Kapitel 1

Einleitung

1.1 Hauptbegriffe

Statistik begegnet uns überall im täglichen Leben, wo Massendaten komprimiert werden.

Mathematische Statistik (auch: schließende Statistik, induktive Statistik, Inferenzstatistik oder inferentielle Statistik) ist das Teilgebiet der Statistik, das die Methoden und Verfahren der Statistik mit mathematischen Mitteln analysiert bzw. mit ihrer Hilfe erst begründet. Die mathematische Grundlage der Mathematischen Statistik ist die *Wahrscheinlichkeitstheorie*: Typischerweise werden die vorhandenen Daten einer Stichprobe (d.h. Beobachtungen, Messungen) als Realisierungen von stochastisch unabhängigen Zufallsvariablen interpretiert, so dass wahrscheinlichkeitstheoretische Methoden zur Untersuchung des stochastischen Verhaltens der Beobachtungen anwendbar sind. Die *Stochastik* fasst als Oberbegriff die Gebiete Wahrscheinlichkeitstheorie und Mathematische Statistik zusammen[1].

Versuchsplanung (Design of Experiments) ist eine Anwendung der Mathematischen Statistik (man sagt daher auch *statistische* Versuchsplanung), die – beispielsweise bei der industriellen Entwicklung von neuen Produkten und Fertigungsprozessen – helfen kann, mit möglichst geringem Aufwand reproduzierbare Ergebnisse zu erhalten.

Die Themen dieses Buchs sind Wahrscheinlichkeitstheorie, Mathematische Statistik und Versuchsplanung, aber nicht Statistische Physik. Letztere ist eine fundamentale physikalische Theorie, mit deren Hilfe u.a. Gesetze der Thermodynamik abgeleitet und begründet werden.

Die *mathematische Statistik* verwendet für manche grundlegenden Begriffe der Naturwissenschaft ihre *eigenen Namen*. Zum Beispiel benutzt der Statistiker für das naturwissenschaftliche „Messen (Beobachten)“ den Terminus „Ziehen einer Stichprobe“, oder er verwendet das Wort „Schätzen“ für das naturwissenschaftliche „Be-

[1] Das Wort *Stochastik* stammt aus der Frühzeit der Wahrscheinlichkeitsrechnung (einer der Bernoullis hat es geprägt [Swoboda 1971, Seite 174]). Es wurde nach dem Zweiten Weltkrieg zum wissenschaftlichen Modewort. Günter Menges definiert in seinem „Grundriß der Statistik, 1“ lapidar: „Stochastik ist alles, was auf der Wahrscheinlichkeitsrechnung aufgebaut ist oder sonstwie mit ihr zu tun hat“ [Menges 1968]

Angewandte Stochastik und Versuchsplanung in den Natur- und Ingenieurwissenschaften: mit Beispielen in R und SAS, First Edition. Christhard Schmid.
© 2023 WILEY-VCH GmbH. Published 2023 by WILEY-VCH GmbH.

COPYRIGHTED MATERIAL

stimmen (determine)“. Statistisches „Schätzen“ hat also nichts zu tun mit dessen umgangssprachlichen Gebrauch im Sinn von „Vermuten“. Was schließlich der Naturwissenschaftler mit „(statistischem) Messfehler“ oder „Rauschen“ bezeichnet, definiert die Statistik quantitativ mit Begriffen wie Konfidenzbereich, Konfidenzintervall, Intervallschätzwert, Reststreuung, innere Varianz.

1.2 Zielgruppen

Das Buch wendet sich an Naturwissenschaftler und Naturwissenschaftlerinnen und an Ingenieure und Ingenieurinnen in der Industrie.

1.2.1 Einsteiger/innen

Für **R-Einsteiger/innen** mit Windows-PC stehen 95 R-Stand-Alone-Beispiele bereit, die unabhängig und größtenteils isoliert voneinander lauffähig sind. (Lediglich das Programm `R_42.R` wird von einigen anderen Programmen eingebunden und verwendet). Steigen Sie direkt ein mit „learning-by-doing“:

- Installieren Sie RStudio (einmalig) – eine gute, gängige Oberfläche für interaktives R. Installationsanleitungen und Anleitungen zur Bedienung von RStudio finden Sie im Internet, z.B. `https://www.youtube.com/watch?v=X_Mxya2Fis0` und `https://www.youtube.com/watch?v=tyvEHQszZJs`.
- Kopieren Sie das R-Skript-Verzeichnis gemäß der Anleitung von Abschnitt 1.6, Seite 12.
- Entscheiden Sie sich für ein attraktives Beispiel aus dem Lehrbuch, etwa in Abschnitt 7.2.3.1, auf S. 274.
- Öffnen Sie das entsprechende Programm `R_69.R` mit R-Studio.
- Starten Sie das Skript zeilenweise mit Ctrl+Enter (der Cursor wandert nach unten).
- Stoppen Sie bei `MEPlot(Fit)`
- Das Ergebnis sollte ungefähr so aussehen wie der Screenshot 1.1 auf Seite 3.

Für **Fortgeschrittene** mit Vorwissen in Höherer Mathematik (d.h. Vektoren und Matrizen, Differential- und Integralrechnung) bietet sich die Chance zu einem vertieften Verständnis der Statistik in den Kapiteln 4 bis 6.

Das Kapitel 4 (Stichprobenanalyse) erläutert Grundbegriffe der Mathematischen Statistik wie Statistische Messfehler, Konfidenzbereich in Abhängigkeit vom Stichprobenumfang, und liefert eine Einführung in Hypothesentests wie t-Test und andere, inklusive der Berechnung der Teststärke (Power-Analyse).

Kapitel 5 behandelt die Grundlagen der Regressionsanalyse: Annahmen, Schätzung, Residuen, Quadratsummen, Verteilungseigenschaften, Hypothesentests, Konfidenzintervalle, Prognoseintervalle, Modelldiagnose, Nichtlineare Regression.

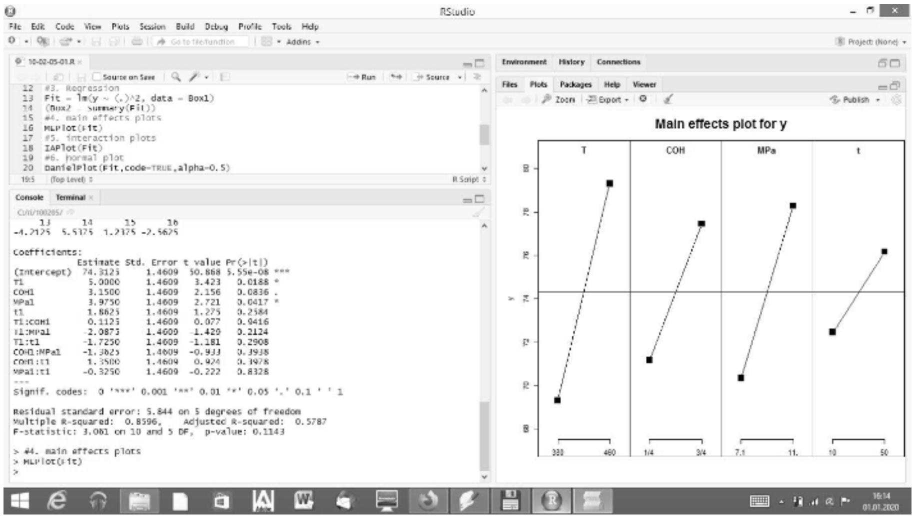

Abb. 1.1: Screenshot nach Durchführung eines Programms in R-Studio

Kapitel 6 ist der Varianzanalyse mit festen Effekten gewidmet. Die Themen sind einfache Varianzanalyse, zweifache Varianzanalyse mit einfacher und mehrfacher Besetzung, Lateinische und Griechisch-Lateinische Quadrate, Kovarianzanalyse.

1.2.2 Physiker/innen

Fortgeschrittene schließlich, die über Höhere Mathematik hinaus mit der **Dirac'schen Deltafunktion** in Berührung gekommen sind, finden im Lehrbuch die Grundlagen der Wahrscheinlichkeitstheorie in knapper Form dargestellt, in den Kapiteln 2 und 3:

Das Kapitel 2 beginnt mit einer kurzen Einführung in die Wahrscheinlichkeitstheorie (Auszug aus dem Vorlesungsmanuskript Theoretische Physik IV, „Thermodynamik und Statistische Physik“ von Prof. Dr. Friederike Schmid, Universität Mainz). Es wird ergänzt durch anwendungsbezogene Themen wie Kombinatorik, Quantile, Symmetrierelationen, quadratische Formen.

Das Kapitel 3 präsentiert acht eindimensionale diskrete Verteilungen (hypergeometrische Verteilung, Binomialverteilung, Poisson-Verteilung, geometrische Verteilung, negative Binomialverteilung u.a.), zehn eindimensionale kontinuierliche Verteilungen (Gauß-Verteilung, Cauchy-Verteilung, Chi-Quadrat-Verteilung, nichtzentrale Chi-Quadrat-Verteilung, Student-Verteilung, nichtzentrale Student-Verteilung, F-Verteilung, nichtzentrale F-Verteilung, Exponentialverteilung und Weibull-Verteilung) und zwei mehrdimensionale kontinuierliche Verteilungen (n-dimensionale und zwei-dimensionale Gauß-Verteilung).

Die Deltafunktion, ein von dem britischen Quantenphysiker Paul Dirac (1902-1984) erfundenes Werkzeug, dient hier dem Zweck, Statistiker von hochabstrakter

Mathematik (Maßtheorie) zu verschonen. Sie ist außergewöhnlich effektiv und elegant:

Diskrete und kontinuierliche Verteilungsdichten verschmelzen durch die δ-Funktion zu einer Einheit. Man könnte z.B. die Tabellen 3.2 und 3.10 auf den Seiten 57 und 84 zu einer einzigen Tabelle zusammenfassen (untereinander schreiben).

Chiquadrat-, Student-, u.a. Verteilungsdichten werden mit der δ-Funktion elegant hergeleitet. Siehe (3.53), (3.73) u.a.

Die Verteilung einer Funktion von Zufallsvariablen ergibt sich mit Hilfe der δ-Funktion wie von selbst (Abschnitt 2.5.1).

Für die Verallgemeinerung auf mehrere Zufallsvariable gilt Ähnliches (Abschnitt 2.5.3).

Die Wahrscheinlichkeitsdichte ist gleich dem Erwartungswert der δ-Funktion, und die kumulative Wahrscheinlichkeitsverteilung ist gleich dem Erwartungswert der Θ-Funktion (Abschnitt 2.5.2).

Eine halbe Seite (Anhang A.3) genügt, um die Definition und die Eigenschaften der Deltafunktion zu formulieren.

Es ist bemerkenswert, dass so gut wie alle Lehrbücher der Wahrscheinlichkeitstheorie auf den Gebrauch dieses effizienten Werkzeugs verzichten.

Schliesslich sei noch auf Beiträge hingewiesen, die einen Bezug zum physikalischen Praktikum haben: Man findet sie auf den Seiten 90 (Fehlerrechnung), 98 (Fehlerbalken), 184, 186 und 189 (ausführliche Beispiele mit einfacher linearer Regression), 193 (Regression von Hand lediglich mit Taschenrechner).

1.2.3 SAS-Anwender/innen

Zum vorliegenden Buch werden 61 SAS-Statistik-Programme zum Download bereitgestellt. Man kann sie nachschlagen in einem Index auf Seite 397. Dieser ist in zwei Gruppen 1 und 2 unterteilt:

Gruppe 1 enthält 37 SAS-Programme zur Erstellung von Grafiken und statistischen Tabellen in Kapitel 2 bis 4.

- SAS-Grafiken zur Wahrscheinlichkeitstheorie finden Sie zwischen Seite 21 und Seite 95.
- In statistischen Tabellen wird auch heute noch aus pädagogischen Gründen nachgeschlagen, um Standard-Aufgaben der schließenden Statistik manuell zu lösen. Solche Tabellen finden Sie zwischen Seite 60 und Seite 76.

Gruppe 2 enthält 24 alternative SAS-Lösungen zu ausgewählten R-Beispielen, (`S_38.sas`) bis (`S_61.sas`). Für Anwender/innen mit geringer R-, aber viel SAS-Erfahrung kann die Gegenüberstellung von R-Lösung und SAS-Lösung hilfreich sein beim Einstieg in die R-Programmierung.

1.3 Anmerkungen zur Notation

R-Code und R-Output hinterlegen wir grau.

Nummern von Gleichungen haben die Form (C.Lfdnr). Dabei ist C die Nummer eines Kapitels (Chapter) oder der Buchstabe eines Appendix, und Lfdnr eine laufende Nummer pro Kapitel. Die Klammern sind obligatorisch. Beispiele: (7.12) oder (A.5).

Nummern von Tabellen haben die Form „Tabelle C.Lfdnr“, also ohne Klammern, aber mit dem Wort Tabelle.

Nummern von Abbildungen haben die Form „Abbildung C.Lfdnr“

Nummern von Abschnitten haben die Form C.S, C.S.sS oder C.S.sS.ssS, ohne Klammern!. Dabei sind S, sS und ssS die Nummern von Section, Subsection und Subsubsection.

Matrizen schreiben wir fett und groß. Beispiele: $\mathbf{X}$, $\boldsymbol{\Sigma}$.

Zufallsvariablen unterstreichen wir. Beispiele: $\underline{x}, \underline{y}, \underline{z}, \underline{t}$. Ausnahmen: siehe unten.

Arithmetische Stichproben-Mittelwerte überstreichen wir. Obwohl sie selbst Zufallsvariablen sind, verzichten wir auf eine extra Unterstreichung. Beispiele: $\bar{x}, \bar{y}, \bar{z}$.

Zufallsvektoren sind klein, fett und unterstrichen. Beispiel: $\underline{\boldsymbol{y}} = (\underline{y}_1 \cdots \underline{y}_n)'$.

Operatoren der Wahrscheinlichkeitsrechnung kennzeichnen wir mit kalligraphischen Lettern $\mathscr{P}$, $\mathscr{E}$, $\mathscr{D}$.

$\mathscr{P}(A)$**:** Wahrscheinlichkeit für Ereignis A

$\mathscr{P}(A|B)$**:** Bedingte Wahrscheinlichkeit

$\mathscr{E}(\underline{z}_i)$**:** Erwartungswert von $\underline{z}_i$, $\mathscr{E}(\underline{\boldsymbol{z}})$: Vektor der Erwartungswerte von $\underline{\boldsymbol{z}}$

$\mathscr{D}^2(\underline{z}) = \mathscr{D}(\underline{z}, \underline{z})$**:** Varianz von $\underline{z}$ (2.31),

$\mathscr{D}(\underline{z}_1, \underline{z}_2)$**:** Kovarianz von $\underline{z}_1$ und $\underline{z}_2$ (2.32);

$\mathscr{D}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}')$**:** Kovarianzmatrix der Vektoren $\underline{\boldsymbol{x}}$ und $\underline{\boldsymbol{z}}$ (2.34)

Griechische Buchstaben verwenden wir (in der Wahrscheinlichkeitstheorie) für unbekannte – d.h. zu schätzende – Parameter. Beispiele: μ, σ^2, β.

Schätzwerte aus einer Stichprobe kennzeichnen wir durch Überdachung. Beispiele: $\hat{\mu}$, $\hat{\sigma}$, $\hat{\mathscr{D}}^2(\underline{z})$, $\widehat{\mathscr{D}}(\underline{x}, \underline{z})$, $\widehat{\mathscr{D}}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}')$. Auf eine extra Unterstreichung verzichten wir, obwohl Schätzwerte immer zufällig sind.

Bemerkung: In diesem Buch werden *nur erwartungstreue Schätzer* verwendet, d.h. wir fordern $\mathscr{E}(\hat{\beta}) = \beta$ für alle Schätzer $\hat{\beta}$ eines Parameters β.
(z.B. $\mathscr{E}(\hat{\mu}) = \mu$, $\mathscr{E}(\hat{\sigma}^2) = \sigma^2$, $\mathscr{E}(\hat{\mathscr{D}}^2(\underline{z})) = \mathscr{D}^2(\underline{z})$,
$\mathscr{E}(\widehat{\mathscr{D}}(\underline{x}, \underline{z})) = \mathscr{D}(\underline{x}, \underline{z})$, $\mathscr{E}(\widehat{\mathscr{D}}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}')) = \mathscr{D}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}')$ usw.)

Beweise: Folgt A=B aus Gleichung x oder aus Abschnitt x, so schreiben wir, um Platz zu sparen, das x direkt über das Gleichheitszeichen. Im einzelnen:
$A \stackrel{(x)}{=} B$ heißt „Beweis siehe Gleichung oder Abschnitt (x)".
$A \stackrel{(x)}{:=} B$ heißt „Definition siehe Gleichung oder Abschnitt (x)" und
$A \stackrel{(x)}{\leadsto} B$ heißt „Aus A folgt B mittels Gleichung oder Abschnitt (x)".
$A \stackrel{5.3.3⑧}{=} B$ heißt „Beweis siehe Abschnitt 5.3.3 Zeile ⑧".
$A \stackrel{\text{Tab.5.3③c}}{=} B$ heißt „Beweis siehe Tabelle 5.3 Zeile ⑦ Spalte e".
$A \stackrel{\checkmark}{=} B$ heißt „Beweis offensichtlich".
$A \stackrel{\text{zz}}{=} B$ heißt „ist zu beweisen („zu zeigen")".

Wortwörtlich übernommene Textpassagen: (z.B. Aufgabentexte) werden in Anführungszeichen gesetzt, und der Kurztitel wird mit dem Hinweis „Ww." versehen. Beispiele

- Seite 101 „Spülmittel", Aufgabe [Wiki 01, Ww.]
- Seite 108 „Betonstahlmatten", Aufgabe [Dürr/Walter 1987, S. 136 Ww.]
- Seite 134 „Forellenbeispiel", Aufgabe [Wiki 02, Ww.]

1.4 Symbole und Formelzeichen

Die Formelzeichen dieses Buches sind ***eindeutig, systematisch*** und können – dadurch bedingt – manchmal ***recht kompliziert*** erscheinen. Die Formelzeichen selbst werden in vier Tabellen vorgestellt:

Die erste Tabelle Tabelle 1.1 S. 7 enthält Symbole und Formelzeichen zu *sieben diskreten eindimensionalen Wahrscheinlichkeitsverteilungen*

Die zweite Tabelle Tabelle 1.2 S. 8 enthält Symbole und Formelzeichen mit Bezug zu *zwölf kontinuierlichen eindimensionalen Wahrscheinlichkeitsverteilungen*. Man beachte das *einheitliche Schema* für Verteilungsdichten, kumulative Verteilungsfunktionen und Quantilfunktionen. Ihre Bezeichnungen sind außerordentlich *effizient*. Mit „Effizienz" ist gemeint, dass die Notation der Quantilfunktionen als eindeutige Umkehrfunktion vermöge der Beziehungen (2.91) und (2.92) quasi „von selbst rechnet", d.h. dem Statistiker streckenweise das Denken abnimmt – ähnlich wie die Diracsche Bra- und Ket-Notation dem Physiker das Denken.

Die dritte Tabelle Tabelle 1.3 S. 9 enthält Formelzeichen zu *Grundbegriffen* wie Wahrscheinlichkeit, Zufallsvariable, Erwartungswert, Mittelwert, Streuung, Varianz und Kovarianz. Der Vollständigkeit halber werden darin auch die etwas weniger systematischen, aber auch gebräuchlichen Bezeichnungen $K_{\underline{z}_1 \underline{z}_2}$ und $\sigma^2_{\underline{z}}$ erwähnt. Bemerkung: Diese Schreibweisen verwenden wir aber nicht.

Die vierte Tabelle Tabelle 1.3 S. 9 enthält Formelzeichen zu *speziellen Begriffen* der Stichproben- und Regressionsanalyse.

Tabelle 1.1: Symbole zur Bezeichnung eindimensionaler diskreter Wahrscheinlichkeitsverteilungen.

Symbol	Name	Def.	Dichte
$\mathrm{Hyp}_{n,N,M}$	Hypergeometrische Verteilung	3.1.1	$f_{\mathrm{Hyp}_{n,N,M}}(z) = \sum_{\substack{x= \\ \max(0,k-n)}}^{\min(m,k)} \delta(z-x)\mathrm{P}_{\mathrm{Hyp}_{n,N,M}}(x)$
$\mathrm{Nhyp}_{n,N,M}$	Negative Hypergeom. Verteilung	3.1.2	$f_{\mathrm{Nhyp}_{n,N,M}}(z) = \sum_{x=m}^{m+N-m} \delta(z-x)\mathrm{P}_{\mathrm{Nhyp}_{n,N,M}}(x)$
$\mathrm{Bin}_{N,p}$	Binomialverteilung	3.1.3	$f_{\mathrm{Bin}_{N,p}}(z) = \sum_{x=0}^{N} \delta(z-x)\mathrm{P}_{\mathrm{Bin}_{N,p}}(x)$
Poi_{λ}	Poisson-Verteilung	3.1.4	$f_{\mathrm{Poi}_{\lambda}}(z) = \sum_{x=0}^{\infty} \delta(z-x)\mathrm{P}_{\mathrm{Poi}_{\lambda}}(x)$
GeoA_{q}	geometr. Verteilung Variante 'A'	3.1.5	$f_{\mathrm{GeoA}_{q}}(z) = \sum_{x=0}^{\infty} \delta(z-x)\mathrm{P}_{\mathrm{GeoA}_{q}}(x)$
GeoB_{q}	geometr. Verteilung Variante 'B'	3.1.5	$f_{\mathrm{GeoB}_{q}}(z) = \sum_{x=0}^{\infty} \delta(z-x)\mathrm{P}_{\mathrm{GeoB}_{q}}(x)$
$\mathrm{Nbin}_{q,v}$	negative Binomialv.	3.1.6	$f_{\mathrm{Nbin}_{q,v}}(z) = \sum_{x=0}^{\infty} \delta(z-x)\mathrm{P}_{\mathrm{Nbin}_{q,v}}(x)$

Tabelle 1.2: Symbole zur Bezeichnung eindimensionaler kontinuierlicher Wahrscheinlichkeitsverteilungen

Symbol	Name der Verteilung	Definition	Dichte (2.23)	kumulative Verteilung (2.12)	Quantilfunktion 2.7
$*$	allgemein		$f_*(z)$	$F_*(z)$	$F_*^{-1}(p)$
$\mathrm{N}_{0,1}$	Normal-Verteilung	(2.87)	$f_{\mathrm{N}_{0,1}}(z)$	$F_{\mathrm{N}_{0,1}}(z)$	$F_{\mathrm{N}_{0,1}}^{-1}(p)$
$\mathrm{N}_{\mu,\sigma}$	Gauß-Verteilung	3.2.1	$f_{\mathrm{N}_{\mu,\sigma}}(z)$	$F_{\mathrm{N}_{\mu,\sigma}}(z)$	$F_{\mathrm{N}_{\mu,\sigma}}^{-1}(p)$
$\mathrm{C}_{\mu,\lambda}$	Cauchy-Verteilung	3.2.2	$f_{\mathrm{C}_{\mu,\lambda}}(z)$	$F_{\mathrm{C}_{\mu,\lambda}}(z)$	$F_{\mathrm{C}_{\mu,\lambda}}^{-1}(p)$
χ^2_n	χ^2-Verteilung	3.2.3	$f_{\chi^2_n}(z)$	$F_{\chi^2_n}(z)$	$F_{\chi^2_n}^{-1}(p)$
$\chi^2_{n,\lambda}$	nichtzentrale χ^2-Verteilung	3.2.4	$f_{\chi^2_{n,\lambda}}(z)$	$F_{\chi^2_{n,\lambda}}(z)$	$F_{\chi^2_{n,\lambda}}^{-1}(p)$
t_n	t-Verteilung	3.2.5	$f_{\mathrm{t}_n}(z)$	$F_{\mathrm{t}_n}(z)$	$F_{\mathrm{t}_n}^{-1}(p)$
$\mathrm{t}_{n,\mu}$	nichtzentrale t-Verteilung	3.2.6	$f_{\mathrm{t}_{n,\mu}}(z)$	$F_{\mathrm{t}_{n,\mu}}(z)$	$F_{\mathrm{t}_{n,\mu}}^{-1}(p)$
$\mathrm{G}_{a,b}$	Gleichverteilung	Tab. 3.10①	$f_{\mathrm{G}_{a,b}}(z)$	$F_{\mathrm{G}_{a,b}}(z)$	$F_{\mathrm{G}_{a,b}}^{-1}(p)$
$\mathrm{F}_{m,n}$	F-Verteilung	3.2.7	$f_{\mathrm{F}_{m,n}}(z)$	$F_{\mathrm{F}_{m,n}}(z)$	$F_{\mathrm{F}_{m,n}}^{-1}(p)$
$\mathrm{F}_{m,n,\lambda}$	nichtzentrale F-Verteilung	3.2.8	$f_{\mathrm{F}_{m,n,\lambda}}(z)$	$F_{\mathrm{F}_{m,n,\lambda}}(z)$	$F_{\mathrm{F}_{m,n,\lambda}}^{-1}(p)$
E_β	Exponential-Verteilung	3.2.9	$f_{\mathrm{E}_\beta}(z)$	$F_{\mathrm{E}_\beta}(z)$	$F_{\mathrm{E}_\beta}^{-1}(p)$
$\mathrm{W}_{T,k}$	Weibull-Verteilung	3.2.10	$f_{\mathrm{W}_{T,k}}(z)$	$F_{\mathrm{W}_{T,k}}(z)$	$F_{\mathrm{W}_{T,k}}^{-1}(p)$

Tabelle 1.3: Formelzeichen für Grundbegriffe wie Wahrscheinlichkeit, Zufallsvariable; Erwartungswert, Mittelwert; Varianz, Kovarianz, Streuung.

Formelzeichen	kurz	Bedeutung	Def.
$\mathscr{P}(A)$		Wahrscheinlichkeit für Ereignis A	(2.8)
$\mathscr{P}(A\|B)$		Bedingte Wahrscheinlichkeit	(2.9)
$\tilde{\alpha}$		Irrtumswahrscheinlichkeit, Signifikanzniveau	4.2.1
$\underline{z}$, $\underline{x}$, $\underline{y}$, etc.		Zufalls-Variable	2.2.3
$\underline{\boldsymbol{z}} = (\underline{z}_1\ \underline{z}_2\ \ldots)'$		Zufalls-Vektor	
$\mathscr{E}(\underline{z})$	$= \mu$	Erwartungswert von $\underline{z}$	(2.29)
$\bar{z} = \frac{1}{n} \sum\limits_{i=1}^{n} \underline{z}_i$	$= \hat{\mu}$	Stichproben-Mittelwert der $\underline{z}_i$	(4.1)
$\mathscr{E}(\underline{\boldsymbol{z}})$	$= \boldsymbol{\mu}$	Vektor der Erwartungswerte von $\underline{\boldsymbol{z}}$	
$\mathscr{D}(\underline{z}_1, \underline{z}_2)$	$\overset{n.v.}{=} K_{\underline{z}_1 \underline{z}_2}$ $\overset{n.v.}{=} K_{\underline{z}_1 \underline{z}_2}$	Kovarianz von $\underline{z}_1$ und $\underline{z}_2$	(2.32)
$\mathscr{D}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}')$	$= \boldsymbol{\Sigma}$	Kovarianzmatrix von $\underline{\boldsymbol{x}}$ und $\underline{\boldsymbol{z}}$, Streuungsmatrix	(2.34)
$\widehat{\mathscr{D}}(\underline{x}, \underline{z})$, $\widehat{\mathscr{D}}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}')$		Schätzer von $\mathscr{D}(\underline{x}, \underline{z})$ bzw. $\mathscr{D}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{z}}')$	
$\mathscr{D}^2(\underline{z}) = \mathscr{D}(\underline{z}, \underline{z})$	$= \sigma^2$ $\overset{n.v.}{=} \sigma^2_{\underline{z}}$	Varianz von $\underline{z}$	(2.31)
$\hat{\mathscr{D}}^2(\underline{z}) = \widehat{\mathscr{D}}(\underline{z}, \underline{z})$	$= \hat{\sigma}^2$	Stichprobenvarianz, Schätzer von $\mathscr{D}^2(\underline{z})$	(4.2)
$\mathscr{D}(\underline{z}) = \sqrt{\mathscr{D}(\underline{z}, \underline{z})}$	$= \sigma$	Standardabweichung von $\underline{z}$	2.31
$\hat{\mathscr{D}}(\underline{z}) = \sqrt{\widehat{\mathscr{D}}(\underline{z}, \underline{z})}$	$= \hat{\sigma}$	Stichprobenstandardabweichung, standard error	Tab.4.1④

Tabelle 1.4: Formelzeichen für spezielle Begriffe der Stichproben- und Regressionsanalyse

Formelzeichen	Bedeutung	Def.
$\hat{\mu} = (\underline{z}_1 + \ldots + \underline{z}_n)/n$	Stichproben-Mittelwert	(4.1)
$\hat{\sigma}$	Stichproben-Standardabweichung, standard error	(4.10)
$\hat{\mathscr{D}}(\hat{\mu})$	standard error of the mean	(4.11)
$\tilde{\alpha}$	Irrtumswahrscheinlichkeit, Signifikanzniveau	4.2.1
$\mathbf{X}$	Design-Matrix bei der linearen Regression	(5.5),(5.16)
$\underline{\boldsymbol{y}} = (\underline{y}_1\ \underline{y}_2\ \ldots\ \underline{y}_n)'$	Zufalls-Vektor der n Meßwerte	(5.28)
$\hat{\boldsymbol{y}} = (\hat{y}_1\ \hat{y}_2\ \ldots\ \hat{y}_n)'$	Schätzer (Fit) für $\underline{\boldsymbol{y}}$	(5.33)
$\underline{\boldsymbol{\varepsilon}} = (\underline{\varepsilon}_1\ \ldots\ \underline{\varepsilon}_n)'$	Statistischer Fehler im Regressionsansatz	(5.1),(5.27)
σ^2	unbekannte Varianz von $\underline{\varepsilon}_i$	(5.8),(5.27)
$\hat{\sigma}^2$	Schätzer von σ^2, mean square error	(5.37)
$\boldsymbol{\beta} = (\beta_0\ \beta_1\ \ldots\ \beta_p)'$	unbekannte Regressionsparameter	(5.1)
$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0\ \hat{\beta}_1\ \ldots\ \hat{\beta}_p)'$	Schätzer von $\boldsymbol{\beta}$	(5.30)
$\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1\ \ldots\ \hat{\varepsilon}_n)'$	Residuen	(5.34)
$\hat{\varepsilon}$	Prognosefehler	Tabelle 5.4 10 p
$\underline{R}^2$	Bestimmtheitsmaß	(5.48)
$\underline{R}^2_{\text{adjusted}}$	korrigiertes Bestimmtheitsmaß	(5.49)
$\xi, \underline{\eta}$	Schwerpunktskoordinaten bei der einfachen linearen Regression	5.8.1
$\lVert\xi\rVert^2, \lVert\underline{\eta}\rVert^2$	Norm von ξ, Norm von $\underline{\eta}$	5.8.1

1.5 Gliederung des Buches

Dieses Lehrbuch spannt einen anwendungsorientierten Bogen von den Grundlagen der Wahrscheinlichkeitstheorie (Kapitel 2 und 3) über zentrale Teile der Mathematischen Statistik (Kapitel 4 bis 6) bis zu praktischen Anwendungen der Statistischen Versuchsplanung (Kapitel 7). Im Inhaltsverzeichnis werden auch die Beispiele aufgelistet, eingebettet in ihren jeweiligen Kontext.

Das Kapitel 2 bildet den ersten Schwerpunkt. Es gibt eine kurze Einführung in die *Wahrscheinlichkeitstheorie*. Darin werden die Begriffe Zufallsversuch, Wahrscheinlichkeit eines Ereignisses, Zufallsvariable, Erwartungswert, u.a. erklärt. Die Wahrscheinlichkeitsverteilung diskreter oder kontinuierlicher Zufallsvariabler sowie von Funktionen derselben wird in eleganter Weise mit Hilfe der Dirac'schen Deltafunktion berechnet. Am Ende von Kapitel 2 findet man weitere für die Mathematische Statistik wichtige Teile der Wahrscheinlichkeitstheorie: Rechenregeln, Symmetrierelationen, Kombinatorik, Quantile. Letztere werden formal als Umkehrfunktionen notiert; das Unbehagen, das einen Statistik-Anfänger befällt, wenn vom Quantil die Rede ist, wird so an einen Formalismus delegiert.

Die Kapitel 3 behandelt die wichtigsten *statistischen Verteilungen*, die zur Beschreibung der (hypothetischen) Verteilung realer Daten verwendet werden. Eindimensionale diskrete Verteilungen, eindimensionale kontinuierliche Verteilungen und mehrdimensionale Gauß-Verteilungen.

Einen zweiten Schwerpunkt dieses Buches bilden die Kapitel 4 und 5 mit den Themen Stichproben und lineare Regression (Ausgleichsrechnung, Fitting).

Ein zentraler Begriff der mathematischen Statistik ist die *Schätzmethode*, nach der Schätzfunktionen für unbekannte Parameter einer statistischen Grundgesamtheit konstruiert werden. Ein klassisches Kriterium für die Beurteilung einer Schätzmethode ist die *Erwartungstreue* (vgl. Gleichungen (4.3), (4.4); ein klassisches Beispiel einer Schätzmethode ist die *Methode der kleinsten Quadrate* (vgl. Gl. (5.11), (5.30), (5.37)). Mit der Hilfe von Schätzmethoden werden Bereichsschätzungen von Parametern der Grundgesamtheit durch *Konfidenzintervalle* (im Abschnitt 4.2) ermittelt. In diesen Kontext gehört auch ein zentrales Thema der Versuchsplanung, die Ermittlung des notwendigen Stichprobenumfangs (sample size): Die Beispiele 4.2.7.1 und 4.2.9.1 demonstrieren den Zusammenhang zwischen sample size und Messfehler (Fehler erster Art, α-Fehler). Der zusätzliche Einfluss von Fehlern *zweiter* Art (β-Fehler) auf die sample size wird im Abschnitt 4.4 (Power-Analyse) behandelt, sowohl analytisch (Abschnitt 4.4.7) als auch mit R-Programmen (Abschnitt 4.4.8).

In den Abschnitten 4.3 und 5.6 wird gezeigt und durch zahlreiche Beispiele (manuell, R, SAS) demonstriert, wie konkrete Vermutungen über die Grundgesamtheit durch geeignete *statistische Tests* bestätigt oder abgelehnt werden.

Im Kapitel 5 wird die Methode der multiplen linearen Regression ausführlich hergeleitet und an Hand von Beispielen mit kommentiertem Computeroutput veranschaulicht.

Kapitel 6 behandelt die Themen Varianzanalyse mit festen Effekten (einfach, zweifach mit einfacher und mehrfacher Besetzung), lateinische bzw. griechisch-lateinische Quadrate und Kovarianzanalyse.

Den dritten, überwiegend praxisbezogenen Schwerpunkt dieses Buches bildet das Kapitel 7 Versuchsplanung mit den Haupt-Themen vollfaktorielle Pläne, teilfaktorielle Pläne und Response Surface Methoden. Weitere Themen sind optimale Pläne und Mixturpläne.

Anhang A enthält mathematische Sätze über Integrale, Matrizen etc., die bei den Beweisen im Hauptteil verwendet werden. Die Anhänge R und S sind Verzeichnisse der 95 R-Programme `R_1.R` bis `R_95.R` und der 61 statistischen SAS-Programme `S_1.sas` bis `S_61.sas`.

1.6 Datenmaterial zu diesem Buch im Internet

Für Leser/innen dieses Buches gibt es eine Möglichkeit, die R-Programme und SAS-Programme zu dem Buch (siehe Auflistung in den Anhängen R und S) aus dem Internet herunterzuladen. Man findet sie auf der Verlagsseite des Buches unter `www.wiley-vch.de/ISBN9783527346295` zum Download.