# Geleitwort

Wer sich zum Studium der Psychologie oder der Soziologie entschließt, interessiert sich vor allem für das Erleben und Handeln der Menschen, er will sich und andere beobachten, verborgene Motive ergründen, Ursachen für Konflikte entdecken, seelische Leiden mindern, durch Beratung helfen usw. Mit Mathematik hat er vielfach nichts im Sinne. Sie erscheint ihm oft trocken, lebensfremd und irrelevant für seinen Beruf. Die Vorlesungen zur mathematischen Psychologie und Statistik stoßen daher zunächst überwiegend auf wenig Gegenliebe. Das erlebte ich in vielen Jahren, in denen ich – gemeinsam mit meinem Kollegen Heinz Ebner – Studierende der Psychologie und Pädagogik in die Statistik einführte. Um die emotionalen Barrieren abzubauen, bemühten wir uns, die Inhalte möglichst ansprechend darzustellen, narrensicher zu erklären und ihre Nützlichkeit an praktischen Beispielen eindringlich zu zeigen. Das ist uns offenbar gelungen; denn die Studierenden besuchten unsere Vorlesungen und Übungen gern und erwarben zumeist eine positive Einstellung zur empirischen Methodik, für die statistische Verfahren unverzichtbar sind.

Die positive Resonanz der Lehrveranstaltungen veranlaßte uns, ein Lehrbuch zu schreiben. 1967 erschienen die „Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen“ erstmals im Berliner Verlag Volk und Wissen. Eine erweiterte Fassung fand Zugang zu deutschen Universitäten, wurde auch in fremde Sprachen übersetzt und erwarb in einem Vierteljahrhundert das Label „Clauß-Ebner“.

Die Langlebigkeit des Titels dürfte vor allem auf den hohen Grad an Verständlichkeit zurückzuführen sein. Wir setzten beim Leser nur das mathematische Schulwissen voraus, behandelten die Kennwerte und Prozeduren in enger Bindung an praktisch relevante Sachverhalte und gaben stellenweise rezeptartige Handlungsanweisungen, die bei der Anwendung statistischer Verfahren zu beachten sind und dann zwangsläufig zum richtigen Ergebnis führen. Eine solche didaktische Vereinfachung mag manchem Mathematiker klippschulenhaft erscheinen; der Mehrzahl der Leser kam die Redundanz und straffe Lenkung sehr entgegen.

Als das Buch zwei Jahrzehnte lang im Handel war, hielten wir es für angebracht, eine gründliche Neubearbeitung vorzunehmen. Dazu kam es jedoch nicht. Heinz Ebner verstarb plötzlich, und ich glaubte, die mittlerweile erschienene Fachliteratur könne unser Buch ablösen. Jedoch die Nachfrage blieb bestehen. Ein unveränderter Nachdruck kam nicht zustande. Um die Marktlücke rasch zu schließen, gewann der Verlag Harri Deutsch zwei erfahrene Dresdner Autoren, Falk-Rüdiger Finze und Lothar Partzsch. Sie sollten in der Tradition des Clauß-Ebner ein Statistiklehrbuch verfassen, das zum Gebrauch an Universitäten sowie zum Selbststudium geeignet ist, die bewährten Verfahren in verständlicher Form behandelt und durch neue Methoden – vor allem multivariate und parameterfreie – ergänzt.

Das vorliegende Buch ist das Produkt. Ich identifiziere mich uneingeschränkt mit seinem Inhalt und der Art der Darstellung. Es ehrt mich, daß mein Name den Autorennamen beigefügt wird, um dadurch eine gewisse Kontinuität der Lehrbuchentwicklung zu signalisieren. Freilich ist das Werk eine durchaus eigenständige Leistung der Verfasser. Sie nötigen den Leser zum selbständigen Mitdenken – mehr, als wir ihm das abverlangten, – geben ihm aber klare Anleitung und die Möglichkeit, an Hand von Beispielen zu prüfen, ob er den Text richtig verstanden hat und anwenden kann. Während wir in den 60er Jahren allenfalls die Nutzung einer Tischrechenmaschine empfehlen konnten, stehen dem Leser heute elektronische Taschenrechner und Computer zur Verfügung. Spezielle Statistikrechner enthalten Programme für Signifikanztests und Varianzanalysen. Ihr Einsatz vereinfacht und beschleunigt die Arbeit außerordentlich. Es wäre töricht, von der Nutzung moderner Rechentechnik abzuraten. Sie kann sehr hilfreich sein und befreit von der Mühsal stumpfsinniger Routine. Aber sie ist und bleibt dienendes Hilfsmittel im Prozeß wissenschaftlichen Problemlösens. Der Nutzer muß mit Einsicht und Sachverstand entscheiden, ob bei einem gegebenen Datensatz die Anwendung eines bestimmten Prüfverfahrens statthaft ist und welche inhaltliche bedeutsame Frage auf diese Weise beantwortet werden kann. Ein solches eindringendes Verständnis setzt voraus, daß man sich einige mathematische Grundbegriffe und Kernaussagen der Wahrscheinlichkeitstheorie aneignet. Sie werden im 3. und 6. Kapitel behandelt, soweit sie für statistische Methodik unentbehrlich sind. Der Leser, der beim ersten Zugriff diese Kapitel überspringt, tut gut daran, sich ihnen später aufmerksam zuzuwenden. Andernfalls läuft er Gefahr, in Praktizismus abzusinken und ernste Fehler zu begehen.

Ich halte das Buch von Finze und Partzsch für eine gut gelungene Einführung, die den heutigen Ansprüchen genügt und dem gegenwärtigen Entwicklungsstand der psychologischen Statistik entspricht. Möge das Buch dankbare Leser finden und dazu beitragen, daß die Studierenden humanwissenschaftlicher Disziplinen ihre eventuell vorhandene Aversion überwinden und die Mathematik in angemessener Weise dafür nützen, methodisch kontrolliert Hypothesen zu prüfen, Aussagen über gesetzmäßige Zusammenhänge zu sichern und Ungewißheit in Wissen zu verwandeln.

Leipzig, Dezember 1993 Günter Clauß

# Vorwort

Die Statistik kann bei der Auswertung empirischer Untersuchungen in der Psychologie, Medizin, Pädagogik, Soziologie und in den angrenzenden Wissenschaften ein hilfreiches methodisches Instrumentarium sein. Das vorliegende Lehrbuch wendet sich in erster Linie an Leser, die die genannten Disziplinen studieren oder auf diesen Gebieten arbeiten, und verfolgt das Ziel, dem Leser in möglichst verständlicher Form die entsprechenden Verfahren vorzustellen und ihn zu deren sachkundiger Anwendung zu befähigen. Es eignet sich auch zum Selbststudium, da die statistischen Verfahren mit vollständig durchgerechneten Zahlenbeispielen behandelt werden und es auf allgemeinen Schulkenntnissen im Fach Mathematik aufbaut.

Die Kapitel 3 und 6 wurden von L. Partzsch, die Kapitel 2 und 4 von F.-R. Finze und die Kapitel 1 und 5 gemeinsam erarbeitet. Die Autoren fühlen sich für den gesamten Text verantwortlich.

Bei der Vorbereitung der 5. Auflage wurden neben der Beseitigung noch vorhandener Druckfehler und Ungenauigkeiten ausführlichere Lösungen der Übungsaufgaben sowie detaillierte Hinweise zum Zugang auf das Computerprogramm SPSS eingearbeitet.

Mit dem Dank für entsprechend eingegangene Verbesserungsvorschläge verbinden wir auch weiterhin unser Interesse an sachdienlicher Kritik.

Dresden, Juli 2004 Falk-Rüdiger Finze, Lothar Partzsch

Verbesserungsvorschläge bitte an:

Verlag Harri Deutsch
Gräfstr. 47
D-60486 Frankfurt am Main
E-mail: verlag@harri-deutsch.de
http://www.harri-deutsch.de/verlag/

# Inhaltsverzeichnis


**1 Einleitung** . . . . . **1**
1.1 Grundanliegen der Statistik . . . . . 1
1.2 Die Relativität statistischer Aussagen . . . . . 3
1.3 Zur Anwendung der Statistik in der Psychologie . . . . . 3
1.3.1 Forderungen an empirische Daten . . . . . 3
1.3.2 Vorteile und Grenzen beim Einsatz der Statistik . . . . . 4

**2 Deskriptive Statistik** . . . . . **6**
2.1 Arten der Daten . . . . . 7
2.1.1 Das Messen . . . . . 7
2.1.2 Klassifikation der Skalen . . . . . 8
2.1.2.1 Nominalskalen . . . . . 8
2.1.2.2 Ordinalskalen . . . . . 9
2.1.2.3 Intervallskalen . . . . . 10
2.1.2.4 Absolut- oder Verhältnisskalen . . . . . 11
2.1.3 Informationsgehalt von Daten . . . . . 12
2.1.4 Genauigkeit der Datenerhebung . . . . . 12
2.2 Monovariable Verteilung . . . . . 13
2.2.1 Darstellung monovariabler Verteilungen . . . . . 13
2.2.1.1 Grafische Darstellung bei Nominal- und Ordinalskalen . . . . . 16
2.2.1.2 Grafische Darstellung metrischer Daten . . . . . 20
2.2.1.3 Gruppierung metrischer Daten . . . . . 24
2.2.2 Kennwerte monovariabler Verteilungen . . . . . 27
2.2.2.1 Mittelwerte . . . . . 27
2.2.2.2 Streuwerte . . . . . 37
2.3 Bivariable Verteilungen . . . . . 54
2.3.1 Grafische Darstellungen bivariabler Verteilungen . . . . . 54
2.3.2 Zusammenhangsmaße bei bivariablen Verteilungen . . . . . 57
2.3.2.1 Abhängigkeitsmaße bei alternativen Daten (Phi-, $\text{Phi}_{\text{COLE}}$- und $Q$-Koeffizient) . . . . . 60
2.3.2.2 Kategoriale Daten (Kontingenzkoeffizienten $C$ und $K$) . . . . . 63
2.3.2.3 Metrische Daten (Maßkorrelationskoeffizient oder auch Produkt-Moment-Korrelationskoeffizient $r$) . . . . . 65
2.3.2.4 Ordinale Daten (Rangkorrelationskoeffizient $R$ und Tau nach KENDALL) 68
2.3.2.5 Gemischtes Datenniveau (tetrachorischer, biserialer und punktbiserialer Korrelationskoeffizient) . . . . . 73
2.3.2.6 Lineare Regression, das Bestimmtheitsmaß . . . . . 77
2.3.2.7 Interpretation von Zusammenhangsmaßen . . . . . 83

**3 Wahrscheinlichkeitstheorie** . . . . . **86**
3.1 Das wahrscheinlichkeitstheoretische Grundmodell . . . . . 87
3.1.1 Stichprobenraum, zufällige Ereignisse . . . . . 87
3.1.2 Relative Häufigkeiten . . . . . 92
3.1.3 Die klassische Wahrscheinlichkeit und die geometrische Wahrscheinlichkeit . . . . 95
3.1.3.1 Kombinatorik . . . . . 95

3.1.3.2 Die klassische Wahrscheinlichkeit . . . 102
3.1.3.3 Die geometrische Wahrscheinlichkeit . . . 104
3.1.4 Die axiomatische Definition der Wahrscheinlichkeit und allgemeine Eigenschaften 105
3.1.5 Die bedingte Wahrscheinlichkeit . . . 107
3.1.6 Unabhängigkeit . . . 111
3.1.7 Die Formel der totalen Wahrscheinlichkeit und die Bayessche Formel . . . 113
3.2 Zufallsgrößen und ihre Verteilung . . . 116
3.2.1 Der Begriff der Zufallsgröße . . . 116
3.2.2 Diskrete Zufallsgrößen . . . 119
3.2.2.1 Diskrete Zufallsgrößen und ihre Verteilung . . . 119
3.2.2.2 Erwartungswert und Varianz diskreter Zufallsgrößen . . . 121
3.2.2.3 Spezielle diskrete Verteilungen . . . 126
3.2.3 Stetige Zufallsgrößen . . . 134
3.2.3.1 Allgemeine Grundlagen zu stetigen Zufallsgrößen und deren Verteilung 134
3.2.3.2 Die gleichmäßige stetige Verteilung . . . 141
3.2.3.3 Die Normalverteilung . . . 143
3.2.3.4 Die Prüfverteilungen . . . 152
3.3 Zufällige Vektoren . . . 157
3.3.1 Der Begriff des zufälligen Vektors . . . 157
3.3.2 Diskrete zufällige Vektoren und Transformationen . . . 160
3.3.3 Unabhängigkeit, Kovarianz, Korrelationskoeffizient . . . 165
3.3.4 Die zweidimensionale Normalverteilung . . . 169
3.4 Statistische Grundbegriffe . . . 171
3.4.1 Grundgesamtheit und Stichprobe . . . 171
3.4.2 Mathematischer Aufbau statistischer Tests . . . 174

**4 Statistische Testtheorie . . . 177**
4.1 Einführung . . . 177
4.1.1 Grundbegriffe . . . 177
4.1.2 Klassifikation statistischer Tests . . . 181
4.2 Anpassungstests . . . 184
4.2.1 Alternative Daten (Binomialtest/$u$-Test) . . . 184
4.2.1.1 Der Binomialtest . . . 185
4.2.1.2 Der $u$-Test . . . 186
4.2.1.3 Der $u_{\text{korr}}$-Test . . . 188
4.2.2 Kategoriale Daten (Polynomialtest/$\chi^2$-Anpassungstest) . . . 189
4.2.2.1 Der Polynomialtest . . . 190
4.2.2.2 Der $\chi^2$-Anpassungstest . . . 191
4.2.3 Zur Frage des Anpassungstests für ordinale Daten . . . 194
4.2.4 Metrische Daten . . . 194
4.2.4.1 Der $\chi^2$-Anpassungstest . . . 194
4.2.4.2 Der David-Test . . . 198
4.2.4.3 Der einfache $t$-Test . . . 199
4.2.4.4 Test des Streuungswertes einer Normalverteilung . . . 200
4.2.4.5 Der KOLMOGOROV-Anpassungstest . . . 201
4.2.5 Übersicht über die Anpassungstests . . . 203
4.3 Unterschiedstests . . . 203
4.3.1 Vergleich zweier Verteilungen mit unabhängigen Stichproben . . . 204
4.3.1.1 Alternative Daten . . . 204

4.3.1.2 Der $\chi^2 - k$ mal 2-Feldertest . . . . . 210
4.3.1.3 Unterschiedstest bei ordinalen Daten und zwei Stichproben . . . . . 213
4.3.1.4 Unterschiedstest bei metrischen Daten und zwei Stichproben . . . . . 221
4.3.1.5 Der Vergleich der Unterschiedstests für 2 Verteilungen mit unabhängigen Stichproben . . . . . 229
4.3.2 Der Vergleich zweier Verteilungen mit abhängigen Stichproben . . . . . 230
4.3.2.1 Der Vergleich zweier Verteilungen mit abhängigen Stichproben bei alternativen Daten . . . . . 230
4.3.2.2 Der Symmetrietest von BOWKER . . . . . 233
4.3.2.3 Der Vorzeichentest . . . . . 235
4.3.2.4 Der Vergleich zweier Verteilungen auf der Grundlage abhängiger Stichproben mit metrischen Daten . . . . . 236
4.3.2.5 Der Vergleich der Unterschiedstests für 2 Verteilungen mit abhängigen Stichproben . . . . . 242
4.3.3 Der Vergleich von mehr als zwei Verteilungen auf der Grundlage unabhängiger Stichproben . . . . . 243
4.3.3.1 Der $\chi^2 - 2 \cdot l$-Feldertest (Globalvergleich) . . . . . 244
4.3.3.2 Nachfolgeauswertung und die Konfigurationsfrequenzanalyse für alternative Daten (multipler Vergleich) . . . . . 245
4.3.3.3 Der $\chi^2$-$k$-mal-$l$-Feldertest (Globalvergleich) . . . . . 246
4.3.3.4 Nachfolgeauswertungen und die Konfigurationsfrequenzanalyse für kategoriale Daten (multipler Vergleich) . . . . . 248
4.3.3.5 Der $H$-Test (Globalvergleich) . . . . . 251
4.3.3.6 Tests für Kontraste (Multipler Vergleich) . . . . . 255
4.3.3.7 Parametrische Unterschiedstest bei Verteilungen mit mehr als 2 unabhängigen Stichproben . . . . . 258
4.3.3.8 Vergleich der Verfahren bei mehr als 2 unabhängigen Verteilungen . . . 267
4.3.4 Vergleich von mehr als 2 Verteilungen bei abhängigen Stichproben . . . . . 267
4.3.4.1 Der $Q$-Test von COCHRAN (Globalvergleich) . . . . . 268
4.3.4.2 Multipler Vergleich bei alternativen Daten und abhängigen Stichproben 270
4.3.4.3 Der FRIEDMAN-Test (Globalvergleich) . . . . . 273
4.3.4.4 Der Test auf Kontraste für korrelierende Stichproben (Multipler Vergleich) . . . . . 275
4.3.4.5 Der Vergleich von mehr als 2 abhängigen Stichproben bei metrischen Daten . . . . . 277
4.3.4.6 Übersicht über die Unterschiedstests bei mehr als zwei abhängigen Stichproben . . . . . 278

**5 Ausblick auf die multivariate Statistik . . . . . 279**
5.1 Die Korrelationsanalyse und die Regressionsanalyse . . . . . 279
5.1.1 Die Korrelationsanalyse bei alternativen Daten . . . . . 280
5.1.2 Korrelationsanalyse bei kategorialen Daten . . . . . 281
5.1.3 Korrelationsanalyse bei ordinalen Daten . . . . . 283
5.1.4 Korrelationsanalyse bei metrischen Daten . . . . . 286
5.1.5 Die Regressionsanalyse . . . . . 288
5.1.5.1 Wahrscheinlichkeitstheoretische Modelle der linearen Regression . . . . 289
5.1.5.2 Die Prüfung für den Regressionskoeffizienten $b$ im Modell I . . . . . 292
5.1.5.3 Die Prüfung für den Achsenabschnitt $a$ . . . . . 293
5.1.5.4 Die Prüfung auf Linearität der Regression . . . . . 294

5.2 Die Faktorenanalyse . . . 295
5.2.1 Einleitung . . . 295
5.2.2 Darstellung und Ansatz der Faktorenanalyse . . . 296
5.2.3 Ein Rechenbeispiel der Faktorenanalyse . . . 301
5.2.4 Hinweise zur Faktoreninterpretation . . . 308
5.3 Die Clusteranalyse . . . 309
5.3.1 Einleitung und Begriffsbestimmung . . . 309
5.3.2 Eigenschaften von Gruppen und methodisches Vorgehen bei der Gruppierung . . . 310
5.3.3 Ähnlichkeits- und Distanzmaße . . . 311
5.3.4 Typen, Kriterien und Verfahren der Gruppierung . . . 313
5.3.5 Ein Rechenbeispiel für eine agglomerative, hierarchische, disjunkte Gruppierung 316
5.3.6 Eine Rechenbeispiel für eine agglomerative, hierarchische, nicht disjunkte Gruppierung . . . 318
5.4 Die einfache Varianzanalyse . . . 320
5.4.1 Die einfache Varianzanalyse für unabhängige Stichproben . . . 321
5.4.1.1 Die Bestimmung der Prüfgröße beim Modell I . . . 322
5.4.1.2 Die Tafel der einfachen Varianzanalyse beim Modell I . . . 324
5.4.1.3 Ein Rechenbeispiel zur einfachen Varianzanalyse beim Modell I . . . 325
5.4.1.4 Die einfache Varianzanalyse beim Modell II . . . 327
5.4.2 Die einfache Varianzanalyse für abhängige Stichproben . . . 328
5.4.2.1 Die Berechnung der Prüfgröße bei korrelierenden Stichproben . . . 328
5.4.2.2 Die Tafel der einfachen Varianzanalyse für korrelierende Stichproben . 330
5.4.2.3 Ein Rechenbeispiel für die einfache Varianzanalyse bei korrelierenden Stichproben . . . 331

**6 Mathematische Grundlagen . . . 334**
6.1 Mengenlehre . . . 334
6.1.1 Der Mengenbegriff . . . 334
6.1.2 Verknüpfungen von Mengen . . . 335
6.1.3 Ausführen mehrerer Mengenoperationen, Rechnen mit Mengen . . . 336
6.1.4 Potenzmenge, kartesisches Produkt . . . 338
6.2 Funktionen . . . 340
6.2.1 Relationen und Funktionen . . . 340
6.2.2 Standardbeispiele reeller Funktionen . . . 343
6.2.2.1 Lineare Funktionen . . . 343
6.2.2.2 Quadratische Funktionen . . . 347
6.2.2.3 Exponentialfunktion und Logarithmusfunktion . . . 350
6.2.2.4 Die Gaußsche Glockenkurve . . . 351
6.3 Matrizen . . . 353
6.3.1 Grundbegriffe . . . 353
6.3.2 Rechnen mit Matrizen . . . 356
6.3.3 Vektoren . . . 359
6.4 Eine Rekursionsformel zur Bestimmung der Einzelwahrscheinlichkeiten . . . 360

**7 Tafelanhang . . . 362**

**8 Aufgaben und Lösungen . . . 456**

**9 Literaturverzeichnis . . . 486**

**Sachwortverzeichnis . . . 489**

# 1 Einleitung

Für wissenschaftliche Untersuchungen in der Psychologie, Pädagogik, Soziologie, Medizin und weiteren artverwandten Disziplinen haben statistische Methoden in zunehmendem Maße an Bedeutung gewonnen. Das Anliegen des vorliegenden Lehrbuches besteht darin, einerseits den Leser mit den Methoden der deskriptiven Statistik und darauf aufbauend mit einem angemessenen Fundus an statistischen Verfahren vertraut zu machen und ihm andererseits die wahrscheinlichkeitstheoretischen Grundlagen, die zum Verständnis der Statistik erforderlich sind, zu vermitteln. Entsprechend dieser Zielstellung ist das Buch aufgebaut.

Die Statistik findet man in den Kapiteln 2, 4 und 5. Wir beginnen in Kapitel 2 mit einer umfassenden Darstellung der Methoden der deskriptiven Statistik, die in der Verwendung von tabellarischen Übersichten, grafischen Darstellungen und geeigneten Kennziffern bestehen. Das Kapitel 4 ist das eigentliche Kernstück dieses Buches und befaßt sich mit den in der Psychologie, Pädagogik und den Sozialwissenschaften gebräuchlichsten statistischen Testverfahren. Die Struktur dieses Kapitels ist nach dem Charakter der vorhandenen Daten aufgebaut. Die einzelnen statistischen Verfahren werden dadurch vorgestellt, daß weniger die theoretischen Details ausgeführt werden, sondern vielmehr anhand repräsentativer Anwendungssituationen die konkrete Testdurchführung „rezeptähnlich“ vorgestellt und an einem konkreten Rechenbeispiel nachvollzogen wird. Um beim Lösen einer statistischen Problemstellung die Suche nach einem geeigneten Testverfahren zu erleichtern, befinden sich am Ende jedes größeren Abschnittes tabellarische Übersichten. In der heutigen Zeit gibt es eine Vielzahl von Computerprogrammen zur Bearbeitung statistischer Problemstellungen. Die Autoren vertreten den Standpunkt, daß das Verständnis der Statistik wesentlich gefördert wird, wenn der Lernende an ausgewählten, typischen Beispielen die erforderlichen Rechenschritte wenigstens einmal „zu Fuß“ ausgeführt hat.

In den Kapiteln 3 und 6 werden, anknüpfend an das Schulwissen, die notwendigen Grundlagen behandelt. In Kapitel 3 verfolgen wir das Ziel, die wichtigsten Begriffe der Wahrscheinlichkeitsrechnung möglichst einfach und klar zu beschreiben und darauf aufbauend ein angemessenes Verständnis für dieses Gebiet zu entwickeln. Im einzelnen handelt es sich hier im wesentlichen um das Grundmodell der Wahrscheinlichkeitsrechnung und den Begriff der Zufallsgröße. Kapitel 6 enthält in zusammenfassender Form Grundkenntnisse aus der Mengenlehre, über reelle Funktionen sowie aus der Kombinatorik, und gibt eine elementare Einführung in die Matrizenrechnung.

Die multivariate Statistik betrachtet in zunehmendem Maße derart komplexe Fragestellungen, daß für den Anwender die Nutzung des Computers erforderlich wird. Aus diesem Grund wird im Kapitel 5 anhand einfacher Beispiele versucht, eine erste Einführung in die Korrelations- und Regressionsanalyse, die Faktorenanalyse, die Clusteranalyse und die Varianzanalyse zu vermitteln. Damit soll zum Verständnis von Lösungen, die man durch die Benutzung von Computerprogrammen erhält, beigetragen werden. Das abschließende Kapitel 7 enthält eine Zusammenstellung der notwendigen Tabellen.

## 1.1 Grundanliegen der Statistik

Bei vielen wissenschaftlichen Untersuchungen, insbesondere unter anderem in der Psychologie, Pädagogik, Soziologie und Medizin, hat man es zum Teil mit großen Anzahlen von Daten zu tun, aus denen man nicht mit Sicherheit auf Aussagen über vorhandene Beziehungen und Phänomene schließen kann. Es handelt sich aber häufig um wiederholbare Erfahrungen, so daß man hoffen kann, immer noch Allgemeingültiges, Gesetzmäßiges in derartigen „Massenerscheinungen“ herauszufinden. Während man z. B. bei einem „gezinkten“ Würfel durch eine einzige Messung sein Gewicht mit ausreichender Genauigkeit zweifelsfrei ermitteln kann, läßt sich aus 10 Wurfergebnissen nicht eindeutig beurteilen, welche Wahrscheinlichkeiten den einzelnen Augenzahlen zugeordnet werden sollen. Es gibt aus prinzipiellen oder pragmatischen Gründen zu viele unkontrollierbare Einflüsse, die eine zuverlässige Aussage über die interessierende Eigenschaft nicht gestatten. Wir sprechen dann davon, daß der „Zufall“ seine Hand im Spiel hat, die Meßergebnisse streuen mehr oder weniger unvorhersehbar. Es entstehen gewisse Unschärfeeffekte und Grauzonen. Trotzdem ist man bestrebt, immer noch Allgemeingültiges zu erkennen, eventuelle Gesetzmäßigkeiten zu entdecken, einen Kern in den vielen Grauzonen streuender Werte zu finden. Man möchte gerne „hinter die Kulissen“ schauen, dem Zufall „auf die Finger sehen“, um mehr Sicherheit in der Unsicherheit zu erzielen. Die Statistik liefert hierzu ein geeignetes methodisches Instrumentarium. Sie gibt **Hilfen zur Entscheidung** bei der Auswertung empirischer Daten.

Wir unterscheiden zwischen der beschreibenden (oder auch deskriptiven) und schließenden Statistik. Das Anliegen der deskriptiven Statistik besteht darin, interessierende Daten von großen Anzahlen von Objekten, Personen usw., die man in diesem Zusammenhang auch Grundgesamtheit nennt, anschaulich, übersichtlich und verständlich darzustellen. Dies erfolgt in Listen und Tabellen, in Grafiken oder mit Hilfe von typischen Maßzahlen wie z. B. Mittelwerten und Streuungen. Dabei ist man bestrebt, auf möglichst umfassenden Erhebungen aufzubauen, wie es beispielsweise bei der Erfassung von Einwohnerzahlen in statistischen Jahrbüchern der Fall ist. Mit Hilfe der Methoden der deskriptiven Statistik soll eine Datenvoranalyse gefördert werden. Auf Grund deren unmittelbarer Verständlichkeit stellen wir die deskriptive Statistik an den Anfang (vgl. Kapitel 2).

Bei der schließenden Statistik wird im Unterschied dazu auf der Grundlage von Informationen aus einer Teilmenge (auch Stichprobe genannt) der Grundgesamtheit auf Aussagen über die Grundgesamtheit geschlossen. Es werden also Aussagen über den empirischen Beobachtungsbereich hinaus getroffen. Genauer heißt dies, man stellt ein theoretisches Modell auf und vergleicht es mit den empirischen Informationen. Mit diesem theoretischen Modell versucht man, das Allgemeingültige des Zufallsgeschehens dadurch zu repräsentieren, daß man bestimmte Wahrscheinlichkeitsverteilungen des Auftretens der Beobachtungswerte postuliert. Dieser Vergleich erfolgt auf zwei Wegen. Zum einen gibt es die Möglichkeit, die postulierte Wahrscheinlichkeitsverteilung oder auch nur interessierende Parameter näherungsweise zu bestimmen. Entsprechende Verfahren werden in der Schätzstatistik entwickelt. Zum anderen stellt man Hypothesen über die Modellverteilung auf und überprüft diese auf der Grundlage der empirischen Daten der Stichprobe. Dieses große Teilgebiet nennt man Inferenzstatistik (oder auch Teststatistik). So könnte man z. B. die Behauptung prüfen, ob die Leistung nach einem Trainingskurs besser ist als vorher. Es sei aber explizit darauf verwiesen, daß die Richtigkeit von Aussagen wegen der unvollständigen Informationen – es wurde ja nur eine Teilmenge untersucht – nicht mit absoluter Sicherheit, sondern nur mit einer bestimmten Wahrscheinlichkeit garantiert werden kann. Überdies sollte man beim Durchführen von statistischen Tests die zu überprüfende Hypothese nicht erst in Abhängigkeit vom vorliegenden Datenmaterial formulieren. Ausgehend von den typischen Erfordernissen der Psychologie, Soziologie und Pädagogik werden wir uns in diesem Lehrbuch vorwiegend mit der Inferenzstatistik (vgl. Kapitel 4) beschäftigen. Deren Ergebnis besteht in der Beurteilung empirischer Daten.

Eine weitere Differenzierung der schließenden Statistik ist durch den Grad der Komplexität gegeben. Wir unterscheiden in diesem Zusammenhang zwischen uni- und multivariater Statistik. Zu letzterer zählt man auch Analysen komplexerer Systeme, die dann im Sinne von sogenannten Datenvoranalysen der explorativen Statistik durchgeführt werden. Kapitel 5 enthält einige Ausführungen über dieses große Gebiet der Statistik.

## 1.2 Die Relativität statistischer Aussagen

**Statistische Aussagen** vermitteln, und das wird leider viel zu oft vergessen, immer nur Erkenntnisse über zufallsabhängige Massenerscheinungen und müssen gerade deshalb im Einzelfall nicht zwingend zutreffen. Die Aussage gilt nur für den Bereich **insgesamt**, über den sie gemacht wird – sie ist eine Globalaussage. Stellen wir uns die Situation vor, daß in einem großen Unternehmen der Industrie von **allen** Mitarbeitern die tatsächliche wöchentliche Arbeitszeit bestimmt wurde, und es ergaben sich im Mittel 40 Stunden. Dann gilt die statistische Aussage: „Der Mittelwert der wöchentlichen Arbeitszeit liegt in diesem Unternehmen bei 40 Stunden." nur dann, wenn man vom gesamten Unternehmen spricht. Es ist leicht einsichtig, daß sie bei den einzelnen Arbeitern, Angestellten, Managern usw. verschieden sein kann. Hier variiert sie vielleicht zwischen 38 und 60 Stunden. Es wäre also falsch, aus der statistischen Aussage auf den Einzelfall zu schließen.

Deshalb unterscheiden wir zwischen statistischen und **kasuistischen** Aussagen, d. h., Aussagen über die Grundgesamtheit und Aussagen über den **Einzelfall**. Eine kasuistische Aussage könnte dann z. B. lauten: „Der Betriebsklempner hat eine wöchentliche Arbeitszeit von 42 Stunden." Sie kann demnach nur eine Information darüber liefern, daß ein bestimmtes Element (Person) einer bestimmten Grundgesamtheit (Unternehmensmitarbeiter) ein bestimmtes Merkmal (wöchentliche Arbeitszeit) in einer bestimmten Ausprägung (42 Stunden) aufweist. Es besteht die Möglichkeit, aus einer Vielzahl kasuistischer Aussagen eine statistische Aussage abzuleiten.

Hervorhebenswert ist an dieser Stelle, daß der Unterschied zwischen einer kasuistischen und einer statistischen Aussage immer nur **relativ** ist. Erheben wir z. B. die wöchentliche Arbeitszeit jedes Mitarbeiters und bilden davon pro Abteilung Mittelwerte, dann sind die Aussagen zur Abteilungsarbeitszeit statistische Aussagen. Erheben wir nun aber nur die Arbeitszeit pro Abteilung und bilden daraus die mittlere Unternehmensarbeitszeit, dann sind die Aussagen zur Abteilungsarbeitszeit kasuistische Aussagen. Ein und dieselbe Aussage kann demnach einmal eine kasuistische und einmal eine statistische Aussage sein. Die Entscheidung darüber, welche Art einer Aussage vorliegt, hängt also immer vom einzelnen Betrachter und von den Randbedingungen der Untersuchung und Interpretation ab.

## 1.3 Zur Anwendung der Statistik in der Psychologie

Die Psychologie ist eine Wissenschaft, die sich mit dem Menschen, seinen Fähigkeiten und Fertigkeiten, seinen Motiven und Zielen, seinen Emotionen und Gefühlen, aber auch mit seinen psychischen Störungen und Beeinträchtigungen auseinandersetzt. Entsprechend breit sind die Methoden wissenschaftlicher Analyse und Synthese gefächert. Die Statistik ist **eine** Möglichkeit, Ergebnisse gegen zufällige Einflüsse und Schwankungen zu sichern, systematische Unterschiede aufzudecken und im Rahmen ihrer Grundannahmen zu verifizieren oder zu falsifizieren. Um die Statistik aber sowohl im Einzelfall als auch insgesamt sinnvoll einsetzen zu können, müssen die Daten, die statistisch weiter verarbeitet werden sollen, einige Bedingungen erfüllen, auf die wir nachfolgend näher eingehen wollen.

### 1.3.1 Forderungen an empirische Daten

Die Statistik, d. h., die in ihr genutzten mathematischen Modelle, benötigt im weitesten Sinne quantifizierbare Daten. Es besteht die Notwendigkeit, daß sich die empirischen Daten und Beobachtungen numerisch, d. h. durch Zahlen, kennzeichnen lassen. Häufig treffen wir aber, speziell im Einzelfall (z. B. der Persönlichkeitspsychologie), auf ausschließlich qualitative Angaben. In solchen Situationen, wie etwa beispielsweise bei einem qualitativen Merkmal wie „soziale Herkunft", benötigen wir wenigstens als quantifizierbare Daten die beobachteten Häufigkeiten der einzelnen „Stufen" des Merkmals.

Eine weitere, notwendige Voraussetzung besteht in einer exakten Definition dessen, was gemessen werden soll. In den sogenannten „exakten Wissenschaften" wie Physik oder Chemie ist das sicher leichter als in den Human- und Sozialwissenschaften. Betrachten wir als Beispiel das Merkmal „soziale Herkunft". Hier sind verschiedene Kategoriensysteme wie etwa Arbeiter, Angestellter, Bauer, Angehöriger der Intelligenz, Unternehmer usw. vorstellbar. Will man aber beispielsweise wissenschaftliche Aussagen aus dem Vergleich verschiedener Untersuchungen ableiten, dann hat das nur einen Sinn, wenn die Inhalte der untersuchten Kategorien vergleichbar, also genau definiert sind. Welche Berufe zählen wir nun zum Beispiel zur Kategorie „Arbeiter"? Der Anwender der Statistik sollte also auch Definitionen hinterfragen.

Schließlich sollen die empirischen Daten durch geeignete Meßverfahren erhoben worden sein. Diese Forderung ist nicht neu und wohl in allen Wissenschaften gleich. Dennoch können wir nicht a priori davon ausgehen, daß sie immer Berücksichtigung findet. Mißt man beispielsweise bei einem Weitsprungwettbewerb die erreichten Weiten mit Hilfe eines Maßbandes aus Gummi, dann würde man offensichtlich an der Seriosität der Ergebnisse zweifeln. Aus diesem Grunde sollten wir, die Anwender, stets darauf achten, daß wenigstens nachfolgende 3 Forderungen an die verwendeten Meßverfahren erfüllt sind:

1. Das Kriterium der Objektivität:
   Damit ist gemeint, daß das jeweilige Meßverfahren unabhängig vom Anwender sein soll. Auch wenn ein anderer Versuchsleiter oder Befrager die Analyse durchführt, sollte dies ohne Einfluß bleiben. Der interessierte Leser findet in der psychologischen Fachliteratur dazu viele Hinweise unter dem Schlagwort „Versuchsleiterfehler".
2. Das Kriterium Reliabilität (auch Zuverlässigkeit):
   Hierbei geht es darum, daß das Meßverfahren reproduzierbar, d. h. wiederholbar sein muß. Jetzt wird der eine oder andere Leser einwenden, daß z. B. gerade sogenannte projektive Tests in der Psychologie diesem Anspruch nicht uneingeschränkt genügen. Das stimmt – aber der qualitative Einzelfall kann ohnehin nicht statistisch ausgewertet werden.
3. Das Kriterium Validität (auch Gültigkeit):
   Die eingesetzten Meßverfahren müssen wirklich das messen, was sie messen sollen und vorgeben zu messen. Es ist eben beispielsweise unsinnig, mit Hilfe der Lösung eines Kreuzworträtsels die Intelligenz eines Probanden messen zu wollen. Auch wenn dieses Beispiel sicher etwas drastisch ist, so verdeutlicht es doch das mitunter zu unkritische Umgehen mit dem Problem der Validität.

*Bemerkung*: *Die Anwendung statistischer Verfahren setzt voraus, daß diese Forderungen erfüllt sind – wobei unterstellt wird, daß es ein wissenschaftstheoretisch weitestgehend ungeklärtes Validitäts-Reliabilitäts-Dilemma gibt.*

### 1.3.2 Vorteile und Grenzen beim Einsatz der Statistik

Wie bereits im Abschnitt 1.2 festgestellt wurde, stellt die Statistik ein methodisches Instrumentarium zur wissenschaftlichen Analyse empirischer Daten bereit. Wir wollen in diesem Abschnitt zusammenfassend sowohl einige wesentliche Vorteile für ihren Einsatz benennen als auch auf Einschränkungen kritisch aufmerksam machen.

1. Vorteile:
   - Die deskriptive Statistik bietet uns die Möglichkeit der Präzision, d. h. der genauen Beschreibung der Beobachtungen und deren Zusammenfassung. Damit können wir unüberschaubare Mengen detaillierter Daten anschaulich darstellen und weiterverarbeiten.

- Durch den Einsatz der schließenden Statistik können wir die auf der Grundlage von Stichproben gewonnenen Aussagen auf die gesamte zugehörige Grundgesamtheit verallgemeinern.
- Wir sind auf Grund der verwendeten mathematischen Modelle in der Lage, selbst Aussagen zur Genauigkeit und zur Sicherheit der festgestellten Schlußfolgerungen zu treffen.
- Es besteht die Möglichkeit, Aussagen sowohl theoretisch als auch empirisch zu überprüfen.
- Schließlich können wir durch den Gebrauch mathematischer Methoden auf der Grundlage der empirisch gewonnenen Daten weitere Modellrechnungen durchführen, d. h., wir können ohne ökonomische, soziale oder andere Konsequenzen weitere Fallbeispiele exemplarisch durchspielen.

2. Kritisch zu beachten sind:
   - Wie immer man die Statistik auch einsetzt, sie liefert niemals Aussagen zur **inhaltlichen Bedeutsamkeit** der durchgeführten Untersuchung. Die Verantwortung dafür liegt ausschließlich beim Anwender.
   - Die Statistik liefert für den Untersuchungsansatz und die -durchführung keine Kriterien darüber, **welche Beobachtungsgrößen** zu verwenden sind. Also selbst wenn das Kreuzworträtsel zur „Messung der Intelligenz“ mißbraucht wird, kann die Statistik darauf keinen Einfluß nehmen.
   - Auch die Frage, welches Meßverfahren im Einzelfall zu verwenden ist, kann mit den Mitteln der Statistik nicht beantwortet werden.
   - Die Statistik kann eine inhaltliche Interpretation nicht leisten. Beispielsweise werden zwar Zusammenhänge und Abhängigkeiten mathematisch ermittelt, aber die Einschätzung ihrer Bedeutung, etwa die Unterscheidung zwischen Ursache und Wirkung, obliegt dem Benutzer.
   - Jede Anwendung von statistischen Verfahren ist abhängig von bestimmten Voraussetzungen. Für deren Beachtung ist der Anwender verantwortlich, d. h., wenn man bei vorliegendem empirischen Datenmaterial den falschen Parameter berechnet oder den falschen Test einsetzt, dann kann sich die Statistik nicht dagegen „wehren“.

# 3 Wahrscheinlichkeitstheorie

In Kapitel 2 hatten wir bereits anhand einer Reihe von Beispielen Methoden und Verfahren zur Auswertung von Versuchsergebnissen kennengelernt. Man verfolgt dabei das Ziel, einerseits einen besseren Überblick über das Datenmaterial zu erhalten und andererseits möglicherweise zugrundliegende allgemeine Gesetzmäßigkeiten zu erkennen und zu studieren. Es handelt sich hier größtenteils um Situationen, in denen das Versuchsgeschehen aufgrund der Wirkung des Zufalls wesentlich beeinflußt wird. Während man zum Beispiel mit hinreichender Genauigkeit unter Verwendung der Gesetze der klassischen Mechanik eine Aufprallgeschwindigkeit beim freien Fall aus gegebener Höhe im voraus berechnen kann, ist es kaum vorstellbar, strenggenommen sogar prinzipiell unmöglich, durch Berechnung das Ergebnis der nächsten Ziehung im Lottospiel vorauszusagen. Hier hat der Zufall einen zu großen Einfluß. Trotzdem ist es auch dort noch möglich, in anderer Weise allgemeingültige Gesetzmäßigkeiten der Wirkung des Zufalls zu erkennen und zu beschreiben. Die Wahrscheinlichkeitsrechnung beschäftigt sich genau mit dieser Aufgabe, wobei man aber prinzipiell beachten muß, daß dies durch Modellbildung erfolgt, und die Interpretation und das Anpassen an eine konkret zu untersuchende Situation weitere Überlegungen erfordern. Man setzt nun im Modell voraus, daß das Auftreten der einzelnen möglichen Versuchsergebnisse in einer betrachteten Situation durch zwischen 0 und 1 gelegene Zahlen, sog. Wahrscheinlichkeiten, bewertet wird und stellt sich methodisch auf den Standpunkt, daß diese Zahlen als Grundinformation bekannt und gegeben sind (sog. axiomatische Methode). Diese Wahrscheinlichkeiten sollen die Chance, den Grad der Bestimmtheit des Auftretens der einzelnen Versuchsergebnisse charakterisieren.

Dabei sind einige Aspekte bei der Interpretation bedenkenswert.

Da ist zum ersten der „logisch-plausible Aspekt“: In speziellen Situationen, z. B. häufig bei Glücksspielen, ist man aufgrund von Symmetrievorstellungen geneigt, jedem Versuchsergebnis die gleiche Wahrscheinlichkeit zuzusprechen. So würde man z. B. bei einem Wurf mit einem idealen Würfel jeder Augenzahl die Chance $\frac{1}{6}$ geben. Oder man hielte es für völlig plausibel, wenn z. B. in einem Leistungstest von Fragen mit Alternativantworten, wobei jeweils genau eine der beiden Antworten richtig ist, für eine Versuchsperson, die völlig ahnungslos ist und willkürlich rät, bei jeder Frage mit 50 %iger Chance die richtige Antwort getroffen wird.

Ein zweiter, und für uns wesentlicher Aspekt ist der „Aspekt der Massenerscheinung“. Wir stellen uns auf den Standpunkt, daß die Gesetzmäßigkeiten des Zufalls bei wachsender Anzahl von Wiederholungen des Versuchs unter Einhaltung konstanter äußerer Versuchsbedingungen erkennbar werden und die Wahrscheinlichkeiten „theoretische Idealwerte“ für die relativen Häufigkeiten des Auftretens der einzelnen Versuchsergebnisse sind. So könnte man sich z. B. vorstellen, daß bei einem gezinkten Würfel sich in langen Versuchsserien die relativen Häufigkeiten für das Auftreten einer „6“ auf einen bestimmten festen Wert, z. B. 0,18, einpegeln. Man akzeptiert hier also die Tatsache, daß auch der gezinkte Würfel im Häufigkeitsverhalten des Auftretens der einzelnen Augenzahlen bei großen Versuchsserien bestimmte Gesetzmäßigkeiten aufweist. Man kann diese „statistischen Wahrscheinlichkeiten“ nun als empirisch gewonnene Erfahrungswerte für das Würfeln mit dem gegebenen gezinkten Würfel auffassen. Dieser empirische Aspekt interpretiert also Wahrscheinlichkeiten als Grenzwerte relativer Häufigkeiten, er ist für die weiteren Betrachtungen eine wesentliche Grundlage.

Ein dritter Aspekt ist der subjektive Aspekt. Hier wird die Wahrscheinlichkeit als Maßzahl für das subjektive „Fürwahrhalten“ des Eintretens des betrachteten Versuchsergebnisses durch den Anwender aufgefaßt. Nimmt man als Beispiele die Reaktion auf Reize bei Versuchspersonen (z. B. Sensibilitätsprüfungen) oder die Einschätzung der Wirksamkeit eines Medikamentes, so stellt man auch hier fest, daß bei der subjektiven Bewertung empirische Erfahrungen und auch „innere Gesetzmäßigkeiten" einen wesentlichen Einfluß haben können.

Aufgrund dieser verschiedenen Interpretationen ist bei Anwendungen in konkreten Situationen kritischer Sachverstand vonnöten.

Wir wollen uns hier vorrangig mit einem **theoretischen Modell zur Beschreibung von Zufallserscheinungen** befassen und in diesem Modell versuchen zu verstehen, wie man mit Wahrscheinlichkeiten umgeht und rechnet. Dieses Modell ist relativ übersichtlich und einfach und aus diesem Grunde wohl auch die Basis für eine Reihe angewandter Disziplinen, wie z. B. für die Statistik, die Informationstheorie, die Biometrie, die statistische Qualitätskontrolle, statistische Versuchsplanung, Bedienungstheorie, Zuverlässigkeitstheorie, Monte-Carlo-Simulation, u. a. Man setzt also voraus, daß im Grundmodell die Wahrscheinlichkeiten für die einzelnen Versuchsausgänge bekannt sind, und bestimmt dann auf dieser Basis kompliziertere Wahrscheinlichkeiten, die von praktischem Interesse sind. Als symptomatisches Beispiel möge hier noch einmal der bereits weiter oben zitierte Leistungstest mit 10 angenommenen Alternativfragen dienen. Wir halten es für völlig einleuchtend, daß bei „willkürlichem" Ausfüllen des Fragebogens je Frage die Chance 50 % besteht, die richtige Antwort zu treffen. Wir können aber nicht intuitiv unmittelbar erraten, wie groß die Wahrscheinlichkeit dafür ist, daß ein auf diese Weise ausgefüllter Fragebogen z. B. genau 3 richtige Antworten enthält. Eine kombinatorische Rechnung ergibt den Wert $\binom{10}{3} \cdot 0{,}5^{10} \approx 11{,}7\ \%$.

## 3.1 Das wahrscheinlichkeitstheoretische Grundmodell

### 3.1.1 Stichprobenraum, zufällige Ereignisse

Es gibt verständlicherweise viele, zum Teil recht unterschiedliche Situationen, in denen aufgrund der Wirkung des Zufalls nichtvorhersagbare Beobachtungen und Ereignisse auftreten. Wir wollen in diesem Abschnitt damit beginnen, erforderliche Bestimmungsstücke für den Aufbau eines Modells zur Beschreibung von Zufallserscheinungen zu präzisieren. Es ist ein Charakteristikum theoretischer Betrachtungen, daß man in der zu untersuchenden Situation das Spektrum der **prinzipiell denkbaren Beobachtungsergebnisse** beachtet. Die praktische Versuchsdurchführung, konkrete Beobachtung und Auswertung ist das eine, das theoretische Bedenken und Einschätzen aller möglichen Eventualitäten das andere. Als Synonym für die zu untersuchende, vom Zufall beeinflußte Situation gebrauchen wir den Begriff zufälliger Versuch. Wir fordern dabei zunächst nur, daß die Menge [1)] der prinzipiell möglichen konkreten Beobachtungsergebnisse von vornherein angegeben werden kann.

Die Menge der möglichen Versuchsergebnisse heiße **Merkmalsraum** oder auch **Stichprobenraum** bzw. **Ereignisraum** und werde mit $\Omega$ bezeichnet. Die Elemente $\omega$ von $\Omega$ nennen wir elementare Versuchsergebnisse.

Jede Beobachtung bzw. Versuchsdurchführung liefert dann ein elementares Versuchsergebnis. Es ist aber aufgrund der Wirkung des Zufalls nicht möglich, vor der Versuchsdurchführung vorherzusagen, welches der möglichen $\omega$ tatsächlich beobachtet werden wird. Man kann die Versuchsdurchführung

[1)] Eine zusammenfassende Darstellung der wichtigsten Grundbegriffe der Mengenlehre findet man im Kapitel 6 (Abschnitt 6.1).

als zufällige Entnahme eines Beobachtungswertes $\omega$ aus der Menge $\Omega$ aller in Frage kommenden Versuchsergebnisse interpretieren.

**Beispiel 3.1/1**: Werfen einer Münze; hier gibt es 2 mögliche Versuchsausgänge, die wir mit Z für „Zahl oben" und W für „Wappen oben" bezeichnen wollen. Also ist $\Omega = \{Z,W\}$.

**Beispiel 3.1/2**: Werfen eines Würfels; wir erhalten offensichtlich $\Omega = \{1,2,3,4,5,6\}$.

**Beispiel 3.1/3**: Werfen von 3 verschiedenen Münzen; hier gibt es pro Münze 2 Möglichkeiten, also insgesamt $2 \cdot 2 \cdot 2 = 8$ Versuchsegebnisse $\omega$, die wir z. B. wie folgt bezeichnen könnten: $\omega = (Z,W,Z)$ heißt dann, daß die 1. und 3. Münze „Zahl oben" zeigt und die 2. Münze „Wappen oben". Wir erhalten $\Omega = \{(Z,Z,Z),(Z,Z,W),(Z,W,Z),(W,Z,Z),(Z,W,W),(W,Z,W),(W,W,Z),(W,W,W)\}$.

**Beispiel 3.1/4**: Werfen zweier (verschiedenfarbiger) Würfel; analog zu Bsp. 3.1/3 erhalten wir $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(2,6),\ldots,(6,6)\}$. Es gibt hier $6 \cdot 6 = 36$ mögliche Versuchsergebnisse $\omega$.

**Beispiel 3.1/5**: Durchführung eines Intelligenztests; das Spektrum der möglichen Punktsumme bewege sich von 50 bis 150, also hat man $\Omega = \{50,51,52,\ldots,150\}$.

**Beispiel 3.1/6**: Ausfüllen eines Fragebogens durch eine Versuchsperson. Wir nehmen der Einfachheit halber an, daß es sich um 5 Fragen zu je 3 Antworten $a,b,c$ handelt. Es gibt somit insgesamt $3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 243$ verschiedene Versuchsergebnisse:

| | $a$ | $b$ | $c$ |
|---|---|---|---|
| 1. | $\times$ | | |
| 2. | $\times$ | | |
| 3. | $\times$ | | |
| 4. | $\times$ | | |
| 5. | $\times$ | | |

| | $a$ | $b$ | $c$ |
|---|---|---|---|
| 1. | $\times$ | | |
| 2. | $\times$ | | |
| 3. | $\times$ | | |
| 4. | $\times$ | | |
| 5. | | $\times$ | |

| | $a$ | $b$ | $c$ |
|---|---|---|---|
| 1. | $\times$ | | |
| 2. | $\times$ | | |
| 3. | $\times$ | | |
| 4. | | $\times$ | |
| 5. | $\times$ | | |

| | $a$ | $b$ | $c$ |
|---|---|---|---|
| 1. | $\times$ | | |
| 2. | $\times$ | | |
| 3. | | $\times$ | |
| 4. | $\times$ | | |
| 5. | $\times$ | | |

...

| | $a$ | $b$ | $c$ |
|---|---|---|---|
| 1. | | | $\times$ |
| 2. | | | $\times$ |
| 3. | | | $\times$ |
| 4. | | | $\times$ |
| 5. | | | $\times$ |

Die Grundmenge $\Omega$ besteht hier aus allen Antwortkonstellationen $\omega = (x_1,x_2,x_3,x_4,x_5)$, wobei jedes $x_i$ die Werte $a,b,c$ annehmen kann.

**Beispiel 3.1/7**: Zufällige Auswahl einer Versuchsperson, die einen Punkt auf einer Rating-Skale markieren soll. Solche Skalen haben gewöhnlich eine Länge von 7 cm bis 10 cm. Wir verwenden eine 10 cm-Rating-Skale und erhalten dann den Stichprobenraum $\Omega$ aller reellen Zahlen $x$ zwischen 0 und 10: $\Omega = \{x\colon 0 \leq x \leq 10\}$.

**Beispiel 3.1/8**: Auswahl einer Versuchsperson und die Messung von deren Körpergröße und Gewicht. Hier besteht $\Omega$ aus allen Meßwertpaaren $\omega = (x,y)$ reeller positiver Zahlen $x,y$: $\Omega = \{(x,y)\colon x > 0, y > 0\}$.

In vielen Situationen ist es möglich, den zufälligen Versuch real oder wenigstens gedanklich beliebig oft unter Einhaltung der Versuchsbedingungen zu wiederholen. Wir werden diese Eigenschaft nicht streng einfordern, orientieren uns aber im weiteren Modellaufbau an der Vorstellung der Wiederholbarkeit zufälliger Versuche. Der nächste Grundbegriff ist der Begriff des zufälligen Ereignisses.

Ein **zufälliges Ereignis** ist eine Teilmenge des Stichprobenraumes $\Omega$. Wir verwenden dementsprechend für zufällige Ereignisse die gleichen Bezeichnungen wie für Mengen, also $A, B, \ldots$ Wir sagen, das zufällige **Ereignis** $A$ **tritt ein**, wenn das Versuchsergebnis in $A$ liegt.

Aufgrund der Wirkung des Zufalls ist es vor der Versuchsdurchführung nicht möglich vorherzusagen, ob ein interessierendes zufälliges Ereignis $A$ eintreten wird oder nicht. Wir betrachten einige Beispiele und knüpfen an die oben erwähnten an:

**Zu Beispiel 3.1/2:** Das zufällige Ereignis $A = \{2,4,6\}$ bedeutet, daß beim Würfeln mit einem Würfel eine gerade Zahl gewürfelt wird.

**Zu Beispiel 3.1/3:** $A = \{(Z,W,W),(W,Z,W),(W,W,Z)\}$ heißt, daß bei dem Münzwurf mit 3 Münzen genau 1mal eine Zahl vorkommt.

**Zu Beispiel 3.1/4:** Das Ereignis $A$, bei einem Wurf mit 2 Würfeln die Würfelsumme 5 zu erhalten, können wir wie folgt angeben: $A = \{(1{,}4),(2{,}3),(3{,}2),(4{,}1)\}$.

**Zu Beispiel 3.1/5:** $A = \{50,51,52,\ldots,90\}$ bedeutet, daß die Versuchsperson beim Intelligenztest nicht mehr als 90 Punkte erreicht, während z. B. $B = \{92\}$ heißt, daß sie genau auf 92 Punkte kommt.

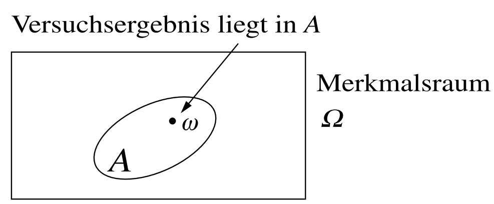
**Zu Beispiel 3.1/6:** $A = \{(x_1,x_2,x_3,x_4,x_5)$ mit $x_i = a$ oder $c$, aber nie $b$ für alle$i\}$ ist das zufällige Ereignis: Die Antwort $b$ wurde kein einziges Mal angekreuzt. Es gibt $2\cdot2\cdot2\cdot2\cdot2 = 32$ verschiedene solche Möglichkeiten, d. h. $A$ enthält 32 Elemente. Wird also z. B. der Fragebogen $\omega = (a,a,c,a,c)$ erhalten, so ist $A$ eingetreten.

**Zu Beispiel 3.1/8:** $A = \{(x,y)\colon x \geq 1{,}70,\ y \leq 68{,}5\}$ bedeutet, daß die ausgewählte Versuchsperson mindestens 1,70 m groß ist und dabei höchstens 68,5 kg wiegt. Auch hier erkennen wir wieder, daß aufgrund der zufälligen Auswahl der Versuchsperson das Ereignis eintreten kann oder nicht.

Zufällige Ereignisse sind also Teilmengen des Merkmalsraumes $\Omega$. Extreme Teilmengen sind die leere Menge $\emptyset$ und die ganze Menge $\Omega$. Die leere Menge $\emptyset$ heißt in diesem Zusammenhang ein **unmögliches Ereignis**. Dieses tritt nie ein, da es kein Element enthält. Auf der entgegengetzten Seite heißt die ganze Menge $\Omega$ ein **sicheres Ereignis**. Es tritt stets bei jedem Versuch ein, da ein beliebiges mögliches Versuchsergebnis zu $\Omega$ gehört.

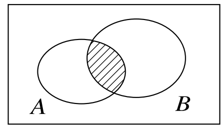
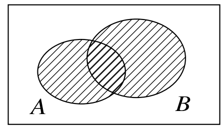
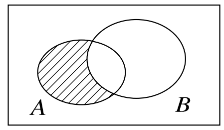
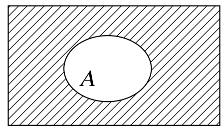
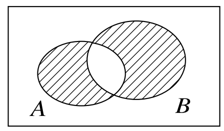
Ist der Merkmalsraum $\Omega$ endlich oder hat er höchstens abzählbar unendlich viele Elemente, so ist **jede** Teilmenge $A$ von $\Omega$ ein zufälliges Ereignis. Enthält $\Omega$ noch mehr Elemente, d. h. ist $\Omega$ überabzählbar, dann muß man aus mathematisch-theoretischen Gründen etwas vorsichtiger sein und kann nur gewisse Teilmengen $A$ von $\Omega$ als zufällige Ereignisse zulassen (man vgl. auch die Bemerkungen am Ende dieses Abschnittes).

Aus den bisherigen Ausführungen erkennen wir, daß beim Aufbau des Modells der Wahrscheinlichkeitsrechnung die Mengenlehre (vgl. auch Abschnitt 6.1) benutzt wird. Man hat hier nur eine andere Sprechweise zu berücksichtigen: Wir sagen nicht, das Element $\omega$, d. h. das Versuchsergenis $\omega$, liegt in der Menge $A$, sondern wir sagen stattdessen, das Ereignis $A$ tritt ein.




Das zufällige Ereignis $A$ tritt ein: Versuchsergebnis liegt in $A$

Man überträgt nun die Operationen zwischen Mengen auf diesen Sachverhalt. Wir betrachten zwei Ereignisse $A$,$B$ und bilden aus ihnen durch Mengenoperationen neue Ereignisse:

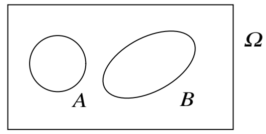
| Bezeichnung | Sprechweise | Interpretation | Symbolische Darstellung |
|---|---|---|---|
| $A \cap B$ | $A$ und $B$ | $A$ und $B$ treten gleichzeitig ein. |  |
| $A \cup B$ | $A$ oder $B$ | Es tritt $A$ oder $B$ ein. |  |
| $A \setminus B$ | $A$ minus $B$ | Es tritt $A$, aber nicht $B$ ein. |  |
| $\overline{A} = \Omega \setminus A$ | nicht $A$ | Es tritt das Gegenteil von $A$ ein. |  |
| $A \triangle B$ | $A$ Diskrepanz $B$ | Es tritt entweder $A$ oder $B$ ein, aber nicht beide zugleich. |  |

Wir betrachten ein Beispiel und verwenden dabei auch gleich eine tabellarische Darstellung. Wir nehmen Beispiel 3.1/4 und betrachten die folgenden beiden Ereignisse:
$A$: Die Augenzahlen beider Würfel sind gleich, d. h., es wird ein Pasch gewürfelt.
$B$: Die Augensumme beträgt 6.

| Operation zwischen $A$ und $B$ | Interpretation | Angabe der zugehörigen Versuchsergebnisse | Anzahl |
|---|---|---|---|
| $A \cap B$ | Man erhält Pasch und gleichzeitig Augensumme 6 | (3,3) | 1 |
| $A \cup B$ | Pasch oder Augensumme 6 | (1,1), (2,2), (3,3), (4,4), (5,5), (6,6), (1,5), (2,4), (4,2), (5,1) | 10 |
| $A \setminus B$ | Pasch, aber nicht Augensumme 6 | (1,1), (2,2), (4,4), (5,5), (6,6) | 5 |
| $\overline{A}$ | kein Pasch | 30 Versuchsergebnisse | |
| $A \triangle B$ | Entweder Pasch oder Augensumme 6 | (1,1), (2,2), (4,4), (5,5), (6,6), (1,5), (2,4), (4,2), (5,1) | 9 |

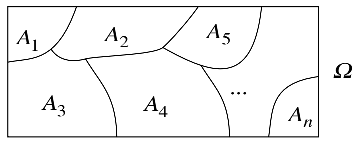
Für das weitere benötigen wir die folgenden beiden Begriffe:

- Man sagt, die zufälligen Ereignisse $A$ und $B$ sind **unvereinbar**, wenn $A \cap B = \emptyset$ gilt. Die Mengen $A$ und $B$ liegen in diesem Fall also „auseinander“:




Unvereinbare Ereignisse können prinzipiell nie zugleich eintreten, weil sie sich „widersprechen". So ist z. B. beim Würfeln mit 2 Würfeln das Ereignis $A$: „Die Augensumme ist ungerade." mit dem Ereignis $B$: „Die Augenzahlen beider Würfel sind gerade." unvereinbar.

- Man sagt, die Ereignisse $A_1, A_2, \ldots, A_n$ bilden ein **vollständiges System von zufälligen Ereignissen**, wenn durch sie eine Zerlegung der Grundmenge $\Omega$ gegeben ist, d. h., wenn sie zusammengelegt ganz $\Omega$ ergeben und sich dabei nicht überschneiden. Formal heißt das: $\Omega = A_1 \cup A_2 \cup \ldots \cup A_n$; je zwei der Ereignisse $A_i$ sind unvereinbar:




Ein vollständiges System von Ereignissen erhält man z. B. immer bei einer „Fallunterscheidung" der möglichen Versuchsergebnisse. So bilden z. B. beim Würfeln mit 2 verschiedenfarbigen Würfeln die folgenden 4 Ereignisse ein vollständiges System:
$A_1$**:** „Beide Augenzahlen sind gerade."
$A_2$**:** „Beide Augenzahlen sind ungerade."
$A_3$**:** „Augenzahl von Würfel 1 ist gerade und Augenzahl von Würfel 2 ist ungerade."
$A_4$**:** „Augenzahl von Würfel 1 ist ungerade und Augenzahl von Würfel 2 ist gerade."

Beim Aufbau eines theoretischen Modells für einen zufälligen Versuch hat man also zuerst die Grundmenge $\Omega$ der in Frage kommenden elementaren Versuchsausgänge anzugeben. Als nächstes werden zufällige Ereignisse betrachtet, sie sind Teilmengen der Grundmenge $\Omega$. Man faßt nun auch die zufälligen Ereignisse wieder zu einer Menge zusammen und nennt diese Menge das **zugehörige Ereignisfeld**. Wir wollen es standardmäßig mit $\mathfrak{A}$ bezeichnen. Ist der Grundraum $\Omega$ endlich oder abzählbar unendlich [1], so enthält das zugehörige Ereignisfeld $\mathfrak{A}$ **alle** Teilmengen von $\Omega$; $\mathfrak{A}$ ist also die Potenzmenge von $\Omega$ (vgl. Abschnitt 6.1). Enthält $\Omega$ mehr als abzählbar unendlich viele Versuchsausgänge, wie z. B. bei „stetigen Messungen", dann muß man etwas vorsichtiger vorgehen: Man betrachtet zunächst ein dem Problem angepaßtes System interessierender Teilmengen $A$ und nimmt dann noch die Ereignisse hinzu, die man aus ihnen durch (u. U. auch unendlich häufiges) Anwenden der Mengenoperationen $\cap, \cup, \setminus, \overline{\phantom{x}}$ erhalten kann. Die so theoretisch gebildete Kollektion von Teilmengen von $\Omega$ ist dann das zugehörige Ereignisfeld $\mathfrak{A}$. Betrachten wir z. B. eine zufällig ausgewählte Glühlampe und messen ihre Brenndauer beliebig genau, so ist jede reelle Zahl $x \geq 0$ als Versuchsergebnis möglich. Wir erhalten als Grundraum $\Omega$ das Intervall $[0,\infty)$, also eine überabzählbar [2] unendliche Menge. Nun könnten uns z. B. alle Ereignisse der Form $A = \{x\colon x \geq a\}$ interessieren, d. h., die Brenndauer beträgt mindestens $a$, wobei $a > 0$ beliebig gewählt wird. Man nimmt nun zu all diesen Ereignissen $A$ noch die durch – auch mehrfaches – Anwenden der Mengenoperationen $\cap, \cup, \setminus, \overline{\phantom{x}}$ entstehenden Teilmengen von $\Omega$ hinzu. Es sei bemerkt, daß das so entstehende System von Teilmengen, das wir zugehöriges Ereignisfeld $\mathfrak{A}$ nennen, nicht sämtliche Teilmengen von $\Omega$ enthält. In jedem Fall ist aber ein Ereignisfeld $\mathfrak{A}$ eine Kollektion von Teilmengen von $\Omega$, in der die Anwendung der Mengenoperationen (höchstens abzählbar unendlich oft) uneingeschränkt möglich ist, d. h. immer wieder zu Elementen der Kollektion $\mathfrak{A}$ führt.

[1] Eine abzählbar unendliche Menge $\Omega$ ist eine solche Menge, die nur so viele Elemente enthält, daß man sie noch durchnumerieren kann, z. B. $\Omega = \{1, 2, 3, 4, \ldots\}$, $\Omega = \{2, 4, 6, 8, \ldots\}$, $\Omega = \{1, 10, 100, 1000, 10^4, \ldots\}$. Es gibt aber auch Mengen, die „wesentlich umfangreicher" sind, so daß ein Durchnumerieren nicht mehr gelingt. Ein Beispiel hierfür ist die Menge aller rellen Zahlen eines Intervalls, wie z. B. $\Omega = \{x \text{ reell}\colon 0 \leq x \leq 10\}$. Solche Mengen nennt man überabzählbar.

### 3.1.2 Relative Häufigkeiten

Zum weiteren Modellaufbau verfolgen wir das Ziel, für jedes zufällige Ereignis $A$ aus dem zugehörigen Ereignisfeld seine Chance des Eintretens durch eine zwischen 0 und 1 gelegene Zahl zu bewerten. Dabei müssen bestimmte Rechenregeln eingehalten werden. Um diese zu finden, orientieren wir uns an der Vorstellung, daß die Gesetze des Zufalls durch Versuchswiederholungen in großen Serien erkennbar werden.
Wir betrachten also einen zufälligen Versuch, bestimmen die zugehörige Grundmenge $\Omega$ und fixieren ein uns interessierendes Ereignis $A$ aus dem Ereignisfeld $\mathfrak{A}$. Nun wird der Versuch wiederholt, also $n$-mal durchgeführt, und dabei wird jedesmal lediglich festgestellt, ob $A$ eintritt oder nicht. Wir bezeichnen mit

$H_n(A)$ die **absolute Häufigkeit** des Eintretens von $A$ in den ersten $n$ Versuchen, d. h. die Gesamtzahl derjenigen Male unter diesen $n$ Versuchen, in denen $A$ eingetreten ist.

Die **relative Häufigkeit** werde mit $h_n(A)$ bezeichnet und entsteht aus der absoluten Häufigkeit mittels Division durch $n$:

$$\boxed{h_n(A) = \frac{H_n(A)}{n}} \qquad \text{(relative Häufigkeit).}$$

**Beispiel**:

| Versuchsnr. | Eintreten von $A$ (markiert mit *) | $H_n(A)$ | $h_n(A)$ |
|---|---|---|---|
| 1 | | 0 | 0 |
| 2 | * | 1 | $\frac{1}{2}$ |
| 3 | * | 2 | $\frac{2}{3}$ |
| 4 | | 2 | $\frac{2}{4}$ |
| 5 | | 2 | $\frac{2}{5}$ |
| 6 | | 2 | $\frac{2}{6}$ |
| 7 | * | 3 | $\frac{3}{7}$ |
| 8 | * | 4 | $\frac{4}{8}$ |
| 9 | | 4 | $\frac{4}{9}$ |
| 10 | * | 5 | $\frac{5}{10}$ |

Man wertet also jeweils die ersten $n$ Versuche. Die Zahlen $H_n(A)$ nehmen zu oder bleiben konstant, während sich die Werte $h_n(A)$ ständig ändern, da man stets durch ein anderes $n$ dividiert.

Wir wollen nun einige Eigenschaften der relativen Häufigkeit diskutieren und beachten dabei getrennt zwei Gesichtspunkte:
1. Wir halten die Anzahl $n$ der Versuche fest und untersuchen die Eigenschaften von $h_n(A)$, wenn über $A$ variiert wird.
2. Wir halten das Ereignis $A$ fest und lassen $n$ wachsen.

Zu 1.: Wir beginnen mit einem Beispiel. Es werde mit 2 verschiedenfarbigen Würfeln gewürfelt, und wir betrachten folgende Ereignisse:
$A_1$ „Die Augenzahlen beider Würfel sind gerade."
$A_2$ „Die Augenzahlen beider Würfel sind ungerade."
$B$ „Die Augensumme beträgt 6, 7, 8 oder 9.",

sowie die abgeleiteten Ereignisse $A_1 \cup A_2, A_1 \cup B, A_1 \cap B, \overline{A_1}$. Wir bemerken, daß $A_1$ und $A_2$ unvereinbar sind.

Versuchsbeispiel:

| Nr. $n$ | Ergebnis | Würfel-summe | $h_n(A_1)$ | $h_n(A_2)$ | $h_n(A_1 \cup A_2)$ | $h_n(\overline{A}_1)$ | $h_n(B)$ | $h_n(A_1 \cup B)$ | $h_n(A_1 \cap B)$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | jeweils für $n = 5,10,15,20$ | | | | | | |
| 1 | (4, 1) | 5 | | | | * | | | |
| 2 | (3, 3) | 6 | | * | * | * | * | * | |
| 3 | (6, 2) | 8 | * | | * | | * | * | * |
| 4 | (1, 5) | 6 | | * | * | * | * | * | |
| 5 | (4, 4) | 8 | * $\frac{2}{5}$ | $\frac{2}{5}$ | * $\frac{4}{5}$ | $\frac{3}{5}$ | * $\frac{4}{5}$ | * $\frac{4}{5}$ | * $\frac{2}{5}$ |
| 6 | (2, 5) | 7 | | | | * | * | * | |
| 7 | (1, 3) | 4 | | * | * | * | | | |
| 8 | (2, 4) | 6 | * | | * | | * | * | * |
| 9 | (6, 2) | 8 | * | | * | | * | * | * |
| 10 | (3 ,1) | 4 | $\frac{4}{10}$ | * $\frac{4}{10}$ | * $\frac{8}{10}$ | * $\frac{6}{10}$ | $\frac{7}{10}$ | $\frac{7}{10}$ | $\frac{4}{10}$ |
| 11 | (5, 4) | 9 | | | | * | * | * | |
| 12 | (3, 6) | 9 | | | | * | * | * | |
| 13 | (1, 6) | 7 | | | | * | * | * | |
| 14 | (2, 2) | 4 | * | | * | | | * | |
| 15 | (4, 1) | 5 | $\frac{5}{15}$ | $\frac{4}{15}$ | $\frac{9}{15}$ | * $\frac{10}{15}$ | $\frac{10}{15}$ | $\frac{11}{15}$ | $\frac{4}{15}$ |
| 16 | (5, 6) | 11 | | | | * | | | |
| 17 | (2, 5) | 7 | | | | * | * | * | |
| 18 | (1, 1) | 2 | | * | * | * | | | |
| 19 | (4, 2) | 6 | * | | * | | * | * | * |
| 20 | (3, 5) | 8 | $\frac{6}{20}$ | * $\frac{6}{20}$ | * $\frac{12}{20}$ | * $\frac{14}{20}$ | * $\frac{13}{20}$ | * $\frac{14}{20}$ | $\frac{5}{20}$ |

Anhand des Beispiels bestätigen wir folgende allgemeine Eigenschaften der relativen Häufigkeiten:

1. Für jedes Ereignis $A$ aus dem Ereignisfeld $\mathfrak{A}$ kann man die relative Häufigkeit $h_n(A)$ bestimmen (einfach durch Abzählen).

   $$\boxed{\text{Es gilt stets} \qquad 0 \leq h_n(A) \leq 1},$$

   genauer, $h_n(A)$ kann einen der Werte $0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n} = 1$ annehmen. Welcher dieser möglichen Werte tatsächlich auftritt, hängt von der konkreten Versuchsserie und damit vom Zufall ab.
2. Für die extremen Ereignisse $\emptyset, \Omega$ erhalten wir offensichtlich

   $$\boxed{h_n(\emptyset) = \frac{0}{n} = 0, \qquad h_n(\Omega) = \frac{n}{n} = 1}.$$

3. Die relativen Häufigkeiten sind **additiv,** d. h., es gilt

   $$\boxed{h_n(A_1 \cup A_2) = h_n(A_1) + h_n(A_2) \ \text{ für beliebige Ereignisse } A_1, A_2 \text{ mit} A_1 \cap A_2 = \emptyset}.$$

Diese Eigenschaft der Additivität ist eine wesentliche Rechenformel. Sind $A_1$ und $A_2$ auseinanderliegend, so zerfallen alle zu $A_1 \cup A_2$ gehörenden Versuchsergebnisse in zwei Gruppen: In diejenigen, die in $A_1$ liegen, und diejenigen, die in $A_2$ liegen. Die Summe ihrer Anzahlen ist damit identisch mit der absoluten Häufigkeit von $A_1 \cup A_2$. Und das überträgt sich unmittelbar auf die relativen Häufigkeiten. So erhalten wir in obigem Beispiel für $n = 20$: $h_{20}(A_1 \cup A_2) = \frac{12}{20} = \frac{6}{20} + \frac{6}{20} = h_{20}(A_1) + h_{20}(A_2)$.

Weiterhin gelten noch folgende Rechengesetze, die man aber auch bereits theoretisch aus den Eigenschaften 1.–3. ableiten kann:
4. $h_n(\overline{A_1}) = 1 - h_n(A_1)$ für jedes Ereignis$A_1$,
5. $h_n(A_1 \cup B) = h_n(A_1) + h_n(B) - h_n(A_1 \cap B)$ für beliebige Ereignisse $A_1, B$.

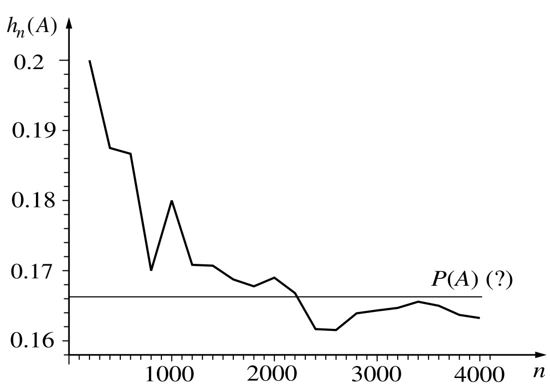
Letztere Eigenschaft kann auch günstig als Rechenkontrolle in obigem Beispiel eingesetzt werden. Wir haben also z. B.

$$\text{für } n = 15: \ \frac{11}{15} = \frac{5}{15} + \frac{10}{15} - \frac{4}{15}, \qquad \text{für } n = 20: \ \frac{14}{20} = \frac{6}{20} + \frac{13}{20} - \frac{5}{20}.$$

Diese Formel läßt sich unmittelbar plausibel erklären: Bei der Bestimmung von $h_n(A_1 \cup B)$ zählen wir alle Versuchsergebnisse, die in $A_1$ oder $B$ liegen. Die in $A_1 \cap B$ liegenden Ergebnisse wurden jedoch sowohl in $h_n(A_1)$ als auch in $h_n(B)$ erfaßt, also doppelt. Wir müssen sie also einmal wieder subtrahieren, d. h. $h_n(A_1) + h_n(B) - h_n(A_1 \cap B)$ bilden.

Zu 2.: Wir werden jetzt $A$ festhalten und die Anzahl $n$ der Versuche erhöhen. Wir betrachten zunächst ein Beispiel, und zwar das einmalige Würfeln mit einem Würfel und dabei das Eintreten des Ereignisses $A = \{6\}$. Eine Computersimulation ergab folgende Ergebnisse:

| Anzahl d. Würfe | absolute Häufigkeit | relative Häufigkeit | Anzahl d. Würfe | absolute Häufigkeit | relative Häufigkeit |
|---|---|---|---|---|---|
| $n$ | $H_n(A)$ | $h_n(A)$ | $n$ | $H_n(A)$ | $h_n(A)$ |
| 200 | 40 | 0,2 | 2200 | 367 | 0,1668182 |
| 400 | 75 | 0,1875 | 2400 | 388 | 0,1615385 |
| 600 | 112 | 0,1866667 | 2600 | 420 | 0,1639285 |
| 800 | 136 | 0,17 | 2800 | 459 | 0,1639286 |
| 1000 | 180 | 0,18 | 3000 | 493 | 0,1643333 |
| 1200 | 205 | 0,1708333 | 3200 | 527 | 0,1646875 |
| 1400 | 239 | 0,1707143 | 3400 | 563 | 0,1655882 |
| 1600 | 270 | 0,16875 | 3600 | 594 | 0,165 |
| 1800 | 302 | 0,1677778 | 3800 | 622 | 0,1636842 |
| 2000 | 338 | 0,169 | 4000 | 653 | 0,16325 |



Am Bild erkennen wir: Bei ständiger Vergrößerung der Anzahl $n$ der Versuche oszillieren die relativen Häufigkeiten zunehmend weniger und pegeln sich auf einen Wert ein. Dieser „Grenzwert" ist die „theoretische Häufigkeit" für das Eintreten des Ereignisses $A$. Wir nennen ihn die Wahrscheinlichkeit von $A$ und bezeichnen diese Zahl mit $P(A)$. Dabei sind aber zwei Fragen kritisch zu stellen: Zum ersten können wir nicht unendlich lange Versuchsserien durchführen und sind somit nicht in der Lage, den genauen Wert $P(A)$ für die theoretische Häufigkeit des Eintretens von $A$ zu ermitteln. Es ist ein wichtiges Anliegen der Statistik, diese Lücke zwischen praktischer Versuchsdurchführung und theoretischen Modellvorstellungen möglichst gut zu überbrücken. Zum zweiten halten wir den Vorgang des Einpegelns der relativen Häufigkeiten, wie er in obigem Beispiel gezeigt wird, für plausibel, es ist aber nicht selbstverständlich, daß bei einer erneuten Würfelserie das Einpegeln auf den gleichen Wert wie vorher erfolgt. Wir werden uns im weiteren auf den Standpunkt stellen, daß diese „Gesetzmäßigkeit des Zufalls" vorliegt und ein entsprechend theoretisches Modell aufbauen, in dem diese Eigenschaft (das sog. Gesetz der großen Zahlen) streng mathematisch bewiesen werden kann.

### 3.1.3 Die klassische Wahrscheinlichkeit und die geometrische Wahrscheinlichkeit

Wie bereits in der Einleitung zu diesem Kapitel unter der Bezeichnung „logisch plausibler Aspekt" erwähnt wurde, handelt es sich bei der klassischen Wahrscheinlichkeit um einen Sonderfall, bei dem man durch plausibles Schließen die zugrundeliegenden Wahrscheinlichkeiten $P(A)$ genau berechnen kann, ohne den zugrunde liegenden zufälligen Versuch jemals durchgeführt zu haben. In allgemeineren Situationen ist dies nicht mehr möglich.

Zu dieser Berechnung benötigt man die Angabe der Anzahlen der in $A$ und in der Grundmenge $\Omega$ liegenden Elemente. Wenn diese Mengen sehr groß und nach „bestimmten Prinzipien" gebildet werden, benutzt man hierzu Formeln der Kombinatorik. Aus diesem Grunde wollen wir an dieser Stelle zunächst die durch die Kombinatorik erfaßten Standardsituationen mit ihren zugehörigen Formeln betrachten.

#### 3.1.3.1 Kombinatorik

In der Wahrscheinlichkeitsrechnung stehen wir des öfteren vor dem Problem, für betrachtete endliche Versuchsergebnismengen $A$ die Anzahl der sich in ihnen befindlichen Elemente zu ermitteln. Man könnte diese Aufgabe natürlich am einfachsten durch Abzählen lösen. Die auftretenden Mengen sind jedoch häufig sehr groß, so daß andere Wege beschritten werden müssen.

**Beispiel 3.1/9**: Es treffen sich 80 Personen, jeder begrüßt jeden. Wie viele Begrüßungen sind das insgesamt? Oder anders ausgedrückt: Wie viele Paare lassen sich aus 80 Personen bilden?

**Beispiel 3.1/10**: Ein Fragebogen enthalte 25 Fragen zu je 3 Antworten, von denen jeweils genau eine angekreuzt werden soll. Wie viele verschiedene Möglichkeiten gibt es, den Fragebogen auszufüllen?

Worin bestehen nun diese „anderen Wege" zur Bestimmung von Anzahlen endlicher Mengen? Wir stellen zunächst einschränkend fest, daß wir nicht das Ziel verfolgen, für **jede** endliche Menge eine Formel für die Anzahl ihrer Elemente anzugeben. Das ist offensichtlich nicht möglich. Wir wollen diese Aufgabe vielmehr für gewisse Mengen $A$ lösen, die nach bestimmten Konstruktionsprinzipien gebildet wurden. Es gibt dabei 6 Standardsituationen, sog. **Permutationen**, **Variationen** und **Kombinationen**, und diese jeweils ohne oder mit Wiederholung. Das grundlegende Problem bei der Anwendung der Kombinatorik besteht stets darin zu erkennen, ob überhaupt eine Standardsituation zutreffend ist, und welche dann die richtige ist. Theoretisch bedeutet dies, daß man bei derartigen Aufgabenstellungen die Elemente der zu untersuchenden Menge $A$ gedanklich in umkehrbar eindeutiger Weise den „Möglichkeiten" einer Standardsituation zuordnet, dadurch auf die Gleichheit der Anzahl der in $A$ enthaltenen

# 4 Statistische Testtheorie

## 4.1 Einführung

### 4.1.1 Grundbegriffe

In Kapitel 2 haben wir uns in der beschreibenden Statistik mit der Analyse empirischer Daten durch geeignete Zusammenfassungen, grafische Darstellungen und Berechnung charakteristischer Kenngrößen beschäftigt. Das Ziel bestand darin, die in ihnen enthaltenen Informationen über die zu untersuchenden Sachverhalte zu erkennen. Das Kapitel 3 diente dazu, angepaßte theoretische Modelle der Wahrscheinlichkeitsrechnung zu formulieren. Wir wollen nun mit der sogenannten schließenden Statistik, speziell mit der sogenannten Testtheorie, beginnen, deren Ziel darin besteht, die empirisch gewonnenen Befunde aus der Sicht wahrscheinlichkeitstheoretischer Modellbildungen zu bewerten und damit erkannte systematische Effekte (sogenannte Signifikanz) zu bestätigen. Man geht nun bei der Testtheorie methodisch in der Weise vor, daß man die z. B. mittels der beschreibenden Statistik gewonnenen Erkenntnisse in Form einer statistischen Hypothese (und häufig zusätzlich einer Alternativhypothese) formuliert und deren Gültigkeit anhand einer neuen Stichprobe testet.

Bevor wir die einzelnen Schritte bei der Durchführung eines solchen Tests diskutieren, wollen wir das allgemeine Anliegen an 2 Beispielen verdeutlichen. Im ersten Beispiel betrachten wir den Intelligenzquotienten ($IQ$) für einen bestimmten Personenkreis, z. B. alle Studienbewerber einer gegebenen Fachrichtung einer bestimmten Universität. Uns interessiert, ob sich der Intelligenzquotient bei dieser Population im Laufe der Zeit verändert hat. Es sei bekannt, daß der Durchschnitts-$IQ$ vor 5 Jahren 120 betrug, und bei der jetzigen Erhebung wurde für $N = 150$ Bewerber ein Mittelwert $\bar{x}$ des $IQ$ von 122 gemessen. Erlauben uns diese Daten, eine systematische Erhöhung des $IQ$ berechtigt anzunehmen? Im zweiten Beispiel betrachten wir das Problem der Modifikation von psychologischen Tests durch sogenannte Paralleltestaufgaben. Das Ziel besteht darin, 2 Testformen (A und B) mit dem gleichen Schwierigkeitsgrad zu entwickeln. Um die Gleichwertigkeit statistisch zu untersuchen, werden Versuchspersonen ($V\,pn$) gebeten, beide Aufgaben zu lösen. Dabei entstand beispielsweise folgendes Ergebnis (mit 316 $V\,pn$):

| | | Aufgabe 1 | | |
|---|---|---|---|---|
| | | gelöst | nicht gelöst | $\sum$ |
| Aufgabe 2 | gelöst | 61 | 92 | 153 |
| | nicht gelöst | 111 | 52 | 163 |
| | $\sum$ | 172 | 144 | 316 |

Man hat dazu den relativen Anteil der gelösten Aufgabe 1 $\left(= \dfrac{61+111}{316}\right)$ mit demjenigen der gelösten Aufgabe 2 $\left(= \dfrac{61+92}{316}\right)$ zu vergleichen, de facto also die Anzahlen derjenigen Fälle, bei denen Aufgabe 1 gelöst und Aufgabe 2 nicht gelöst wurde ($= 111$) mit den Fällen, bei denen Aufgabe 2 gelöst und Aufgabe 1 nicht gelöst ($= 92$) wurde. Wären diese beiden Zahlen identisch, dann wären auch die relativen Häufigkeiten für gelöst/nicht gelöst bei beiden Aufgaben gleich. In diesem Fall wäre die parallele Verwendung der Aufgaben statistisch abgesichert, d. h., sie hätten den gleichen Schwierigkeitsgrad. Nun sind aber im obigen Zahlenbeispiel diese beiden Werte nicht gleich, und es erhebt sich das Problem, zu entscheiden, ob dies als zufällige Schwankung akzeptiert werden kann oder bereits einen systematischen Unterschied zwischen den Schwierigkeiten signalisiert. Wir werden im weiteren Beispiel 4.2/1 ausführlich diskutieren. Die Behandlung von Beispiel 4.2/2 findet der Leser im Abschnitt 4.3.2.1.1.

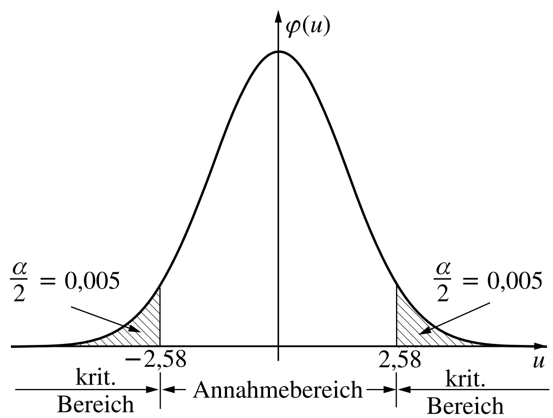
Am Anfang der Durchführung eines Tests steht die Formulierung von statistischen Hypothesen, einer zu prüfenden **Nullhypothese** $H_0$ und einer dem entgegengestellten **Alternativhypothese** $H_1$. Diese Hypothesen sind Aussagen, die die theoretische Verteilung in der zugehörigen Grundgesamtheit betreffen. In obigem Beispiel 4.2/1 würde man beim $IQ$ von einer Normalverteilung ausgehen, wobei wir der Einfachheit halber die Varianz $\sigma^2$ als bekannt voraussetzen wollen, es sei $\sigma^2 = 36$ (Die Bekanntheit und zeitliche Konstanz dieses Wertes könnte z. B. aus langjährigen Messungen gewonnen worden sein. Verzichtet man auf diese Annahme, so führt das auf eine etwas kompliziertere Situation – den $t$-Test, vgl. Abschnitt 4.2.4.3). Wir untersuchen den Mittelwert-Parameter $\mu$ für den $IQ$ der aktuellen Erhebung mit dem Vorsatz nachzuweisen, daß sich dieser systematisch von 120 unterscheidet. Zu diesem Zweck formulieren wir die Nullhypothese $H_0$: $\mu = \mu_0 = 120$ mit dem Ziel, diese abzulehnen und damit einen signifikanten Unterschied nachzuweisen. Mögliche Alternativhypothesen $H_1$ wären: $\mu \neq 120$, bzw. $\mu > 120$ oder $\mu < 120$ (vgl. weiter unten).

Bei jeder statistischen Entscheidung treffen reale Beobachtungen (im Beispiel der Mittelwert $\bar{x} = 122$ bei $N = 150$ Versuchspersonen) und zugrundeliegende Wahrscheinlichkeitsannahmen ($H_0$: $\mu = \mu_0 = 120$) aufeinander. Daraus resultiert immer auch das Risiko einer Fehlentscheidung. In Folge der Betrachtung der Art der Fehlentscheidung differenzieren wir zwischen dem Fehler 1. Art und 2. Art. Ein **Fehler 1. Art** liegt vor, wenn wir die zugrundeliegende **Nullhypothese ablehnen, obwohl sie richtig ist**. Wir sprechen vom Ablehnungsfehler und nennen seine Wahrscheinlichkeit Irrtumswahrscheinlichkeit oder auch Risiko 1. Art und bezeichnen sie mit $\alpha$. Ein solcher Fehler entsteht dann, wenn die konkrete Stichprobe bezüglich der Annahme in der Nullhypothese aufgrund der Wirkung des Zufalls „extrem“ ausgefallen ist. In diesem Fall würde das statistische Prüfverfahren Stichproben- und Grundgesamtheitsverteilung als „verschieden“ analysieren, obwohl sie identisch sind. Die Nullhypothese wird falscher Weise abgelehnt. Ein **Fehler 2. Art** liegt vor, wenn wir die **Nullhypothese beibehalten, obwohl sie falsch ist**. Diesen Fehler nennt man auch Annahmefehler, seine Wahrscheinlichkeit heißt Risiko zweiter Art und wird mit $\beta$ bezeichnet. Solch ein Fehler kann auftreten, wenn die Stichprobe „nur zufällig im Sinne von $H_0$“ ausfällt, obwohl in Wirklichkeit $H_1$ zutrifft. Wir werden im weiteren nur den Fehler 1. Art $\alpha$ festlegen. Man spricht dann von einem **Signifikanztest** zum Signifikanzniveau $\alpha$ bzw. mit der Sicherheitswahrscheinlichkeit $1 - \alpha$. Man spricht von einem signifikanten Untersuchungsergebnis, wenn $H_0$ abgelehnt wird. Wir merken an, daß die Wahrscheinlichkeit für einen Fehler 2. Art nicht angegeben werden kann, da sie von dem unbekannten Parameter abhängt. Wir können nur indirekt Einfluß auf die Größe des Risikos 2. Art nehmen. Signifikanztests sind so konstruiert, daß der Anwender nur $\alpha$ frei wählen kann. Mit der Wahl der Größe von $\alpha$ wird $\beta$ dadurch „automatisch“ mit beeinflußt. Der Zusammenhang besteht darin, daß ein größeres $\alpha$ ein kleineres $\beta$ und umgekehrt ein kleineres $\alpha$ ein größeres $\beta$ gewöhnlich zur Folge hat. Der Anwender des statistischen Verfahrens sollte sich im Vorfeld überlegen, welchen Fehler er eher akzeptieren kann. Bei dieser Entscheidung hilft es ihm, wenn er sich vor Beginn der Untersuchung Gedanken bezüglich der Konsequenzen macht, d. h., will er eher einen Ablehnungs- oder eher einen Annahmefehler zulassen. Dies kann die Statistik nicht entscheiden. Hier ist vom **inhaltlichen** Problem auszugehen – in jedem einzelnen Fall neu. Besteht die Gefahr, daß in dem Fall, da die Nullhypothese abgelehnt wird, obwohl sie richtig ist, daß dann die negativen Konsequenzen überwiegen, so wird der Anwender dieses Risiko möglichst klein halten. Er wird die Hypothesenprüfung auf einem hohen Signifikanzniveau, d. h. mit kleinem $\alpha$ realisieren. Ist im umgekehrten Fall damit zu rechnen, daß ausgesprochen negative Konsequenzen dann zu erwarten sind, wenn wir die Nullhypothese annehmen, obwohl sie falsch ist, so werden wir mit einem niedrigeren Signifikanzniveau, d. h. mit einem größeren $\alpha$ prüfen, womit sich auch der $\beta$-Fehler verkleinert. Ein weiteres Motiv für die Festlegung von $\alpha$ ergibt sich daraus, daß man möglichst „streng gegenüber der angestrebten Entscheidung" verfährt, also möglichst signifikante Aussagen erhält. Ist also die Ablehnung der Nullhypothese die wichtigere Aussage, so wählt man ein kleineres $\alpha$ (z. B. $\alpha = 0{,}01$), im anderen Fall wählt man ein größeres $\alpha$ (z. B. $\alpha = 0{,}05$).

Zur konkreten Durchführung des Tests benötigen wir nun die sogenannte **Testgröße**, die so beschaffen ist, daß sie eine Entscheidung bezüglich der Nullhypothese sinnvoll gestattet. Die mathematische Herleitung der konkreten Form der Testgröße ist mitunter kompliziert und keine Aufgabe des Anwenders. Ihm obliegt es, die am Ende häufig relativ einsichtigen Formeln zur Berechnung der Testgröße richtig zu handhaben. In unserem obigen Beispiel 1 verwendet man als Testgröße den Ausdruck

$$\begin{aligned} u &= \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{N} = \frac{122 - 120}{6} \cdot \sqrt{150} \\ &= 4{,}082. \end{aligned}$$

Aufgrund der Tatsache, daß der $IQ$ normalverteilt ist, folgt, daß die Testgröße $u$ einer standardisierten Normalverteilung entspricht. Die Entscheidung wird nun dadurch getroffen, daß man in geeigneter Weise ein kritisches Gebiet, welches in Intervallform gewählt wird, angibt und die Nullhypothese genau in dem Fall ablehnt, wenn die Testgröße in dieses Gebiet fällt. Von genereller Bedeutung ist dabei weiterhin die Art der Fragestellung. Wir unterscheiden zwischen der **einseitigen** und der **zweiseitigen Fragestellung**. In unserem Beispiel würde eine zweiseitige Fragestellung durch die Alternativhypothese $H_1\colon \mu \neq \mu_0 = 120$ gegeben sein, d. h., sowohl große positive als auch große negative Abweichungen zwischen $\overline{x}$ und $\mu$ sind kritisch. Hier besteht das kritische Gebiet aus 2 Intervallen, die sich am linken und rechten Rand der standardisierten Normalverteilungskurve befinden. Das Risiko 1. Art teilt sich dementsprechend zu je $\frac{\alpha}{2}$ auf. Die nachfolgende Abbildung soll den Gedanken verdeutlichen:

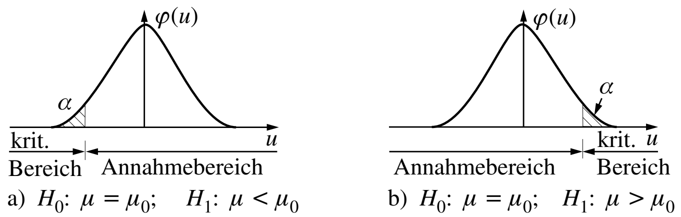



*Die Irrtumswahrscheinlichkeit als Fläche unter der Normalkurve*
*(zitiert nach Clauß/Ebner 1992, S. 195)*

In einem solchen Fall schließen wir aus der Ablehnung der Nullhypothese, daß entweder $\mu < \mu_0$ oder $\mu > \mu_0$ ist. Aus dem Bild ersehen wir, daß in diesem Fall die Nullhypothese genau dann abgelehnt wird, wenn $|u| \geq u_{\alpha;\text{zweis}}$ mit $u_{\alpha;\text{zweis}} = 2{,}58$ erfüllt ist. Man nennt diese Zahl den **kritischen Wert** bei zweiseitiger Fragestellung. Er ergibt sich in unserem Beispiel als Quantil der Ordnung $1 - \frac{\alpha}{2}$ bei der standardisierten Normalverteilung: $u_{\alpha;\text{zweis}} = z_1 - \frac{\alpha}{2}$ (vgl. Abschnitt 3.2.3.3).

Anders stellt sich die Situation bei der **einseitigen Fragestellung** für uns dar. Bei der Formulierung der Nullhypothese ist zu berücksichtigen, daß bei der **Alternativhypothese** entweder $\mu < \mu_0$ oder $\mu > \mu_0$ definiert wird. Mit anderen Worten bedeutet das, daß wir für den Abweichungsfall **vor** Beginn der Untersuchung wohlbegründet davon ausgehen können, daß ein Unterschied nur **in einer Richtung** auftreten kann. In unserem Beispiel ist dies $H_1$: $\mu > \mu_0 = 120$.

Der Begrenzungswert des kritischen Bereiches wird nun nur an einer Seite der Verteilung mit dem Flächenanteil von $\alpha$ festgelegt. In Abhängigkeit davon, ob die Alternativhypothese durch $\mu > \mu_0$ (wie im obigen Beispiel) oder im anderen Fall durch $\mu < \mu_0$ formuliert wird, befindet sich der kritische Bereich auf der rechten oder linken Seite der Prüfverteilung. Die nachfolgende Abbildung soll dies verdeutlichen:




*Annahmebereich und kritischer Bereich bei einseitiger Fragestellung (nach* CLAUSS/EBNER *1992, S.202)*

Der Absolutbetrag des kritischen Wertes $u_{\alpha;\text{eins}}$ der Prüfgröße ist demnach bei einseitiger Fragestellung kleiner als bei zweiseitiger Fragestellung. In unserem Beispiel erhalten wir für $\alpha = 0{,}01$ den kritischen Wert $u_{\alpha;\text{eins}} = z_{1-\alpha} = z_{0,99} = 2{,}326$ (vgl. auch Tafel 2 und die Anmerkung nach Beispiel 7.7/10). Im Ergebnis des Tests lehnen wir die Nullhypothese ab. Die angegebenen Daten erlauben uns also den Schluß, daß eine signifikante Steigerung des $IQ$ in den letzten 5 Jahren zu verzeichnen war.

Zusammenfassend wollen wir im weiteren bei der Durchführung statistischer Tests in einheitlicher Weise nach folgendem 7-Punkte-Schema vorgehen:
1. Bemerkungen zu den Voraussetzungen des Tests: Diese beziehen sich sowohl auf die verwendeten Daten (z. B. Datenniveau, bestimmte testbedingte Einschränkungen) als auch auf die zugehörige Grundgesamtheit.
2. Formulierung der Null- und der Alternativhypothese,
3. Festlegung des Signifikanzniveaus $\alpha$,
4. Angabe der Testgröße (auch Prüfgröße genannt),
5. Angabe des kritischen Bereiches,
6. Berechnung des konkreten Wertes der Testgröße auf der Grundlage der Daten,
7. Entscheidung und Formulierung des Testergebnisses.

Für unser 1. Beispiel erhalten wir:
1. Test auf Mittelwert bei metrischen normalverteilten Daten ($IQ$) mit bekannter Varianz $\sigma^2 = 36$.
2. $H_0$: $\mu = \mu_0 = 120$, $H_1$: $\mu > 120$ (einseitig)
3. $\alpha = 0{,}01$
4. $u = \dfrac{\bar{x} - \mu_0}{\sigma} \cdot \sqrt{N}$
5. $u \geq u_{\text{krit}} = u_{\alpha;\text{eins}} = 2{,}326$ (vgl. Tafel 2)
6. $u = \dfrac{122 - 120}{6} \cdot \sqrt{150} = 4{,}082$
7. $u = 4{,}082 > 2{,}326$, also Ablehnung von $H_0$. Es besteht eine signifikante Erhöhung des $IQ$ gegenüber dem Vergleichswert von 120.

Abschließend wollen wir noch auf eine Besonderheit hinweisen, die beim Umgang mit Computern auftritt. Zur Testentscheidung benötigt man kritische Werte (Grenzen des kritischen Bereiches). Diese werden aus den Quantilen der Verteilungsfunktion, die zur jeweiligen Testgröße gehört, berechnet und liegen gewöhnlich vertafelt vor (vgl. Tafelanhang, Kap. 7). Man ist somit überhaupt erst in der Lage, Tests „per Hand“ durchzuführen. Nun ist es aber wenig sinnvoll, im Computer die vollständigen Tafeln der kritischen Werte zu speichern, da dieser in der Lage ist, nur den jeweils benötigten konkreten Wert auszurechnen. Dies geschieht aber in einer etwas anderen, jedoch dazu äquivalenten Form, bei der nicht die schwieriger zu bestimmenden Quantile, sondern die Verteilungsfunktion der Prüfgröße selbst verwendet wird. Der Computer berechnet die sog. **Überschreitungswahrscheinlichkeit** $\gamma$. Sie ist die Wahrscheinlichkeit dafür, daß im Sinne der Alternativhypothese „noch ungünstigere“ Werte der Prüfgröße, als der konkret erhaltene, auftreten. Wir demonstrieren dies an unserem obigen Beispiel: Dort erhielten wir $u = 4{,}082$, wobei eine standardisierte Normalverteilung zugrunde liegt. Die Überschreitungswahrscheinlichkeit $\gamma$ ergibt sich damit wie folgt:

$$\gamma = P(T^* \geq 4{,}082) = 1 - P(T^* < 4{,}082) = 1 - \Phi(4{,}082)^{1)}$$
$$< 1 - 0{,}9998 = 0{,}0002 \text{ (vgl. Tafel 2).}$$

Dabei bezeichnet $T^*$ – enstsprechend der Vereinbarung im vorhergehenden Abschnitt 3.4.2 – die zu $u$ gehörige normalverteilte **zufällige** Testgröße.

Die Entscheidung wird dann wie folgt gefällt:

Einseitige Fragestellung: Wenn $\gamma \leq \alpha$ gilt, dann lehnen wir $H_0$ ab.
Zweiseitige Fragestellung: Wenn $\gamma \leq \dfrac{\alpha}{2}$ gilt, dann lehnen wir $H_0$ ab.

Im Beispiel erhalten wir $\gamma = 0{,}0001 < \alpha = 0{,}01$, also ist $H_0$ abzulehnen. Die Überschreitungswahrscheinlichkeit hat den Vorteil, daß sie vom Gesichtspunkt der Wahrscheinlichkeit her (und damit in gewissem Sinne in natürlicher Weise) eine Einschätzung des konkret erhaltenen Wertes der Prüfgröße erlaubt.

*Anmerkung*: *Mitunter spricht man im Falle von*
$\gamma \leq 0{,}001$ *von einer hochsignifikanten (symbolisch:* $^{***}$*) Abweichung,*
$\gamma \leq 0{,}01$ *von einer sehr signifikanten (symbolisch:* $^{**}$*) Abweichung,*
$\gamma \leq 0{,}05$ *von einer signifikanten (symbolisch:* $^{*}$*) Abweichung.*

### 4.1.2 Klassifikation statistischer Tests

Die Klassifikation von statistischen Verfahren ist deshalb notwendig, weil diese wesentlich von den konkreten Bedingungen und Voraussetzungen abhängen. Außerdem erleichtert die Klassifikation im Sinne einer Lernhilfe das Finden der geeigneten Tests für die jeweils vorliegenden Bedingungen. Insgesamt erfolgt dabei eine Ordnung der Tests nach der Art der Daten, die sie verarbeiten.

Wir unterscheiden eine Reihe von Fragestellungen, deren Antworten das Suchfeld für den jeweiligen Test systematisch eingrenzen, bis schließlich nur noch am „besten geeignete“ Tests für das entsprechende Urmaterial zur Wahl stehen.

1. Die Grundfragestellung
   Wir unterscheiden zwischen zwei Arten von Fragestellungen:
   a) Kann man annehmen, daß eine vorliegende Stichprobe, d. h. die empirischen Daten der Untersuchung aus einer Population, aus einer Grundgesamtheit mit einer bestimmten, bekannten Verteilung stammt? Beispielsweise können wir fragen: Liegt eine Normalverteilung der Daten vor?

[1)] Tafel 2 reicht nur bis zum $u$-Wert von 3,49 und ergibt 0,9998. Der Wert 4,082 ist noch größer als 3,49, so daß der Tafelwert garantiert über 0,9998 liegt. Damit ist $1 - \Phi(4{,}082)$ kleiner als $1 - 0{,}9998 = 0{,}0002$.

Diese Frage ist deshalb besonders interessant, weil wir nach ihrer Bejahung mit parametrischen Tests (vgl. Frage 5) arbeiten dürfen. Diese Frage nach der Übereinstimmung bedeutet also den Vergleich einer empirischen (Stichprobe) mit einer theoretischen (Grundgesamtheit) Verteilung. Tests, die Antwort auf diese Frage geben, sind die sogenannten **Anpassungstests** (vgl. Kapitel 4.2)

Als Beispiel für einen Anpassungstest mag uns nachfolgende Fragestellung dienen: Wir untersuchten im Rahmen einer Qualitätskontrolle eine Warensendung (Stichprobe) hinsichtlich des Anteils von Ausschußteilen und stellten fest, daß 3 % nicht in Ordnung waren. Nun wissen wir, daß die Gesamtproduktion (Grundgesamtheit) eines Werkes durchschnittlich 0,5 % Ausschußteile beinhaltet. Mit Hilfe eines Anpassungstestes können wir unter Akzeptanz einer festgelegten Irrtumswahrscheinlichkeit die Frage beantworten, ob die untersuchte Warensendung tatsächlich aus diesem Unternehmen stammt.

b) Kann man annehmen, daß zwei oder mehr vorliegende Stichproben aus der gleichen Grundgesamtheit stammen, oder unterscheiden sich die Grundgesamtheiten, aus denen die Stichproben entnommen wurden, signifikant voneinander? Hier vergleichen wir zwei oder mehr empirische Verteilungen miteinander und untersuchen, ob sie unter den gleichen Bedingungen erhoben wurden. Wir suchen Antwort auf die Frage, ob für die Stichproben die gleiche statistische Bedingung vorliegt. Tests, die diese Frage beantworten, bezeichnen wir als **Unterschiedstests** (vgl. Kap. 4.3). Eine praktische Fragestellung für einen Unterschiedstest können wir uns wie folgt vorstellen: Wir untersuchten eine Gruppe von Lehramtskandidaten in Hamburg (Stichprobe 1) und eine Gruppe in München (Stichprobe 2) hinsichtlich ihrer Fähigkeit zur sozialen Kompetenz beim Umgang mit Schülern. Mit Hilfe eines Unterschiedstests können wir z. B. Antwort auf die Frage geben, ob sich die beiden Gruppen bzgl. der sozialen Kompetenz unterscheiden oder nicht.

   Bei den Unterschiedstests trennen wir drei inhaltliche Aspekte:

   a) Wir fragen nach Unterschieden in der zentralen Tendenz, d. h. nach Mittelwertsunterschieden. Dann helfen uns die sogenannten **Lokationstests**.
   b) Wir suchen nach Unterschieden in den Streuungen zwischen den Stichproben, d. h., wir betrachten die Verteilung der Daten um den Mittelwert herum. In diesem Fall nutzen wir die sogenannten **Dispersionstests**.
   c) Uns interessieren Unterschiede sowohl bzgl. der zentralen Tendenz als auch bzgl. der Streuung. Dabei setzen wir die sogenannten **Omnibustests** ein.

   Anmerkung: Messen wir an einem Objekt mehrere Variable, dann interessiert uns die Frage, ob ein Zusammenhang zwischen diesen Variablen besteht. Ist ein solcher vorhanden, dann interessiert überdies, welcher Art dieser Zusammenhang zwischen den Variablen ist. Antwort auf die erste Frage gibt uns die sogenannte **Korrelationsanalyse** (vgl. Abschnitt 5.1), und Antwort auf die zweite Frage gibt uns die sogenannte **Regressionsanalyse** (vgl.Abschnitt 2.3.2.6 und 5.1.5.4).

2. Die Art der Stichprobenerhebung

   Hinsichtlich der Art der Stichprobenerhebung unterscheiden wir zwei große Gruppen von Stichproben. Wir differenzieren zwischen den unabhängigen und den abhängigen (auch korrelierend genannten) Stichproben.

   **Unabhängige Stichproben** entstehen z. B., wenn jede Stichprobe an anderen Objekten, die nach einem Zufallsprinzip völlig unabhängig voneinander ausgewählt wurden, erhoben wird. Zwischen Objekten verschiedener Stichproben existiert dann keinerlei Informationsverbindung, und es kann auch keine gegenseitige Zuordnung hergestellt werden. Unabhängige Stichproben liegen beispielsweise vor, wenn wir eine 4. Klasse in Dresden und eine 4. Klasse in Frankfurt hinsichtlich ihres Wortschatzes untersuchen. Das Zufallsprinzip der Auswahl bestand z. B. darin, daß in einem Lostopf von allen Städten Deutschlands mit mehr als 500 000 Einwohner je eine benannte 4. Klasse enthalten war und wir rein zufällig die Klassen aus Dresden und Frankfurt gezogen haben.

**Abhängige Stichproben** stehen uns dann zur Verfügung, wenn an denselben Objekten zwei oder mehr verschiedene Bedingungen untersucht wurden. Dann entstehen mindestens Meßwertpaare pro Objekt. Wir sprechen von abhängig, weil die erhobenen Daten auf das gleiche Objekt zurückführbar sind. Wenn wir also ein überdurchschnittlich intelligentes Kind hinsichtlich seiner Leistungsfähigkeit in Mathematik und Physik untersuchen, dann wird sich diese hohe Intelligenz in beiden Bereichen widerspiegeln. Es existiert eine Art übergeordnete Abhängigkeit der Untersuchungsergebnisse.
Mit abhängigen Stichproben haben wir es häufig im Bereich von Lern- und Entwicklungsuntersuchungen, aber auch im Bereich der Kontrolle von Arbeitsgestaltungsmaßnahmen zu tun. Stellen wir uns vor, wir untersuchen die Belastung an einem Arbeitsplatz vor und nach einer Gestaltungsmaßnahme. Um zu prüfen, ob z. B. die Belastung reduziert wurde, müssen wir die Daten beider Erhebungen als abhängige Stichproben behandeln.

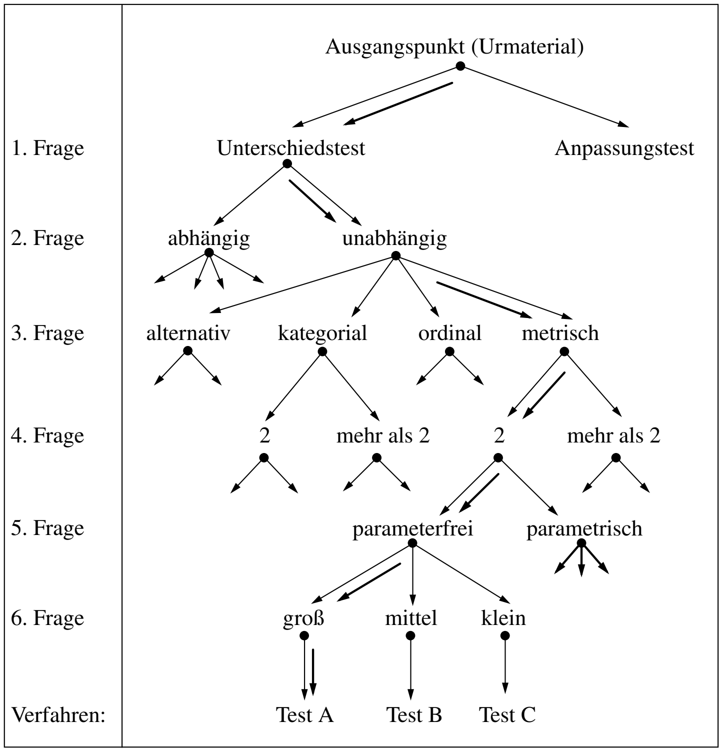
3. Die Arten der Daten
   Die Frage nach der Art der Daten ist die Frage nach dem Informationsgehalt. Wie wir im Kapitel 3.1.3. erfahren haben, unterscheiden wir **alternative**, **kategoriale**, **ordinale** und **metrische Daten**. Da jedes statistische Verfahren die genaue Bestimmung des Informationsgehaltes zur Voraussetzung hat, differenzieren wir bei der Suche nach dem geeigneten Test entsprechend dem Datenniveau.

4. Die Anzahl der Stichproben
   Hinsichtlich der Anzahl der Stichproben klassifizieren wir in drei Gruppen von Tests:
   a) Es liegt nur **eine** Stichprobe vor. In diesem Fall handelt es sich im wesentlichen bzgl. der Fragestellung um einen Anpassungstest.
   b) Uns stehen **zwei** Stichproben zur Verfügung. Dann interessiert uns entweder ein Unterschiedstest, oder wir interessieren uns für eventuelle Korrelationen zwischen ihnen.
   c) Ist die Anzahl der Stichproben **größer als zwei**, so unterscheiden wir zwischen dem **Globalvergleich** und dem multiplen Vergleich. Beim Globalvergleich ist die Frage zu beantworten, ob es zwischen allen Verteilungen **insgesamt** einen Unterschied gibt. Beim **multiplen Vergleich** fragen wir, zwischen welchen der **einzelnen** Verteilungen es einen Unterschied gibt?

5. Die Verteilungsannahme
   Diese Unterscheidung wird in vielen Lehrbüchern als die wichtigste dargestellt. Wir trennen hier in die verteilungsabhängigen – auch parametrischen – Tests und in die verteilungsunabhängigen – auch verteilungsfreien oder nichtparametrischen – Tests.
   Bei **parametrischen Tests** werden typische Parameter der Verteilung untersucht. Meist sind das der Mittelwert und die Varianz. So können z. B. parametrische Tests an die Voraussetzung der Normalverteilung gebunden sein. Parametrische Tests sind hinsichtlich der Genauigkeit ihrer Aussage stärker als parameterfreie Tests, da sie ja auch wesentlich mehr Voraussetzungen erfordern.
   **Parameterfreie Tests** erfordern keine Annahme über die Art der Verteilung der Daten. Sie finden immer dann Anwendung, wenn z. B. die Voraussetzungen für eine Verteilungsannahme nicht unmittelbar ersichtlich sind. Diese Tests arbeiten auf der Grundlage aller Untersuchungsergebnisse der Stichproben, d. h., bei ihnen können wir nicht mit typischen Maßzahlen von Verteilungen arbeiten.

6. Der Umfang der Stichprobe
   Als letzte Klassifikationshilfe steht uns die Größe der Stichprobe zur Verfügung. Hier gibt es für die einzelnen statistischen Verfahren keine festen Anzahlen. Es wird zwischen großen, mittleren und kleinen Stichproben unterschieden. In vielen Fällen ist es so, daß mit wachsendem Stichprobenumfang eine immer bessere Annäherung an die theoretische Verteilung erfolgt. Deshalb dürfen wir dann mit Näherungsformeln, die im allgemeinen wesentlich anwendungsfreundlicher sind, arbeiten.

Fassen wir die Teilfragestellungen zur Testklassifikation zusammen, dann können wir das empfohlene Vorgehen am nachfolgenden logischen Baum grafisch gut verdeutlichen:




## 4.2 Anpassungstests

Wie bereits in Abschnitt 4.1.2 erwähnt wurde, besteht eine der beiden großen Untergruppen von Tests aus sogenannten Anpassungstests. Sie beantworten die Frage, ob die empirisch gefundene Verteilung innerhalb einer Stichprobe mit einer theoretischen Verteilung übereinstimmt oder nicht. Anpassungstests prüfen die Hypothese, ob eine Stichprobe einer vorgegebenen Verteilungsfunktion entspricht bzw. zu einer ganz bestimmten Klasse vorgegebener Verteilungsfunktionen gehört. Es erfolgt der Vergleich einer empirischen Stichprobe und ihrer Verteilung mit einer theoretisch angenommenen Verteilungsfunktion.

Wir werden im weiteren in Abhängigkeit vom Informationsgehalt der Daten „am besten" geeignete Anpassungstests betrachten. Entsprechend wird in den nachfolgenden Abschnitten die Gliederung von der Art der Urdaten bestimmt.

### 4.2.1 Alternative Daten (Binomialtest/$u$-Test)

Stehen uns nur zwei Ausprägungen eines Merkmals zur Verfügung, dann prüfen wir die Hypothese über die Wahrscheinlichkeit einer dieser Ausprägungen. Die zugrundeliegende Wahrscheinlichkeitsverteilung ist in diesem Fall die Binomialverteilung. Uns steht eine Stichprobe vom Umfang $N$ zur Verfügung. Diese Stichprobe ist aufgeteilt auf die Kategorien 1 und 2 (da alternative Daten), wobei wir folgende Bezeichnungen einführen wollen: