

Univ. Prof. Dr. Willi Wolf

Beschreibende und schließende Statistik

kultur- und
sozialwissenschaften



FernUniversität in Hagen

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

INHALTSVERZEICHNIS

Didaktisch-methodische Hinweise	VII
1. Einige Vorbemerkungen	VII
2. Generelle Zielsetzungen	IX
3. Formaler Aufbau des Textes	XI
4. Mathematische Voraussetzungen, Black-Box-Verfahren, Hilfsmittel und Einsatz der EDV	XI
5. Literaturhinweise	XV

Block I: Einführung in die deskriptive Statistik

Teil I: Eindimensionale Häufigkeitsverteilungen	1
I.A: Zum Einstieg	3
1. Von der Urliste zur Häufigkeitsverteilung	5
2. Klassierung von Meßwerten	10
3. Graphische Darstellungen eindimensionaler Verteilungen	16
3.1 Vorbemerkungen	16
3.2 Kreis- und Rechteckdiagramme	18
3.3 Stab- und Balkendiagramme	19
3.4 Histogramme und Liniendiagramme	20
3.5 Manipulationsmöglichkeiten mit graphischen Darstellungen	22
4. Eine Grundaufgabe der deskriptiven Statistik: Definition und Berechnung von Parametern	26
ME 1: Mathematischer Exkurs: Ordnungsrelation und Intervalle	30
I.B: Einige Grundlagen und Voraussetzungen	31
5. Begriffe und Hypothesen - Der Operationalisierungsprozeß	33
5.1 Vorbemerkungen	33
5.2 Begriffe	34
5.2.1 Einteilung von Begriffen	34
5.2.2 Arten von Objekten	35
5.2.3 Arten von Merkmalen	35
5.2.4 Arten von Merkmalsausprägungen	36
A) Quantitativer Gehalt	36
B) Stetig versus diskret	39
5.2.5 Zusammenfassung und Ausblick	40
5.3 Hypothesen in der empirischen Sozialforschung	41
5.3.1 Arten von Hypothesen	41
5.3.2 Empirisch-wissenschaftliche Hypothesen	42

5.4	Operationalisierung und Operationalisierungsprozeß.....	43
5.4.1	Beispiel	43
5.4.2	Die operationale Definition und die Struktur des Operationalisierungsprozesses.....	46
5.4.3	Überlegungen zur Güte von Operationalisierungsprozessen	48
6.	Messen in den Sozialwissenschaften - Meßniveaus	52
6.1	Vorläufiger Begriff des Messens	52
6.2	Erweiterung des Meßbegriffes	53
6.3	Meßniveaus	64
7.	Das Rechnen mit dem Summenzeichen	68
7.1	Bestandteile der Summenschreibweise	68
7.2	Die Summationsvariable bezieht sich auf eine Meßwerttabelle	70
7.3	Die Summationsvariable stellt eine Rechenvorschrift dar	71
7.4	Ein wichtiger Sonderfall	71
7.5	Regeln zum "Rechnen" mit dem Summenzeichen	72
7.6	Häufige Fehler im Umgang mit dem Summenzeichen	73
7.7	Zwei abschließende Beispiele	73
I.C:	Modelle und Parameter	75
8.	Das h- und das H-Modell	77
8.1	Das h-Modell - die Häufigkeitsverteilung	77
8.1.1	Entwicklung des h-Modells	77
8.1.2	Interpretation und Eigenschaften des h-Modells	79
	A) Grundinterpretation: Häufigkeiten und Meßwertgröße	79
	B) Symmetrische und unsymmetrische Verteilungen	80
	C) Spannweite und Modalwert	81
	D) Typen von Verteilungen	83
	E) Fläche unter der Verteilungskurve	84
	F) Verteilungsschwergewicht	85
8.2	Das Konzept der Summenhäufigkeit	87
8.3	Das H-Modell - die Summenhäufigkeitsverteilung	89
8.3.1	Die Entwicklung des H-Modells	89
8.3.2	Zusammenhänge zwischen h- und H-Modell	92
8.3.3	Weitere Anwendungsmöglichkeiten des H-Modells	93
9.	Lokalisation als zentrale Fragestellung der deskriptiven Statistik	95
9.1	Zielsetzungen von Kap. 9 und 10	95
9.2	Der arithmetische Mittelwert	97
9.2.1	Zielsetzung, Definition und Berechnung	97
9.2.2	Einige Eigenschaften von \bar{x}	99
9.2.3	Auswirkungen der Klassenbildung auf die Exaktheit von arithmetischen Mittelwerten	100
9.3	Der Median	101
9.4	Erweiterung des Mediankonzepts: relative Positionen	105
9.5	Der Modus	107
9.6	Abschließende Bemerkungen zu den Lokalisationsparametern	109

10. Dispersion als zentrale Fragestellung der deskriptiven Statistik	112
10.1 Einleitung	112
10.2 Varianz und Standardabweichung	116
10.2.1 Zielsetzung und Definitionen	116
10.2.2 Interpretationsmöglichkeiten von s^2/s	119
10.2.3 Berechnung von s^2/s	121
10.2.4 Eigenschaften der Varianz	124
10.2.5 Auswirkung der Klassenbildung auf die Exaktheit von s^2/s	124
10.3 Quartilabstand und Häufigkeitsverteilung	125
10.3.1 Quartilabstand QA	125
10.3.2 Die Häufigkeitsverteilung zur Interpretation der Dispersion bei nominalem Meßniveau	125
10.4 Verteilungsbeschreibungen mit Quartildarstellungen	126
10.5 Standardisierung von Meßwerten	129
ME 2: Mathematischer Exkurs: Betragszeichen und Beträge von Zahlen	133
Teil II: Zweidimensionale Häufigkeitsverteilungen	135
II.A: Zum Einstieg	137
11. Zweidimensionale Häufigkeitstabellen und ihre Interpretation	139
11.1 Allgemeine Struktur zweidimensionaler Häufigkeitstabellen.....	139
11.2 Prozentuierung und Interpretation zweidimensionaler Tabellen	141
12. Graphische Darstellungen zweidimensionaler Verteilungen	146
12.1 Graphische Darstellungen bei einer $m \times n$ -Tabelle mit $m, n \geq 3$	146
12.2 Graphische Darstellungen bei einer $m \times n$ -Tabelle mit m und/oder $n = 2$	149
II.B: Modell und Parameter	151
13. Das Punktwolken-Modell	153
13.1 Vorbemerkungen	153
13.2 Entwicklung des Modells	155
13.3 Anwendungen des Punktwolken-Modells	159
13.3.1 Lineare und nichtlineare Korrelation	159
13.3.2 Positive und negative Korrelation	160
13.3.3 Stärke einer Korrelation	161
13.3.4 Prognosen mit Hilfe von Korrelationen	164
14. Korrelation als Fragestellung der deskriptiven Statistik	166
14.1 Vorbemerkungen	166
14.2 Definition des Produkt-Moment-Korrelationskoeffizienten r	167
14.3 Berechnungsformel für r und Berechnungsbeispiel	173
14.4 Zur Interpretation von Korrelationen	175
14.5 Das Konzept der partiellen Korrelation	178
14.6 Erweiterungen des Korrelationskonzeptes	179
15. Regression als Fragestellung der deskriptiven Statistik	180
15.1 Vorbemerkungen	180
15.2 Die Gleichungen der Regressionsgeraden	181

15.3	Berechnungsformeln und Berechnungsbeispiel	185
15.4	Erste Interpretationen und Eigenschaften	186
15.5	Erweiterungen des Regressionskonzeptes	188
16.	Zusammenhänge zwischen Korrelation, Regression und Varianz	189
16.1	Vorbemerkungen	189
16.2	Varianz "um" die Regressionsgerade und Varianz "auf" der Regressionsgeraden ...	190
16.3	Zusammenhänge zwischen Korrelation, Regression und Varianz	193
ME 3:	Mathematischer Exkurs: Geraden und Geradengleichungen	197

Das Zwischenspiel: Ausflüge in Wahrscheinlichkeitsrechnung und schließende Statistik

17.	Einige Grundlagen aus der Wahrscheinlichkeitsrechnung	203
17.1	Vorbemerkungen	203
17.2	Unser Alltagsverständnis und erste Begriffe	204
17.3	Basisbeispiel	207
17.4	Berechnung von Wahrscheinlichkeiten - Sätze und Formeln	209
17.4.1	Geometrische Definition der Wahrscheinlichkeit und Wahrscheinlichkeiten bei einfachen Ereignissen	209
17.4.2	Ereignisverknüpfungen und ihre Wahrscheinlichkeiten	212
17.4.3	Ein komplizierter Fall	220
18.	Ausbau der Wahrscheinlichkeitsrechnung und Übergänge zur analytischen Statistik	223
18.1	Vorbemerkungen	223
18.2	Ausbau der Wahrscheinlichkeitsrechnung	224
18.2.1	Die Zufallsvariable X	224
18.2.2	Die Wahrscheinlichkeitsdichte $f(x)$	225
18.2.3	Die Verteilungsfunktion $F(x)$	230
18.3	Übergänge zur analytischen Statistik und Verteilungsmodelle	235
18.3.1	Das Δ -Beispiel	236
18.3.2	Grundgesamtheit und Stichprobe	237
18.3.3	Grundaufgabe der analytischen Statistik: Schlüsse von Stichproben auf Grundgesamtheiten	239
18.3.4	Verteilungsmodelle	241
19.	Die Normalverteilung	243
19.1	Die Funktionsgleichung für die Wahrscheinlichkeitsdichte einer Normalverteilung	245
19.2	Wahrscheinlichkeitsdichte und Verteilungsfunktion der Standard-Normalverteilung	247
19.3	Wahrscheinlichkeitsdichte und Verteilungsfunktion einer beliebigen Normalverteilung $N(\mu, \sigma)$	252
19.4	Zwei Regelgruppen zur Normalverteilung	256
19.5	Die Bedeutung der Normalverteilung	259
19.5.1	Welche Variablen sind normalverteilt?	260
19.5.2	Die Normalverteilung als Grenzverteilung	262
19.6	Verteilungsmodelle für Stichprobenparameter	262

20. t-, χ^2- und F-Verteilung	270
20.1 Vorbemerkungen	270
20.2 Die t-Verteilung	274
20.3 Die χ^2 -Verteilung	276
20.4 Die F-Verteilung	279
ME 4: Mathematischer Exkurs: Symmetrische Funktionen	282

Block II: Grundlagen der analytischen Statistik und statistische Tests

Teil III: Grundlagen der analytischen Statistik 285

21. Grundbegriffe der analytischen Statistik	287
21.1 Vorbemerkungen	287
21.2 Wahrheit und Falschheit bei All-Aussagen und bei statistischen Aussagen	288
21.3 Das RTP-Beispiel	291
21.4 Statistische Hypothesen	295
21.5 Die "Logik" der statistischen Entscheidung	300
21.6 Die Irrtumswahrscheinlichkeiten α und β	304
21.7 Der statistische Test	311
22. Signifikanzprüfungen und das konservative Modell der statistischen Entscheidung	313
22.1 α und β bei einer zusammengesetzten Alternativhypothese	313
22.2 Signifikanzprüfungen und Signifikanzniveaus	319
22.3 Konservatives Vorgehen und das K-Modell der statistischen Entscheidung	321
22.4 Statistische Signifikanz durch Vergrößerung des Stichprobenumfangs	323
23. Ausblick: Das Fundamentalmodell der statistischen Entscheidung	326
23.1 Vorbemerkungen	326
23.2 Praktische Bedeutsamkeit und Effektgröße	327
23.3 Zusammenhänge zwischen α , β , n und der Effektgröße E	331
23.4 Das Fundamentalmodell der statistischen Entscheidung	335

Teil IV: Auswahl und Anwendung statistischer Tests 337

24. Kriterien für die Auswahl statistischer Tests	339
24.1 Vorbemerkungen	339
24.2 Klassifizierung der Verfahren der analytischen Statistik	341
24.3 Kriterien für die Testauswahl	343
24.3.1 Anzahl der Stichproben	343
24.3.2 Die Art der zu vergleichenden Stichproben	349
24.3.3 Das Meßniveau der (abhängigen) Variablen	350
24.4 Zum Aufbau der Testbeschreibungen	350
A. Freizeit in R: Fragestellungen - Daten - Testauswahl	351
B. Ausgewählte χ^2-Tests	357
1. Einführende Informationen	357
2. Der χ^2 -Test für eine 2x2-Tabelle	362

3.	Der χ^2 -Test für eine $m \times 2$ - bzw. $2 \times n$ -Tabelle	364
4.	Der χ^2 -Test für eine $m \times n$ -Tabelle	366
5.	Der χ^2 -Test nach McNemar für zwei abhängige Stichproben	369
6.	Der χ^2 -Test für die Güte der Anpassung	372
7.	Kontingenzkoeffizienten	377
C.	Ausgewählte Rangsummentests	379
1.	Einführende Informationen	379
2.	Berechnung von Rangsummen	380
3.	Rangsummentest nach Wilcoxon für $k = 2$ unabhängige Stichproben	382
4.	Rangsummentest nach Kruskal-Wallis für $k \geq 3$ unabhängige Stichproben: H-Test .	386
5.	Rangsummentest nach Dunn für Kontrastvergleiche bei $k \geq 3$ unabhängige Stichproben	390
D.	t-Test und F-Test	395
1.	Einführende Informationen	395
2.	t-Test für den Vergleich zweier Mittelwerte	396
3.	F-Test für den Vergleich zweier Varianzen	399

Aufgabenblätter	403
------------------------------	------------

Lösungen	431
-----------------------	------------

Tabellenanhang	463
-----------------------------	------------

VT 1	Wahrscheinlichkeitsdichte der standardisierten Normalverteilung	464
VT 2	Verteilungsfunktion der standardisierten Normalverteilung	466
VT 3	Verteilungstabelle der t-Verteilung	468
VT 4	Verteilungsfunktion der χ^2 -Verteilung	470
VT 5	Verteilungsfunktion der F-Verteilung	475

Didaktisch-methodische Hinweise

1. Einige Vorbemerkungen

„So eine umfangreiche Einführung in die Statistik!“ mögen Sie vielleicht gedacht haben, als Sie diesen Band zum ersten Mal in die Hand nahmen. Und vielleicht sind dabei auch noch einige weitere Vorbehalte gegenüber der Statistik wieder aus der Erinnerung aufgetaucht. Aus diesem Grund sollen vorab in drei Vorbemerkungen solche Vorbehalte angesprochen werden. Ich hoffe, auf diese Weise eventuell vorhandene Befürchtungen zerstreuen, zumindest aber mildern zu können.

Die **erste Vorbemerkung** zielt auf den Umfang des Textes ab. Der Umfang hängt damit zusammen, daß es sich nicht um ein knapp gefaßtes Lehrbuch, sondern um ein Arbeitsbuch handelt, das für das **Selbststudium** geeignet ist. Zu den Elementen, die dieses ermöglichen sollen, gehören beispielsweise:

- eine durchaus beabsichtigte Textredundanz
- sehr viele Abbildungen zur Erläuterung der behandelten Sachverhalte
- viele Beispiele im Text
- Übungsaufgaben mit ihren vollständigen Lösungen

Auch die Verwendung der Marginalien „kostet“ viele zusätzliche Seiten.

Der inhaltliche Umfang des Kurses entspricht in etwa den Inhalten einer einsemestrigen zweistündigen Lehrveranstaltung im (längeren) Wintersemester oder, anders formuliert, er entspricht etwa 150 Textseiten eines konventionellen Lehrbuches. Also, lassen Sie sich bitte durch den Umfang des Buches nicht abschrecken.

Im gleichen Kontext soll in einer **zweiten Vorbemerkung** eine weitere, immer wieder zu hörende Befürchtung angesprochen werden: Statistik habe sehr viel mit Mathematik zu tun, sie bestehe aus meist unverständlichen Formeln und sei, ganz allgemein, sehr schwierig.

Zwar ist zutreffend, daß auch diese „Einführung in die Statistik“ nicht auf mathematische Grundlagen verzichten kann. Aber diese Grundlagen sind, wie unten in 4. näher ausgeführt wird, sehr bescheiden und beschränken sich auf elementare Kenntnisse, wie sie beispielsweise bis zur 10. Klasse im allgemeinbildenden Schulwesen erworben werden. Die Kenntnisse, die Sie aus Ihrer Schulzeit mitbringen, sind also sicher für unsere „Einführung“ ausreichend.

Aber es geht in den obigen Vorbehalten ja noch um ein weiteres Problem: das Verstehen und Anwenden der in mathematische Formeln gefaßten Quantifizierungen. Dieses Problem ist ein gravierendes Problem, spiegelt es doch den Sachverhalt, daß oft - selbst kluge - Menschen, aus welchen Gründen auch immer, nicht gelernt haben, alltägliche Probleme auch unter einem quantitativen Gesichtspunkt zu betrachten, sie in häufig sehr einfache mathematische Beziehungen und Formeln abzubilden und damit „rechenfähig“ zu machen.

statistisches Argumentieren

Im Rahmen unserer Einführung wird versucht, dieser Schwierigkeit durch „**statistisches Argumentieren**“ zu begegnen. Damit ist Folgendes gemeint: Ausgehend von Situationen, die vom Alltag her als „vertraut“ oder vom Verlauf der „Einführung“ her als „inzwischen vertraut“ eingeschätzt werden, wird das jeweilige Problem zunächst in sprachlicher Form entwickelt und mit Plausibilitätsüberlegungen zu einer vorläufigen Lösung geführt. In diesem Zusammenhang spielen auch viele Abbildungen eine wichtige Rolle. Dann erst werden die gewonnenen Ergebnisse, oft schrittweise, in die mathematische Sprache, in Formeln übersetzt.

Mit einem solchen Vorgehen sind die Hoffnungen verknüpft, den Weg zu den erforderlichen Quantifizierungen leichter begehbar zu machen und Vorbehalte gegenüber einer in Formeln gefaßten mathematischen Sprache abzumildern.

Eine **letzte Vorbemerkung** zielt in eine andere Richtung. Die Statistik ist eine Disziplin ohne eigenen Inhaltsbereich, die in vielen anderen Disziplinen, wie beispielsweise den Sozial-, den Sprach- oder den historischen Wissenschaften, angewendet wird. Das Ausmaß dieser Anwendungen außerhalb der Naturwissenschaften hat in den letzten Jahrzehnten deutlich zugenommen.

Statistik ist also eine **formale Wissenschaft**. Diese Eigenschaft hat Konsequenzen für die erforderlichen **Lehr-Lern-Prozesse**. Man könnte die Statistik durchaus auf ihren formalen Kern reduzieren und diesen in einer streng systematisierten Form darbieten. Damit würden sich die oben genannten 150 Seiten eines konventionellen Lehrbuches um deutlich über die Hälfte reduzieren lassen. Aber was hier an Kürze gewonnen würde, ginge im Regelfall an Verständlichkeit verloren. Außerdem müßten die Studierenden alle Transferleistungen im Hinblick auf Anwendungen in dem sie interessierenden Bereich allein erbringen: eine Aufgabe, die kaum zu lösen ist. Um diese Schwierigkeiten zu vermeiden, sind Einführungen in die Statistik hinsichtlich des erwähnten statistischen Argumentierens sowie der ausgewählten Beispiele meist **fachspezifisch** konzipiert. Im Falle unserer „Einführung“ sind unter diesem Aspekt die Sozialwissenschaften, vor allem aber die Erziehungswissenschaft angesprochen.

Oben wurden die Lehr-Lern-Prozesse angesprochen. Ganz ohne Zweifel lernen Studierende der Sozialwissenschaften Statistik dann am intensivsten, sehen am ehesten den Nutzen statistischer Quantifizierungen ein, wenn sie eigene empirische Forschungen durchführen. Überträgt man diese Aussage auf das Studium, so wäre ein **Studienprojekt** die für die Statistik beste Aneignungsform. Das ist zwar richtig, doch setzt ein sinnvolles Studienprojekt den permanenten Kontakt, die persönliche Interaktion zwischen Lernenden und Lehrenden voraus. Das aber ist eine Voraussetzung, die sich in einem Buch zum Selbststudium nicht realisieren läßt.

Aus diesem Grunde ist unsere „Einführung“ **systematisch** aufgebaut. Dieser Aufbau spricht aber nicht dagegen, daß sie auch in Studienprojekten als Basislektüre verwendet werden kann. Der systematische Aufbau hat jedoch bei den gewählten **Illustrationsbeispielen** folgende Konsequenz: Sie sind meist konstruiert und fiktiv. Sie sprechen zwar implizit ein Problem an, sind aber so aus dem Kontext herausgelöst und inhaltlich reduziert worden, daß sie nur für die gerade stattfindende Argumentation funktional sind. Deswegen ist es auch hinreichend, wenn zur Vermeidung von Rechenaufwand mit kleinen und überschaubaren Zahlen gearbeitet wird. Gleiches gilt überwiegend auch für die Aufgaben.

Art der Illustrationsbeispiele

Nach diesen allgemeinen Vorbemerkungen wenden wir uns nun den didaktisch-methodischen Hinweisen im engeren Sinne zu, die über

2. Generelle Zielsetzungen
3. Formaler Aufbau des Textes
4. Mathematische Voraussetzungen, Black-Box-Verfahren, Hilfsmittel und Einsatz der EDV
5. Literaturhinweise

informieren.

2. Generelle Zielsetzungen

Wer eine Einführung in Methoden und Konzepte der quantitativen empirischen Sozialforschung studiert, müßte anschließend zumindest in der Lage sein, vorliegende, sein Studienfach betreffende empirische Untersuchungen „kritisch“ lesen und deren Ergebnisse „kritisch“ anwenden zu können. Dazu müßten Kenntnisse in den folgenden Bereichen erworben werden:

- a) **Grundkonzepte der empirischen sozialwissenschaftlichen Forschung**
wie beispielsweise Operationalisierung, Messen, Repräsentativität, Validität, Reliabilität
- b) **Erhebungstechniken oder Verfahren zur Datensammlung**
wie beispielsweise Beobachtung, Befragung/Interview, Inhaltsanalyse

**Curriculum:
Einführung in die
empirische
Sozialforschung**

c) Verfahren der deskriptiven (beschreibenden) Statistik

wie beispielsweise Berechnung von Durchschnittswerten, Dispersionsparametern oder Korrelationskoeffizienten, Anwendung von Varianz- und Regressionskonzept

d) Verfahren der analytischen (schließenden) Statistik

wie beispielsweise Auswahl und Anwendung statistischer Tests, Beurteilung der statistischen und der praktischen Bedeutsamkeit von Ergebnissen

Solche Kenntnisse sollten vor allem folgenden Zielsetzungen dienen:

- Beurteilung der Güte von empirischen Untersuchungen
- Beurteilung der Relevanz der Ergebnisse empirischer Untersuchungen für bestimmte Verwendungszwecke
- Steigerung der Kommunikationsfähigkeit, beispielsweise mit solchen Personen, die empirische Untersuchungen durchführen
- Bereitstellung notwendiger Voraussetzungen für jene Studierenden, die selbst empirische Untersuchungen durchführen wollen

Dieses Programm werden wir im Rahmen unserer „Einführung“ nur in bestimmten Teilen abdecken können, wobei vor allem die Punkte c) und d) im Mittelpunkt stehen. Eine Kurzcharakterisierung der insgesamt dreiteiligen Gliederung läßt sich so beschreiben (s. hierzu auch das Inhaltsverzeichnis):

**Beschreibung der
Grobstruktur der
„Einführung“**

- In einem „**Block I**“ geht es im Schwerpunkt um grundlegende Verfahren und Konzepte der deskriptiven Statistik, die durch notwendige Überlegungen zur Operationalisierung und zum Messen in den Sozialwissenschaften ergänzt werden. Zu den statistischen Konzepten gehören Lokalisation, Dispersion, Korrelation und Regression.
- Im zweiten, „**Zwischenspiel**“ genannten Teil werden wichtige Ergebnisse aus der Wahrscheinlichkeitsrechnung und Statistik aufbereitet und systematisch zusammengefaßt. Dieser Teil wird sich als relativ leicht erweisen, da wir hier im Schwerpunkt mit Plausibilitätsüberlegungen arbeiten und außerdem auf Kenntnisse aus Block I zurückgreifen können.
- Im abschließenden „**Block II**“ steht dann die analytische Statistik im Zentrum, wobei wir uns wieder auf grundlegende Konzepte, wie statistische Hypothesen und statistische Entscheidung beschränken. Dieser Abschnitt schließt mit der exemplarischen Vorstellung einiger statistischer Tests.

Zu unserer „Einführung“ gehören außerdem noch **Aufgabenblätter** und die vollständigen **Lösungen** zu den Aufgaben.

Soweit zu den inhaltlichen Zielen der „Einführung“. Detailliertere Informationen über die Inhalte finden sich unter der Überschrift „Kurzcharakteristik“ am Anfang der einzelnen Teile I.A, I.B usw.

3. Formaler Aufbau des Textes

Die „Einführung“ besteht aus vier durch unterschiedliche Kopfzeilen gekennzeichneten Komponenten.

a) Textteil (schwarze Kopfzeile)

Der Textteil ist in 24 Kapitel gegliedert. Die einzelnen Kapitel beginnen jeweils mit einer Auflistung der Lernziele, die in dem Kapitel erreicht werden sollen. In wichtigen Fällen wird vorab auch noch auf Voraussetzungen verwiesen, die man sich dann zweckmäßigerweise vorher kurz ansehen sollte.

Ob die jeweils angegebenen Lernziele erreicht wurden, können Sie beurteilen, wenn Sie sich im Anschluß an das Kapitel den Lernzielkatalog noch einmal ansehen. Eine weitere und wichtigere Überprüfungsmöglichkeit bilden die jeweils zu einem Kapitel gehörenden Aufgabenblätter.

Im Textteil werden **Marginalien** (Randauszeichnungen) verwendet, die ein späteres Wiederfinden bestimmter Passagen erleichtern sollen. Der hier noch reichlich vorhandene Platz sollte von Ihnen für Ihre Anmerkungen genutzt werden.

**Textteil
(schwarze Kopfzeile)**

**Dies ist eine
Marginalie!**

b) Aufgabenblätter AB (rote Kopfzeile)

Zu allen wichtigen Aspekten der Einführung gibt es ein oder mehrere Aufgabenblätter, die sich im Regelfall auf die vorher diskutierten Sachverhalte beschränken und mit denen Sie Ihren Lernerfolg selbst kontrollieren können. In einigen Ausnahmefällen vermitteln sie zusätzlich die eine oder andere Vertiefung oder Weiterführung.

**Aufgabenblätter
(rote Kopfzeile)**

c) Aufgabenlösungen LÖ (grüne Kopfzeile)

Aufgaben mit der Funktion der Selbstkontrolle sind immer nur so gut wie die Lösungen, anhand derer das eigene Arbeiten kontrolliert werden kann. Aus diesem Grund sind die Lösungen weitgehend vollständig im Lösungsteil wiedergeben, den Sie aber selbstverständlich erst **nach** Ihren Lösungsversuchen benutzen sollten.

**Lösungen
(grüne Kopfzeile)**

d) Verteilungstabellen (VT) (braune Kopfzeile)

Hier sind die erforderlichen statistischen Tabellen zusammengestellt.

**Verteilungstabellen
(braune Kopfzeile)**

e) Testbeschreibungen (blaue Kopfzeile)

Hier werden drei Gruppen von statistischen Tests vorgestellt.

**Testbeschreibungen
(blaue Kopfzeile)**

Zusätzlich zu einer durchgehenden **Seitennumerierung** sind die einzelnen Kapitel, die Aufgabenblätter und die Lösungen noch in der Kopfzeile genannt und „intern“ durchnummeriert.

**Hinweise zur
Seitennumerierung**

4. Mathematische Voraussetzungen, Black-Box-Verfahren, Hilfsmittel und Einsatz der EDV

Wie schon oben erwähnt, beschränken sich die mathematischen Voraussetzungen auf elementare Kenntnisse, die das Niveau der zehnten Klasse im allge-

meinbildenden Schulwesen nicht überschreiten. Diese Feststellung lässt sich in folgender Weise konkretisieren:

Konkretisierung der mathematischen Voraussetzungen

a) Kenntnisse der vier Grundrechenarten Addition, Subtraktion, Multiplikation und Division im Bereich der reellen Zahlen

b) Kenntnisse im Umgang mit Quadratwurzeln und Potenzen, beispielsweise:

$$\sqrt{32,5} = 5,70 \text{ (Ermittlung mit dem Taschenrechner oder einer Tabelle)}$$

$$\sqrt{0,85} = 0,92$$

$$5^2 = 5 \cdot 5 = 25$$

$$a^3 = a \cdot a \cdot a \quad (\text{und nicht } = 3a)$$

c) Elementare Kenntnisse algebraischer Umformungen und linearer Gleichungen, beispielsweise:

$$\frac{ax + ay}{2a} = \frac{a(x + y)}{2a} = \frac{x + y}{2}$$

$$(u + v)^2 = (u + v)(u + v) = u^2 + 2uv + v^2$$

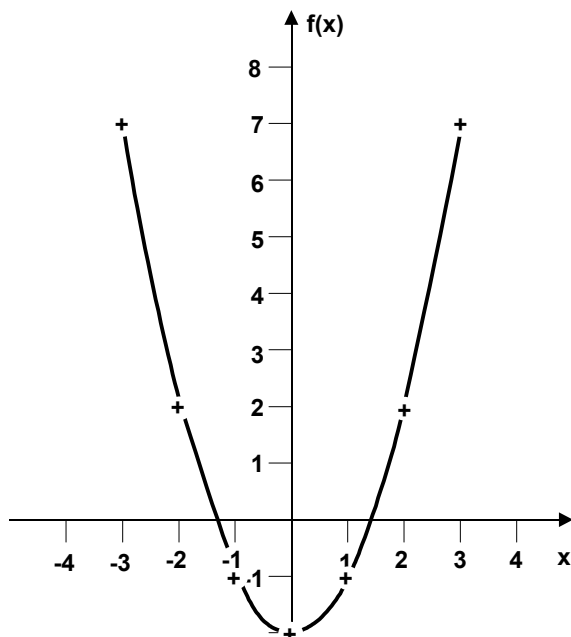
$$7x + 3 = 2x \quad \Leftrightarrow \quad 5x = -3 \quad \Leftrightarrow \quad x = -\frac{3}{5}$$

d) Kenntnis des rechtwinkligen Koordinatensystems und Fertigkeit, die Bilder einfacher Funktionen zeichnen zu können

Soll beispielsweise das Bild von $f(x) = x^2 - 2$ im Intervall von $x = -3$ bis $x = +3$ gezeichnet werden, so führt das über die Wertetafel

x	-3	-2	-1	0	1	2	3
f(x)	7	2	-1	-2	-1	2	7

zu folgendem Funktionsbild:

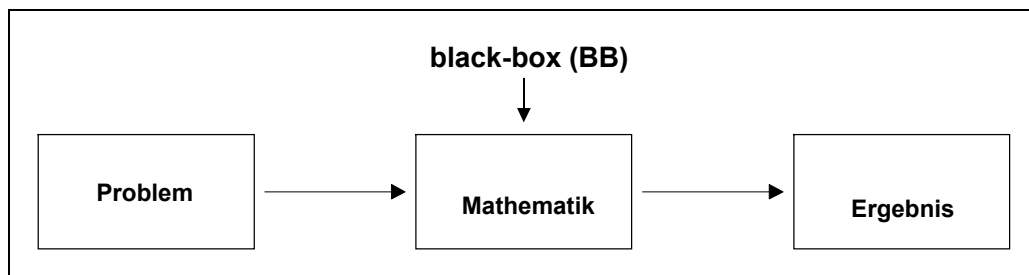


Weitere wichtige mathematische Grundlagen, wie beispielsweise das Rechnen mit dem Summenzeichen oder die Beschreibung von Zahlenintervallen, werden in eigenen Kapiteln oder in kurzen mathematischen Exkursen erläutert.

Ab und zu erfolgen Hinweise auf mathematische Sachverhalte, die erst ab Klasse 11 vermittelt werden. Diese Hinweise, die jeweils am Ende von Plausibilitätsüberlegungen stehen, sollen Studierende mit fundierteren Mathematikkenntnissen darüber informieren, wie das gerade angesprochene Problem mathematisch angegangen werden kann.

Beschränkt man die mathematischen Voraussetzungen auf das oben gekennzeichnete Niveau, so ergibt sich folgende Situation: Mit fundierteren Voraussetzungen ließen sich die entsprechenden Sachverhalte beweisen. Dieser Ansatz kann aber nicht Zielsetzung einer allgemeinen Einführung sein. Die Konsequenz besteht darin, daß wichtige Sachverhalte nicht bewiesen werden, sondern nur mitgeteilt werden können.

Sachverhalte dieser Art müssen, da mit ihnen im weiteren Verlaufe der „Einführung“ gearbeitet wird, zur Kenntnis genommen werden und - wie die Worte einer fremden Sprache - gelernt werden und angewendet werden können. Will man diesen Vorgang beschreiben, so kann das mit Hilfe des folgenden **Black-Box-Modells** geschehen:



Die Mathematik ist der mittlere „schwarze Kasten“, von dem wir zwar nicht wissen, wie er genau funktioniert, der uns aber dennoch relevante Ergebnisse liefert. Wenn immer wir das Black-Box-Verfahren anwenden, wird es am Rande durch die Marginalie **BB** gekennzeichnet. Leserinnen und Leser wissen dann, daß ihnen ein Ergebnis mitgeteilt wird, mit dem weitergearbeitet wird, dessen Herleitung aber die vorhandenen Mathematikkenntnisse vermutlich übersteigen würde.



Zwar ist dieses Vorgehen bedauerlich, es kennzeichnet dennoch eine alltägliche Situation. Jeder von Ihnen wird, so nehme ich an, die Fläche eines Kreises nach der Formel „ $F = r^2\pi$ “ berechnen können. Aber, „Hand aufs Herz“, wer kann diese Formel noch herleiten?

Wir verwenden das Black-Box-Verfahren aber noch in einer zweiten Situation. Es kommt immer wieder vor, daß auch elementare algebraische Umformungen

sehr aufwendig sein können, ohne daß damit ein Erkenntnisgewinn verbunden wäre. Auch in solchen Fällen werden wir oft die Ergebnisse nur mitteilen.

**Hilfsmittel:
Taschenrechner**

Welche **Hilfsmittel** sind erforderlich? Zwar wird der Rechenaufwand möglichst gering gehalten, dennoch vermindert ein einfacher **Taschenrechner**, mit dem man auch Quadratwurzeln berechnen kann und der über einen Zwischenspeicher und eine Prozentautomatik verfügt, den Arbeitsaufwand oft spürbar.

Rolle der EDV

Welche Rolle spielt die **EDV**? Es ist nicht die Absicht dieser „Einführung“, die Studierenden zu Experten für die Durchführung statistischer Tests zu machen: erstens gibt es zu viele Tests, und zweitens übernimmt diese Aufgabe immer mehr die EDV. Deswegen werden auch im Block II nur zwei Tests exemplarisch vorgestellt. Leserinnen und Leser sollten aber anschließend bei einfachen Problemen in der Lage sein, mit Hilfe weiterführender Literatur den für ein Problem geeigneten statistischen Test auswählen und auch durchführen zu können.

Allerdings erfolgen solche Durchführungen immer seltener mit einem Taschenrechner 'per Hand', sondern mit Hilfe der **EDV**. In den letzten Jahren sind sehr komfortable Statistikprogramme für den PC auf den Markt gekommen, mit denen man (ggf. zu Hause mit einem eigenen PC) die Verfahren der deskriptiven und der analytischen Statistik (einschließlich graphischer Darstellungen) ohne jeglichen Rechenaufwand durchführen kann. Man muß nur noch lernen, solche Programme zu 'steuern'; der Zeitaufwand hierfür ist klein im Vergleich zu dem Zeitaufwand, den man bei einer Auswertung ohne EDV hätte. Ein Problem besteht allerdings: Diese Programme machen alles, was ihnen befohlen wird, und sei es noch so unsinnig. So berechnen sie beispielsweise auf Auftrag kommentarlos den Mittelwert von qualitativen Daten. Leserinnen und Leser können in dieser „Einführung“ die theoretischen Kenntnisse erwerben, die für einen sinnvollen Umgang mit solchen Auswertungsprogrammen erforderlich sind.

Oben war von „einfachen“ Problemen die Rede. Darüber hinaus gibt es aber auch Anwendungsfälle, die den Einsatz komplexer statistischer Auswertungsmethoden erfordern und die ebenfalls mit der EDV bearbeitet werden. In solche 'multivariate Verfahren', zu denen beispielsweise die Faktorenanalyse oder die Clusteranalyse gehören, muß man sich im Bedarfsfall jeweils gesondert mit Hilfe weiterführender oder spezieller Literatur einarbeiten. Ziel unserer „Einführung“ ist es nun auch, Grundlagen für solche weiterführenden Beschäftigungen zu schaffen.