

GUSTAV GRÖNNÄS

Statistik und Wahrscheinlichkeit

leicht gemacht

GUSTAV GRÖNNÄS

Statistik und Wahrscheinlichkeit

leicht gemacht

2., erweiterte Auflage

© 2007, 2023 Gustav Grönnäs

www.gustav-gronnas.de

ISBN Softcover: 978-3-384-10407-6

Druck und Distribution im Auftrag des Autors:

tr edition GmbH

An der Strusbek 10

22926 Ahrensburg

Germany

Das Werk, einschließlich seiner Teile, ist urheberrechtlich geschützt. Für die Inhalte ist der Autor verantwortlich.

Jede Verwertung über die Grenzen des Urheberrechtsgesetzes ist unzulässig. Die Anfertigung von Kopien, Vervielfältigungen oder Verarbeitung in elektronischen Medien (auch auszugsweise) bedarf der ausdrücklichen, schriftlichen Genehmigung des Autors.

Inhalt

Vorwort	1
Vorwort zur zweiten Auflage	2
Anwendungshinweise	3
Verwendete Symbole	4
1 Datenanalyse	6
1.0 Datentypen	7
1.1 Analyse qualitativer Daten	8
1.2 Analyse quantitativer Daten	11
1.2.1 Mittelwert	11
1.2.2 Zentralwert und Quantil	14
1.2.3 Standardabweichung	19
1.2.4 Praktisches Vorgehen zur Ermittlung von Mittelwert und Standardabweichung	21
1.3 Regression	23
1.3.1 Lineare Regression	23
1.3.2 Praktisches Vorgehen in der linearen Regression	25
1.3.3 Pseudolineare Regression	29
1.3.4 Polynomiale Regression	30
1.3.5 Praktisches Vorgehen in der polynomialen Regression	32
1.3.6 Pseudopolynomiale Regression	35
1.4 Übungsaufgaben zur Datenanalyse	38
1.4.1 Eine Kaffeepreisanalyse	38
1.4.2 Eine Schulnotenanalyse	38
1.4.3 Ein Arzneimitteltest	39
2 Wahrscheinlichkeit	40
2.0 Definition der Wahrscheinlichkeit	40
2.1 Die Negation – das Gegenereignis	41
2.2 Verknüpfungen von Ereignissen	42
2.2.1 Die UND-Verknüpfung – das logische AND	42
2.2.2 Die ODER-Verknüpfung – das logische OR	43
2.2.3 Die ausschließende ODER-Verknüpfung – das logische XOR	44
2.3 Grafische Darstellung logischer Verknüpfungen	45
2.3 Übungsaufgaben zur Wahrscheinlichkeit	48
2.3.1 Die Frischmilchpackungen	48
2.3.2 Die Schulabbrecher	48
2.3.3 Die Autofarben	48
2.3.4 Die guten Noten	48
3 Kombinatorik	49
3.1 Entnahme unterscheidbarer Elemente aus einer konstanten Gesamtheit	51
3.2 Entnahme unterscheidbarer Elemente aus einer variablen Gesamtheit	52

3.3 Entnahme gleicher Elemente aus einer variablen Gesamtheit	54
3.4 Entnahme nicht vollständig unterscheidbarer Elemente aus einer variablen Gesamtheit	55
3.5 Entnahme unterscheidbarer Elemente aus einer endlichen Gesamtheit bei Mehrfachbelegung von Plätzen	56
3.6 Entnahme nicht unterscheidbarer Elemente aus einer endlichen Gesamtheit bei Mehrfachbelegung von Plätzen	57
3.7 Übungsaufgaben zur Kombinatorik	58
3.7.1 Ein Buchregal	58
3.7.2 Ein Passwort	58
3.7.3 Der Gen-Code	58
3.7.4 Die IP-Adressen	58
4 Dichte- und Verteilungsfunktionen	59
4.1 Begriffsbestimmung	59
4.1.1 Dichte- und Verteilungsfunktion auf diskreten Mengen	60
4.1.2 Dichte- und Verteilungsfunktion auf stetigen Mengen	63
4.2 Hypergeometrische Funktion	65
4.3 Geometrische Funktion	68
4.4 Binomiale (Bernoullische) Funktion	70
4.5 Poisson-Funktion	73
4.6 Exponentielle Funktion	75
4.7 Die Gauss-Funktion	77
4.7.1 Die Gauss-Normal-Funktion für stetige Daten	77
4.7.2 Approximation der binomialen Verteilung mittels der Gauss-Verteilungsfunktion	79
4.8 Die Log-Normal-Funktion	82
4.9 Die Weibull-Funktion	85
4.10 Auswahl einer geeigneten Wahrscheinlichkeitsfunktion	88
4.11 Übungsaufgaben zu Dichte- und Verteilungsfunktionen	91
4.11.1 Hamburger Volkslauf	91
4.11.2 Ein Kindergartenausflug	91
4.11.3 Eine Mathematikprüfung	91
4.11.4 Ein Spielautomat	92
4.11.5 Viele Laboruntersuchungen	92
4.11.6 Einige Schulabbrecher	92
4.11.7 Einige Meteoriteneinschläge	92
4.11.8 Eine Warteschlange	93
5 Parameter-Schätzungen	94
5.1 Parameter-Schätzung in der hypergeometrischen Dichtefunktion	94
5.1.1 Schätzung der Gesamt-Element-Anzahl	94
5.1.2 Schätzung der Gesamt-Günstigen-Elemente-Anzahl	96
5.2 Parameter-Schätzung in der binomialen Dichtefunktion	96
5.2.1 Schätzung der Grundwahrscheinlichkeit	96
5.2.2 Schätzung der Gesamt-Element-Anzahl	98

5.2.3 Schätzung der Grundwahrscheinlichkeit und der Entnahmemenge	99
5.3 Konfidenzintervalle	101
5.3.1 Tschebyschewsche Ungleichung	103
5.3.2 Schätzung der Entnahmemenge	104
5.4 Parameterschätzung in der Weibull-Funktion	106
5.5 Übungsaufgaben zu Parameter-Schätzungen	108
5.5.1 Die Tiefkühlpackungen	108
5.5.2 Die Klausur-Beurteilungen	108
5.5.3 Die faulen Äpfel	108
5.5.4 Das Füllvolumen	108
6 Untersuchungen und Testkonzepte	109
6.1 Planung	109
6.1.1 Festlegung der Testsituation	109
6.1.2 Wahl der Aussagesicherheit	109
6.1.3 Wahl der Aussagegenauigkeit	110
6.1.4 Aufbau einer Datentabelle	114
6.1.5 Ermittlung der Mindestanzahl der Messungen	115
6.2 Ablaufdiagramm einer Testplanung	117
6.3 Testauswertung	119
6.4 Übungsaufgaben zu Testkonzeptionen	121
6.4.1 Ein Erkennungssystem	121
6.4.2 Der Kraftstoffverbrauch	121
6.4.3 Der Leistungstest	121
6.4.4 Der Werkstofftest	121
7 Hypothesentests	122
7.1 Einfache Tests	123
7.2 Chi-Quadrat-Test	126
7.3 Student-t-Test	136
7.4 Fisher-F-Test	139
7.5 ANOVA-Variationen	146
7.5.1 Fisher-ANOVA-Test	151
7.5.2 Turkey-ANOVA-Test	152
7.5.3 Newman-Keuls-ANOVA-Test	153
7.5.4 Duncan-ANOVA-Test	154
7.6 ANOVA-Tests Kurzreferenz	157
7.6.1 Ermittlung der Testgrößen	157
7.6.2 Testentscheidung	161
7.6.3 Warnhinweis	161
7.7 Übungsaufgaben zu Hypothesentests	164
7.7.1 Eine Servicebeurteilung	164
7.7.2 Die Warteschlange	164
7.7.3 Die Halsschmerzen	164
7.7.4 Ein Ja-Nein-Test	165

8 Bedingte Wahrscheinlichkeit	166
8.1 Praktisches Vorgehen zur Ermittlung der bedingten Wahrscheinlichkeit	170
8.2 Grafische Darstellung der bedingten Wahrscheinlichkeit	173
8.3 Übungsaufgaben zur bedingten Wahrscheinlichkeit	175
8.3.1 Ein Krankheitstest	175
8.3.2 Ein Mail-Filter	175
8.3.3 Ein Alkoholproblem	175
8.3.4 Der Kraftstoffverbrauch	175
9 Datenfilterung	177
9.1 Filterung quantitativer Daten	178
9.1.1 Filterung einzelner quantitativer Daten	178
9.1.2 Filterung quantitativer Regressions-Daten	179
9.2 Filterung qualitativer Daten	182
9.3 Praktische Vorgehen zur Datenfilterung	184
9.4 Übungsaufgaben zur Datenfilterung	187
9.4.1 Das Schlafbedürfnis	187
9.4.2 Die Schulnotenausreißer	187
9.4.3 Die Luftstrahlung	187
9.4.4 Die Regressionsdaten	187
A Lösungen der Übungsaufgaben	188
A.1 Lösungen der Übungsaufgaben zur Analyse vorhandener Daten	188
A.1.1 Kaffeepreisanalyse	188
A.1.2 Schulnotenanalyse	191
A.1.3 Arzneimitteltest	193
A.2 Lösungen der Übungsaufgaben zur elementaren Wahrscheinlichkeit	194
A.2.1 Die Frischmilchpackungen	194
A.2.2 Die Schulabbrecher	194
A.2.3 Die Autofarben	195
A.2.4 Die guten Noten	196
A.3 Lösungen der Übungsaufgaben zur Kombinatorik	197
A.3.1 Ein Buchregal	197
A.3.2 Ein Passwort	197
A.3.3 Der Gen-Code	198
A.3.4 Die IP-Adressen	198
A.4 Lösungen der Übungsaufgaben zu Dichte- und Verteilungsfunktionen	200
A.4.1 Der Hamburger Volkslauf	200
A.4.2 Der Kindergartenausflug	201
A.4.3 Die Mathematikprüfung	202
A.4.4 Der Spielautomat	202
A.4.5 Die Laboruntersuchungen	204
A.4.6 Die Schulabbrecher	205

A.4.7 Die Meteoriteneinschläge	206
A.4.8 Die Warteschlange	207
A.5 Lösungen der Übungsaufgaben zu Parameterschätzungen	208
A.5.1 Die Tiefkühl-Packungen	208
A.5.2 Die Klausur-Beurteilungen	208
A.5.3 Die faulen Äpfel	209
A.5.4 Das Füllvolumen	210
A.6 Lösungen der Übungsaufgaben zu Testkonzeptionen	211
A.6.1 Ein Erkennungssystem	211
A.6.2 Der Kraftstoffverbrauch	212
A.6.3 Der Leistungstest	213
A.6.4 Der Werkstofftest	213
A.7 Lösungen der Übungsaufgaben zu Hypothesentests	215
A.7.1 Die Servicebeurteilungen	215
A.7.2 Die Warteschlangen	219
A.7.3 Die Halsschmerzen	222
A.7.4 Ein Ja-Nein-Test	224
A.8 Lösungen der Übungsaufgaben zur bedingten Wahrscheinlichkeit	227
A.8.1 Ein Krankheitstest	227
A.8.2 Ein Mail-Filter	228
A.8.3 Ein Alkoholproblem	229
A.8.4 Der Kraftstoffverbrauch	230
A.9 Lösungen der Übungsaufgaben zur Datenfilterung	232
A.9.1 Das Schlafbedürfnis	232
A.9.2 Die Schulnotenausreißer	235
A.9.3 Die Luftstrahlung	236
A.9.4 Die Regressionsdaten	237
B Tabellen	240
B.1 Gauss-Normal-Verteilung	242
B.2 Chi-Quadrat-Verteilung	245
B.3 Student-t-Verteilung	247
B.4 Fisher-F-Verteilung	249
Index	253

Vorwort

Ebenso im Alltag, wie in wissenschaftlichen Arbeiten sind Statistiken gegenwärtig. Sei es nun, dass eine Umfrage deutliche Zustimmung zu einer Aussage ergibt, oder ein neues Medikament eine bessere Wirksamkeit aufweist, stets werden die Methoden der Statistik eingesetzt um die Ergebnisse zu erzielen.

Gelegentlich sind Sätze, wie 'mit Statistik lässt sich alles beweisen' zu hören. Doch die 'Lügen der Statistik' entstehen aus der beabsichtigt oder unbeabsichtigt falschen Anwendung der statistischen Methoden. Fehler – und damit auch Lügen – sind aber stets erkennbar. Es ist stets nur eine Frage des 'Gewusst wie'.

Diese Schrift richtet sich ebenso an Schülerinnen und Schüler, wie an Studierende oder beruflich Interessierte und bietet einen einfachen, aber dennoch tiefgehenden Zugang zu allen wichtigen, mit der Wahrscheinlichkeitsrechnung oder der Statistik verbundenen Fragen.

Eine vollständige Darstellung aller Verfahren und Aufgabenstellungen ist natürlich nicht möglich. Hier ist eine Beschränkung auf das häufigst Benötigte erforderlich. Es wird aber, an den entsprechenden Stellen, auf weiterführende Methoden verwiesen.

Es wird bewusst auf Beweisführung verzichtet. Statt dessen wird ausführlich auf alle Lösungsprobleme eingegangen. Ein Verzicht auf Fachbegriffe ist aber leider nicht möglich. Dem Leser, der Leserin wird empfohlen, sich diese Begriffe so bald wie möglich zu eigen zu machen.

Vorrangiges Ziel ist es, einen programmartigen Ablauf der Vorgehensweisen so darzustellen, dass alle Verfahren auch ohne umfangreiche Vorkenntnisse anwendbar werden.

Besonderer Dank gebührt den Studenten der *Fachhochschule Bremerhaven*, die mit ihren Fragen und Vorschlägen an der Entstehung dieses Buches mitgewirkt haben!

Möge allen das nötige Handwerkszeug für eine erfolgreiches Arbeiten gegeben sein.

Bremerhaven, im Februar 2007

Vorwort zur zweiten Auflage

Der Text wurde vom Originalautor HELGE NORDMANN übernommen und erheblich überarbeitet und erweitert.

Aus vielen Kommentaren und Leser- oder Leserinnenanfragen ergaben sich viele kleine Textveränderungen, die die Verständlichkeit verbessern sollen. Druckfehler und Layoutfehler konnten entsprechend auch behoben werden.

Insbesondere aber wurden viele Wahrscheinlichkeitsfunktionen neu aufgenommen. Hier sind die Log-Normal-Funktion und die WEIBULL-Funktion beispielhaft zu nennen. Gerade für Semester- oder Abschlussarbeiten wurden diese Funktionen wiederholt nachgefragt.

Auch die Bildung und Darstellung von Quantilen wurde vielfach gewünscht und neu aufgenommen.

Zur besseren Auffindbarkeit wurden die Tabellen an das Ende des Buches verschoben.


Und schließlich wurden viele Graphiken und Funktionsgraphen neu eingefügt. Hier gebührt RENE GROTHMAN, dem Autor des Mathematikprogrammes *Euler Math Toolbox* besonderer Dank. Ohne sein Programm wären die Darstellungen kaum so einfach zu realisieren gewesen. Und auch die Überprüfung vieler Aussagen oder Beispiele auf Richtigkeit konnte effektiv nur mittels nachrechnen mit diesem Programm durchgeführt werden.


Trotz der vielen Erweiterungen kann dieses Werk natürlich keine vollständige Darstellung aller Verfahren der Statistik oder Wahrscheinlichkeitstheorie sein. Für die weitaus meisten Anwender oder Anwenderinnen – und für diese ist das Buch geschrieben – sollten die benötigten Verfahren jedoch nachvollziehbar dargestellt sein.


Ich wünsche allen Lesern und Leserinnen ein erfolgreiches Arbeiten mit dieser Wegleitung.

Mora, im Dezember 2023

Anwendungshinweise

 Alle Kapitel dieses Buches bauen auf einander auf. Deshalb ist es für jene, die dieses Buch als Lehrbuch anwenden wollen, erforderlich, die Kapitel in der vorgegebenen Reihenfolge zu lesen.

 Wer hingegen nur bei Bedarf etwas nachschlagen möchte, findet zusätzlich zum Inhaltsverzeichnis einen Index am Ende dieses Buches, so dass sich jede Fragestellung, jede Methode und jedes Verfahren schnell auffinden lässt. Es sei aber noch einmal darauf hingewiesen, dass alle Kapitel auf einander aufbauen. Damit kann ein Verständnis eines Kapitels, nur in Kenntnis der vorigen Kapitel erwartet werden.

 Beispiele sind nach ihrer Art unterschiedlich dargestellt. Vollständige Beispiele tragen eine Abschnitsnummerierung, die mit einem **B** gekennzeichnet ist.

Teilbeispiele werden eingerückt dargestellt.

Verwendete Symbole

α :	Irrtumswahrscheinlichkeit ('alpha')
α, β :	Parameter der WEIBULL-Funktion ('alpha', 'beta')
κ :	Freiheitsgrad ('kappa')
ϕ :	Wahrscheinlichkeit aus der GAUSS-Verteilung ('Phi')
μ :	Mittelwert, theoretisch
σ :	Standardabweichung, theoretisch
χ :	Testgröße 'Chi'
Σ :	Summe ('Sigma')
\int :	Integral
Ω :	Menge von Ereignissen, Ereignisraum ('Omega')
E :	Ereignis
f :	Testgröße des FISHER-Tests, theoretisch
F :	Testgröße des FISHER-Tests, ermittelt
G :	GINI-Koeffizient
h :	Häufigkeit
H :	Hypothese
i, j :	Index, zumeist veränderlich $i, j \in \{0; 1; \dots n\}$
k :	Anzahl der Abstufungen oder Klassen
K :	Anzahl Kombinationsmöglichkeiten
m :	Anzahl (falls n bereits im Gebrauch)
n :	Anzahl
N :	Anzahl in einer großen Menge
p :	Wahrscheinlichkeit
R :	Anzahl unterscheidbarer Elemente in einer großen Menge
$r; s$:	Anzahl unterscheidbarer Elemente (Sorten)
s :	Standardabweichung, errechnet
t :	Testgröße des STUDENT-Tests, theoretisch
T :	Testgröße des STUDENT-Tests, ermittelt

u :	Faktor zur Intervalleingrenzung
\bar{x} :	Mittelwert, errechnet
x_s :	Standardabweichung, errechnet
x,y :	Variable
z :	Transformierte Variable

1 **Datenanalyse**

Häufig treten bei der Untersuchung beliebiger Phänomene große Datenmengen auf, die auf Grund ihres Umfanges zu unübersichtlich sind, um noch verstanden zu werden. Die Zusammenfassung und Interpretation dieser Daten (*Analyse*) wird in der *Statistik* versucht. Entsprechend der Vielfalt der Fragestellungen, ist eine Fülle von Fachbegriffen entstanden, die aber zumeist zu wenigen Begriffen zusammenfassbar sind.

Im Folgenden wird weitestgehend die Vielzahl der Begriffe vermieden und statt dessen eine einheitliche Bezeichnungsweise bevorzugt. Auf feinere Begriffsunterscheidungen wird dann, soweit erforderlich, hingewiesen.

1.0 Datentypen

Grundsätzlich müssen bei der Analyse vorhandener Daten zwei Typen von Daten unterschieden werden: *qualitative Daten* und *quantitative Daten*.

- Qualitative Daten sind Daten, die eine nicht größensortierbare Eigenschaft beschreiben. Beispiele hierfür sind: Das Geschlecht der Besucher einer Ausstellung, Produktnamen etc.
- Quantitative Daten sind Daten, die ihrer Größe nach sortierbar sind (sie sind eben in ihrer Quantität messbar). Beispiele hierfür sind Längen, Preise für Waren etc.

Quantitative Daten werden wiederum in zwei Typen unterschieden: *diskrete Daten* und *stetige Daten*.

- Diskrete Daten sind Daten, die schrittweise größenveränderlich sind. Beispiele hierfür sind: Anzahlen (natürliche Zahlen), aber auch beliebige andere Größen, etwa die Verkaufsmenge pro Kunde eines Produktes das in 0,25kg Paketen verkauft wird etc.
- Stetige Daten sind Daten, die aus den reellen Zahlen stammen. Beispiele hierfür sind: Körper'gewichte' von Schülern, jährlich gemessene Fahrstrecken bestimmter Fahrzeugtypen etc.

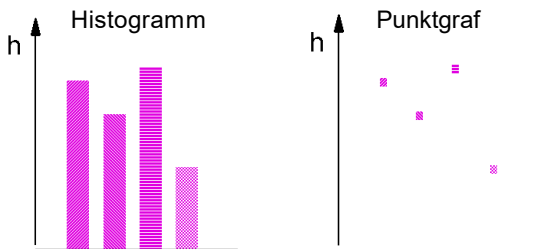
1.1 Analyse qualitativer Daten

Unabhängig davon, ob die vorliegenden Daten quantitativer oder qualitativer Natur sind, kann die Anzahl eines jeden gemessenen Datums, die *absolute Häufigkeit* h_i (kurz: die *Häufigkeit*) angegeben werden.

Gelegentlich ist nicht die absolute Anzahl der Daten, sondern der Anteil eines bestimmten Datums an der Gesamtanzahl $\sum h_i$ aller Daten von Interesse. Daher wird die *relative Häufigkeit* $h_{i\text{rel}}$ definiert (vgl. Wahrscheinlichkeit, S. 40):

$$h_{i\text{rel}} = \frac{1}{\sum h_i} h_i .$$

Die absolute oder relative Häufigkeit lässt sich anschaulich in Grafen darstellen. Für die Häufigkeit quantitativer Daten sind nur *Histogramme* (*Blockdiagramme*) oder *Punktgraf* geeignet, so dass die Häufigkeit über dem Messwert aufgetragen wird (d.h.: die Häufigkeit wird als Funktionswert und der Messwert als Argument einer Funktion angesehen):



Qualitative Daten lassen sich auch in weiteren Grafen, etwa *Sektorgraf*en ('*Tortendiagrammen*') oder beliebig gestalteten Symbolen unterschiedlicher Größe darstellen:

