,746	8,466	8,260	8,102	7,976	7,874	7,790	7,718	7,657	7,605
7	12,25	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,719	6,620	6,538	6,469	6,410	6,359
8	11,26	8,649	7,591	7,006	6,632	6,371	6,178	6,029	5,911	5,814	5,734	5,667	5,609	5,559
9	10,56	8,022	6,992	6,422	6,057	5,802	5,613	5,467	5,351	5,257	5,178	5,111	5,055	5,005
10	10,04	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,942	4,849	4,772	4,706	4,650	4,601
11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,744	4,632	4,539	4,462	4,397	4,342	4,293
12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,388	4,296	4,220	4,155	4,100	4,052
13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,191	4,100	4,025	3,960	3,905	3,857
14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	4,030	3,939	3,864	3,800	3,745	3,698
15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,004	3,895	3,805	3,730	3,666	3,612	3,564
16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,780	3,691	3,616	3,553	3,498	3,451
17	8,400	6,112	5,185	4,669	4,336	4,101	3,927	3,791	3,682	3,593	3,518	3,455	3,401	3,353
18	8,285	6,013	5,092	4,579	4,248	4,015	3,841	3,705	3,597	3,508	3,434	3,371	3,316	3,269
19	8,185	5,926	5,010	4,500	4,171	3,939	3,765	3,631	3,523	3,434	3,360	3,297	3,242	3,195
20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,457	3,368	3,294	3,231	3,177	3,130
21	8,017	5,780	4,874	4,369	4,042	3,872	3,640	3,506	3,398	3,310	3,236	3,173	3,119	3,072
22	7,945	5,719	4,817	4,313	3,988	3,758	3,587	3,453	3,346	3,258	3,184	3,121	3,067	3,019
23	7,881	5,664	4,765	4,264	3,939	3,710	3,539	3,406	3,299	3,211	3,137	3,074	3,020	2,973
24	7,823	5,614	4,718	4,218	3,895	3,667	3,496	3,363	3,256	3,168	3,094	3,032	2,977	2,930
25	7,770	5,568	4,675	4,177	3,855	3,627	3,457	3,324	3,217	3,129	3,056	2,993	2,892	2,850
26	7,721	5,526	4,637	4,140	3,818	3,591	3,421	3,288	3,182	3,094	3,021	2,958	2,904	2,857
27	7,677	5,488	4,601	4,106	3,785	3,558	3,388	3,256	3,149	3,062	2,988	2,926	2,872	2,824
28	7,636	5,453	4,568	4,074	3,754	3,528	3,358	3,226	3,120	3,032	2,959	2,896	2,842	2,795
29	7,598	5,420	4,538	4,045	3,725	3,499	3,330	3,198	3,092	3,005	2,931	2,868	2,814	2,767
30	7,562	5,390	4,510	4,018	3,699	3,473	3,305	3,173	3,067	2,979	2,906	2,843	2,789	2,742
60	7,077	4,977	4,126	3,649	3,339	3,119	2,953	2,823	2,718	2,632	2,559	2,442	2,394	2,352
120	6,851	4,787	3,949	3,480	3,174	2,956	2,792	2,663	2,559	2,472	2,399	2,336	2,282	2,234
240	6,742	4,695	3,864	3,398	3,094	2,878	2,714	2,586	2,482	2,395	2,322	2,260	2,205	2,157

$\alpha$ -Quantil  $F_{m,n;\alpha}$  der  $F_{m,n}$ -Verteilung mit den Freiheitsgraden m (Zähler), n (Nenner)

$\alpha = 0,5\%$       Beispiele:       $F_{5,14,0,5\%} = 5,562$        $F_{14,5,0,5\%} = 13,215$



$n$	$m$													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	16212	19997	21614	22501	23056	23440	23715	23924	24091	24222	24427	24505	24572	24632
2	198,50	199,01	199,16	199,24	199,30	199,33	199,36	199,38	199,39	199,42	199,42	199,42	199,43	199,48
3	55,55	49,80	47,47	46,20	45,39	44,84	44,43	44,13	43,88	43,68	43,52	43,39	43,27	43,17
4	31,33	26,28	24,26	23,15	22,46	21,98	21,62	21,35	21,14	20,97	20,82	20,70	20,60	20,44
5	22,78	18,31	16,53	15,56	14,51	14,20	13,96	13,77	13,62	13,49	13,38	13,29	13,21	13,15
6	18,63	14,54	12,92	12,03	11,46	11,07	10,79	10,57	10,39	10,25	10,13	10,03	9,950	9,878
7	16,24	12,40	10,88	10,05	9,522	9,155	8,885	8,678	8,514	8,380	8,270	8,176	8,097	8,028
8	14,69	11,04	9,597	8,805	8,302	7,952	7,694	7,496	7,339	7,211	7,105	7,015	6,938	6,872
9	13,61	10,11	8,717	7,956	7,471	7,134	6,885	6,693	6,541	6,417	6,314	6,227	6,153	6,089
10	12,83	9,427	8,081	7,343	6,872	6,545	6,303	6,116	5,968	5,847	5,746	5,661	5,589	5,526
11	12,23	8,912	7,600	6,881	6,422	6,102	5,865	5,682	5,537	5,418	5,320	5,236	5,165	5,103
12	11,75	8,510	7,226	6,521	6,071	5,757	5,524	5,345	5,202	5,085	4,988	4,906	4,836	4,775
13	11,37	8,186	6,926	6,233	5,791	5,482	5,253	5,076	4,935	4,820	4,724	4,643	4,573	4,513
14	11,06	7,922	6,680	5,998	5,562	5,257	5,031	4,857	4,717	4,603	4,508	4,428	4,359	4,299
15	10,80	7,701	6,476	5,803	5,372	5,071	4,847	4,674	4,536	4,424	4,329	4,250	4,181	4,122
16	10,58	7,514	6,303	5,638	5,212	4,913	4,692	4,521	4,384	4,272	4,179	4,099	4,031	3,972
17	10,38	7,354	6,156	5,497	5,075	4,779	4,559	4,389	4,254	4,142	4,050	3,971	3,903	3,844
18	10,22	7,215	6,028	5,375	4,956	4,663	4,445	4,276	4,141	4,030	3,938	3,860	3,793	3,734
19	10,07	7,093	5,916	5,268	4,853	4,561	4,345	4,177	4,043	3,933	3,841	3,763	3,696	3,638
20	9,944	6,987	5,818	5,174	4,762	4,472	4,257	4,090	3,956	3,847	3,756	3,678	3,611	3,553
21	9,829	6,891	5,730	5,091	4,681	4,393	4,179	4,013	3,880	3,771	3,680	3,602	3,536	3,478
22	9,727	6,806	5,652	5,017	4,609	4,322	4,109	3,944	3,812	3,703	3,612	3,535	3,469	3,411
23	9,635	6,730	5,582	4,950	4,544	4,259	4,047	3,882	3,750	3,642	3,551	3,474	3,408	3,351
24	9,551	6,661	5,519	4,890	4,486	4,202	3,991	3,826	3,695	3,587	3,497	3,420	3,354	3,296
25	9,475	6,598	5,462	4,835	4,433	4,150	3,939	3,776	3,645	3,537	3,447	3,370	3,304	3,247
26	9,406	6,541	5,409	4,785	4,384	4,103	3,893	3,730	3,599	3,492	3,402	3,325	3,259	3,202
27	9,342	6,489	5,361	4,740	4,340	4,059	3,850	3,687	3,557	3,450	3,360	3,284	3,218	3,161
28	9,284	6,440	5,317	4,698	4,300	4,020	3,811	3,649	3,519	3,412	3,322	3,246	3,180	3,123
29	9,230	6,396	5,276	4,659	4,262	3,983	3,775	3,613	3,483	3,376	3,287	3,211	3,145	3,088
30	9,180	6,355	5,239	4,623	4,228	3,949	3,742	3,580	3,451	3,344	3,255	3,179	3,113	3,056
60	8,495	5,795	4,729	4,140	3,760	3,492	3,291	3,134	3,008	2,904	2,817	2,742	2,677	2,620
120	8,179	5,539	4,497	3,921	3,548	3,285	3,087	2,933	2,808	2,705	2,618	2,544	2,479	2,423
240	8,027	5,417	4,387	3,816	3,447	3,187	2,991	2,837	2,713	2,610	2,524	2,450	2,385	2,329

## Boxenverzeichnis

Rechen-Box 1 . . . . .	10
Rechen-Box 2 . . . . .	27
Rechen-Box 3 . . . . .	32
Rechen-Box 4 . . . . .	60
Konfidenz-Box 1 . . . . .	140
Test-Box 1 . . . . .	152
Power-Box 1 . . . . .	155
Planungs-Box 1 . . . . .	157
Test- und Konfidenz-Box 2 . . . . .	164
Power-Box 2 . . . . .	166
Planungs-Box 2 . . . . .	167
Konfidenz-Box 3 . . . . .	173
Test-Box 3 . . . . .	177
Power-Box 3 . . . . .	180
Planungs-Box 3 . . . . .	182
Konfidenz-Box 4 . . . . .	187
Test-Box 4 . . . . .	192
Power-Box 4 . . . . .	193
Planungs-Box 4 . . . . .	195
Test-Box 5a . . . . .	202
Konfidenz- und Test-Box 5b . . . . .	203
Power-Box 5a . . . . .	205
Prognose-Box 5a . . . . .	209
Test-Box 5c . . . . .	216
Konfidenz- und Test-Box 5d . . . . .	217
Power-Box 5b . . . . .	218
Prognose-Box 5b . . . . .	219
Test-Box 6a . . . . .	234
Test-Box 6b . . . . .	235
Power- und Planungs-Box 6a . . . . .	236
Power- und Planungs-Box 6b . . . . .	237
Test-Box 7 . . . . .	246
Power- und Planungs-Box 7a . . . . .	250
Power- und Planungs-Box 7b . . . . .	258

# Index

Der Index enthält Begriffe aus dem *methodischen Textteil* (also nicht aus *allen Beispielen* und *Anwendungen*) und aus dem *Anhang*. Für jedes Stichwort sind nur die *wichtigsten* (also nicht *alle*) Textstellen aufgeführt, an denen es erwähnt wird.

- absolute Häufigkeit, 3
- äquidistant, 205
- äquivalent, 38
- Alternative, 143
- alternatives Merkmal, 2
- Amplitude, 54
- Anfangswert, 54
- Anpassungstest, 222, 242
- Anstieg, 202, 216
- arithmetisches Mittel, 6
- asymptotische Konfidenzgrenze, 133, 172
- asymptotischer t-Test, 191
- asymptotischer Test, 164
  - einer Wahrscheinlichkeit, 152
- Augensumme, 85
- Auswertung, 68
- b( $-|-, -$ ), 88
- B(n,p)-Verteilung, 88
- balanciert, 249
- bedingt, 242, 243
- bedingte Wahrscheinlichkeit, 79
- Beobachtung, 1
- Bernoulli-Verteilung, 113
- Bestimmtheitsmaß, 21, 22, 27, 34, 60,
  - 200
  - der Regressionsfunktion, 37
  - gewichtet, 32
- Binomial-Verteilung, 87, 88, 106, 113,
  - 121
- Binomial-Wahrscheinlichkeiten, 89
- Binomialkoeffizient, 88, 93
- Bogenmaß, 52
- Chance, 48
- Chebychev, 113
- Chiquaret-Statistik, 224, 228
- Chiquaret-Test, 224, 230, 245, 254
- Chiquaret-Testwert, 245, 254
- Chiquaret-Verteilung, 239
- codiert, 2
- Daten, 1
- deduktive Statistik, 123
- dekadischer Logarithmus, 38
- Dichte, 95
  - der Normalverteilung, 100
  - Funktion, 95
- disjunkt, 77
  - paarweise, 77
- diskrete Gleichverteilung, 85
- diskrete Zufallsvariable, 74, 84, 109
- diskretes Merkmal, 2
- Dosis-Wirkungs-Analyse, 51
- dualer Logarithmus, 38
- Durchschnitt, 77
- Ein-Punkt-Verteilung, 114
- einfache Nullhypothese, 223, 224
- Einflußvariable, 17, 210
- Eintritts-Wahrscheinlichkeit, 87, 124
- Elementar-Ereignis, 71
- Elementar-Wahrscheinlichkeit, 84
- empirische Standardabweichung, 9, 108
- empirische Varianz, 8, 108
- empirischer Mittelwert, 108
- Ereignis, 71, 74
- Ereignis-Rate, 48
- Erfolg, 87
- Ergebnis, 69

- Ergebnis-Wahrscheinlichkeit, 84  
Ergebnisraum, 69  
Erwartungswert, 108–110  
    Normalverteilung, 100, 103, 112  
    Schätzung, 114  
Eulersche Gammafunktion, 184  
Eulersche Zahl, 38  
exakte Konfidenzgrenze, 136, 140, 172, 190  
    Berechnung, 131  
Binomial  
    obere, 128  
    untere, 127  
    zweiseitige, 130  
exakter  
    t-Test, 190  
    Test einer Wahrscheinlichkeit, 149, 152  
Excel, 5, 13, 30, 34, 57, 67, 93, 94, 107, 122, 139, 179, 214, 241, 260  
Experiment, 1  
exponentieller Zusammenhang, 35, 42  
F-Verteilung, 131  
Fakultät, 93  
falsch-negativ, 144  
falsch-positiv, 144  
Fehler 1. Art, 144  
Fehler 2. Art, 144  
Fehlerrisiko 1. bzw. 2. Art, 144  
Fehlerrisiko 2. Art, 153, 165, 180, 193, 204, 230, 231, 249, 256  
Fisher, 131  
Freiheitsgrad, 131, 183, 198, 224, 245  
Frequenz, 3  
Funktion  
    konstante, 37  
    kubische, 65  
    lineare, 17, 36  
    quadratische, 58  
Gammafunktion, 133, 184, 240  
Gauß'sche Glockenkurve, 101  
Geburtsgewicht, 97  
gewichtete  
    kleinste Quadrate, 31  
    Regressionsanalyse, 215  
    Regressionsgerade, 31  
Gewichtsfaktor, 30, 36, 59  
Gewichtung, 30, 215  
Gleichverteilung, 85  
GRAD, 52  
Grad des linearen Zusammenhangs, 21, 22  
Graphik, 1  
Grenzwert  
    von Histogrammen, 95  
Grenzwertsatz  
    von de Moivre und Laplace, 106  
    Zentraler, 100, 104, 115  
grobe Konfidenzgrenze, 137, 140  
grobes Konfidenzintervall, 137  
Häufigkeit, 3, 48  
    absolute, 3  
    relative, 3, 30, 48, 71, 124  
Häufigkeitsinterpretation, 71  
Häufigkeitstabelle, 3, 4  
Hardy-Weinberg Gleichgewicht, 222, 228, 229, 233  
Histogramm, 3–5, 85, 95  
homogen, 192  
homogene Varianzen, 187, 190, 198  
homogener Würfel, 71  
Hypothesen, 143  
Indikator, 87  
Indikator-Variable, 74  
Intervall-Ereignis, 95, 98  
Intervall-Wahrscheinlichkeit, 95  
klassifiziert, 2  
Klassifizierung, 5  
kleinste Quadrate, 19, 36, 59

- gewichtet, 31
- Kodierung, 74
- Kolmogorov, 75
- komplementäre Wahrscheinlichkeit, 76
- komplementäres Ereignis, 76
- Konfidenz-Box 1, 140
- Konfidenz-Box 2, 164
- Konfidenz-Box 3, 173
- Konfidenz-Box 4, 187
- Konfidenz-Box 5, 203, 217
- Konfidenzgrenze, 147, 159, 164, 173, 185, 187, 203, 209, 217, 219
  - für den Erwartungswert, 171
  - für Wahrscheinlichkeit, 126, 133
- Konfidenzintervall, 140, 160, 164, 173, 186, 187, 199, 209, 219
- konstante Funktionen, 37
- Kontingenztafel, 242
- Korrelationskoeffizient, 22, 27, 38, 200
  - gewichtet, 32
- Kreis-Diagramm, 3, 4
- Kreuztabelle, 14
- Kreuzung von Monohybriden, 222, 226, 232
- kritische Grenze, 143
- kritischer Wert, 145
- kubische Funktion, 65
- Lagemaß, 6
- Lageparameter, 100, 112
- Laplace-Verteilung, 85
- Lebensalter, 96
- Leukämiefälle, 120
- Likelihood-Quotienten
  - Statistik, 228, 234
  - Test, 226, 229, 234, 245, 254
  - Testwert, 226, 245, 254
- lineare
  - Funktion, 17, 36
  - Regressionsanalyse, 197
  - Skala, 39
- Transformation, 10
- linearer Zusammenhang, 17, 35, 63
- lineares Regressionsmodell, 197
- linearisierende Transformationen, 46
- log-Chance, 48
- Log-Skala, 39
- logarithmische Skala, 39
- logarithmische Transformation, 38, 39
- logarithmischer Zusammenhang, 35, 40
- Logarithmus, 38
  - dekadischer, 38
  - dualer, 38
  - natürlicher, 38
- Logarithmus-Funktion, 39
- logistischer Zusammenhang, 48, 66
- Logit-Skala, 49
- Logit-Transformation, 48
- LQ = Likelihood-Quotient, 226
- LQ-Testwert, 226
- Maßzahl, 1
- Maximalwert, 54
- Median, 6
- Merkmal, 1
- Mindestumfang, 156, 167, 182, 195, 232
- Mißerfolg, 87
- Mittelwert, 6, 10, 108, 117, 118
- $N(-,-)$  Normalverteilung, 100
- $\mathbb{N}_0$  Menge nicht-neg. ganzer Zahlen, 84
- natürlicher Logarithmus, 38
- negativer Trend, 18
- Nicht-Eintritts-Wahrscheinlichkeit, 87
- nichtzentrale Chiquadrat-Verteilung, 230, 231, 240, 249, 256
- Nichtzentralität, 231
- Normal-Approximation
  - Binomial-Verteilung, 106
  - Poisson-Verteilung, 121
- normale Konfidenzgrenze, 134, 137, 140
- normaler Test einer Wahrscheinlichkeit, 146, 152

- normales Konfidenzintervall, 134  
Normalverteilung, 100, 102–104, 112, 115, 172, 187, 190, 199, 208  
Dichte, 100  
Nullhypothese, 143  
ordinales Merkmal, 2  
P-Wert, 147, 225  
paarweise disjunkt, 77  
Parameter, 18, 36  
Pearson-Clopper, 126  
Pearson-Statistik, 224  
Periode, 54  
periodischer Zusammenhang, 52  
Phase, 54  
Planung, 68  
Planungs-Box 1, 157  
Planungs-Box 2, 167  
Planungs-Box 3, 182  
Planungs-Box 4, 195  
Planungs-Box 6, 236, 237  
Planungs-Box 7, 250, 258  
Plot, 17  
Poisson-Verteilung, 119  
Polygenie, 142  
Polynom, 65  
    2. Grades, 58  
positiver Trend, 18  
Potenz-Zusammenhang, 35, 44  
Power, 153, 165, 180, 193  
Power-Box 1, 155  
Power-Box 2, 166  
Power-Box 3, 180  
Power-Box 4, 193  
Power-Box 5, 205, 218  
Power-Box 6, 236, 237  
Power-Box 7, 250, 258  
Produkt-Ereignis, 80, 82  
Produktraum, 80, 82  
Produktregel, 81, 83  
Prognose, 19, 208, 209, 219  
Prognose-Box 5, 209, 219  
proportionale Varianzen, 215  
Prozentzahl, 48  
Punktwolke, 17  
quadratische Funktionen, 58  
quadratischer  
    logistischer Zusammenhang, 66  
    Zusammenhang, 35, 58, 63  
qualitatives Merkmal, 1, 14, 16  
Quantil der  
    Chiquaret-Verteilung, 239, 269–271  
    F-Verteilung, 131, 132, 272–276  
    Standard-Normalverteilung, 133, 146, 267, 268  
    t-Verteilung, 172, 267, 268  
quantitatives Merkmal, 2, 16, 17  
 $\mathbb{R}$  Menge reeller Zahlen, 39, 75  
RAD, 52  
Rangliste, 6  
Realisierung, 74, 84  
Rechen-Box 1, 10  
Rechen-Box 2, 27  
Rechen-Box 3, 32  
Rechen-Box 4, 60  
Rechengenauigkeit, 11, 28  
reelle Zahl, 39, 75  
Regressionsanalyse, 197, 198, 202, 203, 209, 216, 217, 219  
gewichtete, 215  
Regressionsfunktion, 36, 41, 43, 45, 48, 55, 197  
Regressionsgerade, 19, 27, 36, 198  
gewichtet, 31  
Regressionsmodell, 197  
relative Häufigkeit, 3, 30, 48, 71, 124  
Residuenanalyse, 199, 218  
Residuenplot, 199  
Residuum, 199, 218  
Roulette-Spiel, 69, 71, 78, 86  
Rundung, 2

- Säulen-Diagramm, 3–5, 85  
Schärfe, 154  
Schätzen einer Wahrscheinlichkeit, 124  
Schätzung, 114  
Scheiteldarstellung, 58  
Scheitelkoordinaten, 59  
sichere Ereignis, 76  
Signifikanz der Beobachtung, 147  
Signifikanzniveau, 225  
simultaner Vorgang, 80, 82  
Skala  
    lineare, 39  
    Log-, 39  
    logarithmische, 39  
    Logit-, 49  
Skalen-Transformationen, 46  
Skalenparameter, 100, 112  
Skalenverschiebung, 10, 28  
Sonntagsfrage, 70  
Spannweite, 9  
Stabdiagramm, 85  
Standard-Normalverteilung, 102, 110, 111  
Standardabweichung, 8, 10, 110, 111  
    einer Zufallsvariablen, 108  
    empirische, 9, 108, 114  
    Normalverteilung, 100, 103, 112  
    Schätzung, 114  
Standardisierung, 101  
StatLab-Auswahl 1985, 4, 5, 11, 14–16, 25, 28, 29, 64, 210, 211, 262, 263  
StatLab-Daten, 1, 105, 261  
stetige Zufallsvariable, 75, 95, 110  
stetiges Merkmal, 2  
Stetigkeitskorrektur, 106, 121  
Stichprobe, 6, 114  
Stichproben-Mittelwert, 115  
Stichprobenraum, 69  
Stichprobenumfang, 3  
Stochastik, 68  
stochastischer Vorgang, 68, 69  
Streuung um Mittelwert, 21, 37  
Streuung um Regressionsfunktion, 37  
Streuung um Regressionsgerade, 21, 37  
Streuungsmaß, 6, 8, 9  
Streuungsreduktion, 21  
Student's t-Test, 177, 202, 203, 216, 217  
Student'sche t-Verteilung, 183  
Studie, 1  
Summenregel, 84  
t-Test, 175, 177, 189, 190, 192, 202, 203, 216, 217  
    asymptotischer, 192  
    exakter, 192  
t-Verteilung, 179, 183  
Tabelle, 1  
Test, 141  
    auf Unabhängigkeit, 252  
    einer Wahrscheinlichkeit, 141, 150, 152  
    eines Erwartungswerts, 171, 174  
Test-Box 1, 152  
Test-Box 2, 164  
Test-Box 3, 177  
Test-Box 4, 192  
Test-Box 5, 202, 203, 216, 217  
Test-Box 6, 234, 235  
Test-Box 7, 246  
Testen  
    des Anstiegs, 200  
    zweier Erwartungswerte, 189  
    zweier Wahrscheinlichkeiten, 161  
Testschärfe, 153, 165, 180, 193, 204, 231  
Torten-Diagramm, 3  
Träger, 84  
Transformation  
    lineare, 10  
    linearisierende, 46  
    logarithmische, 38, 39  
    Skalen-, 46  
Trend-Gerade, 19

- unabhängig, 112
- unabhängige Wiederholungen, 83
- Unabhängigkeit
  - von Zufallsvariablen, 82
  - zweier Merkmale, Test, 252
  - zweier Zufallsvariablen, 80
- Ungleichung von Chebychev, 113
- unmögliches Ereignis, 76
- Unterschied
  - zweier Erwartungswerte, 185
  - zweier Wahrscheinlichkeiten, 159
- Untersuchungobjekt, 1
- Urne, 70, 82
- Variable, 1
- Varianz, 8–10, 30, 110, 111
  - einer Zufallsvariablen, 108
  - empirische, 8, 108, 114
  - Schätzung, 116
- Vereinigung, 77
- Vergleich
  - von Verteilungen, 243
  - zweier Erwartungswerte, 185
  - zweier Wahrscheinlichkeiten, 158
- Versuchsplanung, 156, 166, 181, 194, 204, 230, 232, 249, 257
- Verteilung, 84
- Verteilungsfunktion, 90, 98
  - Standard-Normalverteilung, 102, 265, 266
- Vorhersage-Fehler, 19, 20, 31, 36
- Wahlumfrage, 91, 141, 154, 157
- Wahrscheinlichkeit, 48, 75
  - bedingte, 79
  - Eintritts-, 87
  - Elementar-, 84
  - Ergebnis, 84
  - Häufigkeitsinterpretation, 71
  - Intervall-, 95
  - komplementäre, 76
  - Nicht-Eintritts-, 87
- Wahrscheinlichkeitstheorie, 68
- Wiederholungen
  - unabhängige, 83
- Winkelmaß, 52
- Würfel, 69, 71, 80, 81, 85, 86, 109, 111, 112, 117, 118
- Zahlen-Lotto, 79, 86
- zentrale Chiquadrat-Verteilung, 239
- Zentraler Grenzwertsatz, 100, 104, 115
- zentraler Wert, 6
- Ziehung, 242
- Zielvariable, 17, 197
- Zufall, 68
- Zufallsvariable, 74
  - diskrete, 74, 84, 109
  - stetige, 75, 95, 110
- Zufallsvariablen
  - Unabhängigkeit, 80, 82
  - zusammengesetzte Nullhypothese, 223
- Zusammenhang, 14, 17, 35
  - bei qualitativen Merkmalen, 242
  - exponentieller, 35, 42
  - linearer, 35, 63
  - logarithmischer, 35, 40
  - logistischer, 48, 66
  - periodischer, 52
  - Potenz-, 35, 44
  - quadratischer, 35, 58, 63
  - quadratischer logistischer, 66

Vertrieb der Hefte 4, 14, 23, 26 durch Universitätsbuchhandlung, Bibliothekstr. 3, D-28359 Bremen.  
Vertrieb der übrigen Hefte (soweit nicht vergriffen) durch die Autoren oder FB 3 Mathematik/Informatik  
Universität Bremen, Postfach 330440, D-28334 Bremen.

1. Ulrich Krause (1976): Strukturen in unendlichdimensionalen konvexen Mengen, 74 S.
2. Fritz Colonius, Diederich Hinrichsen (1976): Optimal control of hereditary differential systems. Part I, 66 S.
3. Günter Matthiessen (1976): Theorie der heterogenen Algebren, 88 S.
4. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1976): Skript zur Analysis, Band 1 (13. Auflage 2004), 286 S.
5. Wolfgang Schröder (1977): Operator-algebraische Ergodentheorie für Quantensysteme, 59 S.
6. Rolf Röhrlig, Michael Unterstein (1977): Analyse multivariabler Systeme mit Hilfe komplexer Matrixfunktionen, 216 S.
7. Horst Herrlich, Hans-Eberhard Porst, Rudolf-Eberhard Hoffmann, Manfred Bernd Wischnewsky (1976): Nordwestdeutsches Kategorienseminar, 193 S.
8. Fritz Colonius, Diederich Hinrichsen (1977): Optimal Control of Hereditary Differential Systems. Part II: Differential State Space Description, 36 S.
9. Ludwig Arnold (1977): Differentialgleichungen und Regelungstheorie, 185 S.
10. Rudolf Lorenz (1977): Iterative Verfahren zur Lösung großer, dünnbesetzter symmetrischer Eigenwertprobleme, 104 S.
11. Konrad Behnen, Hans-Peter Kinder, Gerhard Osius, Rüdiger Schäfer, Jürgen Timm (1977): Dose-Response-Analysis, 206 S.
12. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1978): Minimalbasen polynomialer Moduln, Strukturindizes und BRUNOVSKY-Transformationen, 53 S.
13. Konrad Behnen (1978): Vorzeichen-Rangtests mit Nullen und Bindungen, 53 S.
14. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer, Eberhard Oeljeklaus (1978): Skript zur Linearen Algebra, Band 1 (15. Auflage 2004), 249 S.
15. Günter Ludyk (1978): Abtastregelung zeitvarianter Einfach- und Mehrfachsysteme, 54 S.
16. Momme Johs Thomsen (1977): Zur Theorie der Fastalgebren, 146 S.
17. Klaus Horneffer, Horst Diehl (1978): Modellrechnungen zur anaeroben Reduktionskinetik des Cytochroms P-450, 34 S.
18. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst, Manfred Bernd Wischnewsky (1979): Structure of Topological Categories, 252 S.
19. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1979): Geometric and moduletheoretic approach to linear systems. Part I: Basic categories and functors, 28 S.
20. Hans-Friedrich Münzner, Dieter Prätzel-Wolters (1979): Geometric and moduletheoretic approach to linear systems. Part II: Moduletheoretic characterization and reachability, 28 S.
21. Eckart Beutler, Hans Kaiser, Günter Matthiessen, Jürgen Timm (1979): Biduale Algebren, 165 S.
22. Horst Diehl, Detlef Harbach, Jürgen Timm (1980): Planung und Auswertung von Atomabsorptions-Spektrometrie-Untersuchungen mit der Additionsmethode, 44 S.
23. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1981): Skript zur Analysis, Band 2 (8. Auflage 2003), 299 S.

24. Horst Herrlich (1981): Categorical Topology 1971-1981, 105 S.
25. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst, Manfred Bernd Wischnewsky (1981): Special Topics in Topology and Category Theory, 108 S.
26. H. Wolfgang Fischer, Jens Gamst, Klaus Horneffer (1984): Skript zur Linearen Algebra, Band 2 (9. Auflage 2005), 257 S.
27. Rudolf-Eberhard Hoffmann (1982): Continuous Lattices and Related Topics, 314 S.
28. Horst Herrlich, Rudolf-Eberhard Hoffmann, Hans-Eberhard Porst (1987): Workshop on Category Theory, 169 S.
29. Harald Boehme (1987): Zur Berufspraxis des Diplommathematikers, 16 S.
30. Jürgen Timm (1986): Mathematische Modelle der Dosis-Wirkungsanalyse bei den experimentellen Untersuchungen der Arbeitsgruppe zur karzinogenen Belastung des Menschen durch Luftverunreinigung, 65 S.
31. Dieter Denneberg (1988): Mathematik für Wirtschaftswissenschaftler. I. Lineare Algebra, 97 S.
32. Peter E. Crouch, Diederich Hinrichsen, Anthony J. Pritchard, Dietmar Salamon (1988, previous edition University of Warwick 1981): Introduction to Mathematical Systems Theory, 244 S.
33. Gerhard Osius (1989): Some Results on Convergence of Moments and Convergence in Distribution with Applications in Statistics, 27 S.
34. Dieter Denneberg (1989): Verzerrte Wahrscheinlichkeiten in der Versicherungsmathematik, Quantilsabhängige Prämienprinzipien, 24 S.
35. Eberhard Oeljeklaus (1989): Birational splitting of homogeneous Albanese bundles, 30 S.
36. Gerhard Osius, Dieter Rojek (1989): Normal Goodness-of-Fit Tests for Parametric Multinomial Models with Large Degrees of Freedom, 38 S.
37. Dieter Denneberg (1990): Mathematik zur Wirtschaftswissenschaft. II. Analysis, 59 S.
38. Ulrich Krause, Cornelia Zahlten (1990): Arithmetik in Krull monoids and the cross number of divisor class groups, 29 S.
39. Dieter Denneberg (1990): Subadditive Measure and Integral, 39 S.
40. Ulrich Krause, Peter Ranft (1991): A limit set trichotomy for monotone nonlinear dynamical systems, 31 S.
41. Angelika van der Linde (1992): Statistical analyses with splines: are they well defined? 22 S.
42. Dieter Denneberg (1992): Lectures on non-additive measure and integral (new edition: Non-additive measure and integral. TDLB 27, Kluwer Academic, Dordrecht (1994)), 114 S.
43. Gerhard Osius (1993): Separating Agreement from Association in Log-linear Models for Square Contingency Tables With Applications, 23 S.
44. Hans-Peter Kinder, Friedrich Liese (1995): Bremen-Rostock Statistik Seminar, 5. - 7. März 1992, 110 S.
45. Dieter Denneberg (1995): Extension of a measurable space and linear representation of the Choquet Integral, 30 S.
46. Dieter Denneberg, Michael Grabisch (1996): Shapley value and interaction index, 20 S.
47. Angelika Bunse-Gerstner, Heike Faßbender (1996): A Jacobi-like method for solving algebraic Riccati equations on parallel computers, 24 S.
48. Hans-Eberhard Porst editor (1997): Categorical methods in algebra and topology - a collection of papers in honour of Horst Herrlich, 498 S.
49. Angelika van der Linde, Gerhard Osius (1997): Estimation of nonparametric risk functions In matched case-control studies, 28 S.

50. Angelika van der Linde (1997): Estimating the smoothing parameter in generalized spline-based regression, 46 S.
51. Ursula Müller, Gerhard Osius (1998): Asymptotic normality of goodness-of-fit statistics for sparse Poisson data, 15 S.
52. Ursula Müller (1999): Nonparametric regression for threshold data, 18 S.
53. Gerhard Osius (2000): The association between two random elements – A complete characterization in terms of odds ratios, 32 S.
54. Horst Herrlich, Hans-E. Porst (2000): CatMAT 2000, Proceedings of the Conference: Categorical Methods in Algebra and Topology, 490 S.
55. Gerhard Osius (2001): A formal derivation of the conditional likelihood for matched case-control studies, 30 S.
56. Angelika van der Linde (2002): Dimension reduction and linear discriminant functions based on an odds ratio parameterization, 46 S.
57. Angelika van der Linde (2002): On the association between a random parameter and an observable, 26 S.
58. Walter Schill, Pascal Wild (2002): Design Optimisation for Logistic Regression in Two-Phase Studies, 53 S.
59. Gerhard Osius (2002): Statistik in den Naturwissenschaften (Digitale Neuauflage als PDF: Sept. 2005), 291 S.