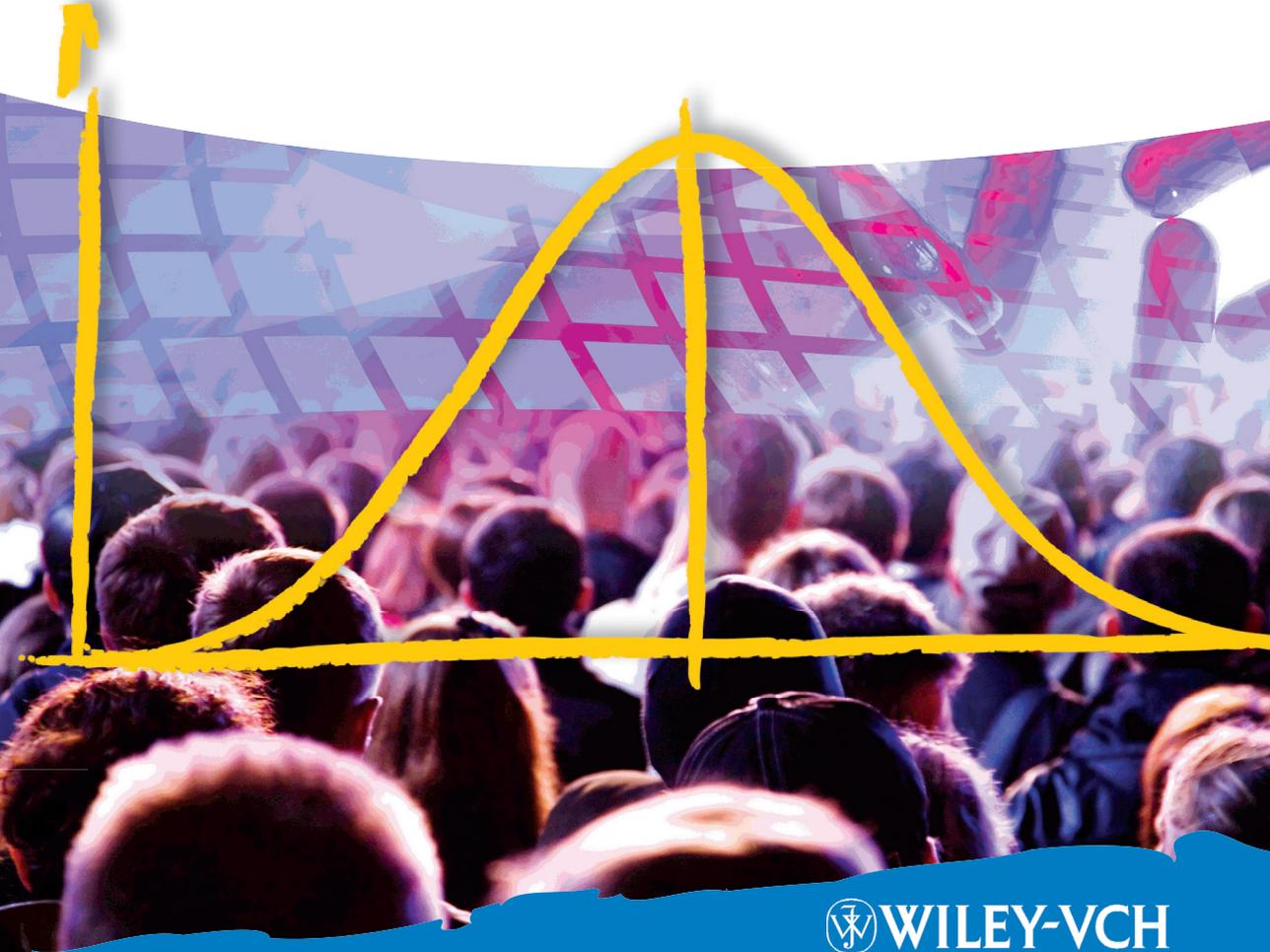


Verdammt clever!



Philip Rowe

Statistik für Mediziner und Pharmazeuten



 **WILEY-VCH**

Philip Rowe

**Statistik für Mediziner
und Pharmazeuten**

***Beachten Sie bitte auch
weitere interessante Titel
zu diesem Thema***

Ansorge, R., Oberle, H. J., Rothe, K.,
Sonar, T.

Mathematik für Ingenieure 1

Lineare Algebra und analytische
Geometrie, Differential- und
Integralrechnung einer Variablen

2010

ISBN: 978-3-527-40980-8

Rumsey, D.

Statistik für Dummies

2010

ISBN: 978-3-527-70594-8

Zachmann, H. G., Jünger, A.

Mathematik für Chemiker

2007

ISBN: 978-3-527-30315-1

Cann, A. J.

Mathe für Biologen

2004

ISBN: 978-3-527-31183-5

Philip Rowe

Statistik für Mediziner und Pharmazeuten



**WILEY-
VCH**

WILEY-VCH Verlag GmbH & Co. KGaA

Autor

Philip Rowe

Liverpool John Moores University, UK

Übersetzer

Dipl.-Phys. Carsten Heinisch, Kaiserslautern
redaktor.de

1. Auflage 2012

Alle Bücher von Wiley-VCH werden sorgfältig erarbeitet. Dennoch übernehmen Autoren, Herausgeber und Verlag in keinem Fall, einschließlich des vorliegenden Werkes, für die Richtigkeit von Angaben, Hinweisen und Ratschlägen sowie für eventuelle Druckfehler irgendeine Haftung

**Bibliografische Information
der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <<http://dnb.d-nb.de>> abrufbar.

© 2012 Wiley-VCH Verlag & Co. KGaA,
Boschstr. 12, 69469 Weinheim, Germany

Alle Rechte, insbesondere die der Übersetzung in andere Sprachen, vorbehalten. Kein Teil dieses Buches darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Photokopie, Mikroverfilmung oder irgendein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsmaschinen, verwendbare Sprache übertragen oder übersetzt werden. Die Wiedergabe von Warenzeichnungen, Handelsnamen oder sonstigen Kennzeichen in diesem Buch berechtigt nicht zu der Annahme, dass diese von jedermann frei benutzt werden dürfen. Vielmehr kann es sich auch dann um eingetragene Warenzeichen oder sonstige gesetzlich geschützte Kennzeichen handeln, wenn sie nicht eigens als solche markiert sind.

Print ISBN: 978-3-527-33119-2

Umschlaggestaltung Simone Benjamin,
McLeese lake, Canada

Satz Beltz Bad Langensalza GmbH,
Bad Langensalza

Druck und Bindung Markono Print Media Pte
Ltd, Singapore

Gedruckt auf säurefreiem Papier.

Für Carol, Joshua und Nathan

Inhaltsverzeichnis

Vorwort XIII

Teil 1 Datentypen

- 1 Datentypen** 3
- 1.1 Kommt es wirklich darauf an? 3
- 1.2 Daten auf einer Intervallskala 3
- 1.3 Daten auf einer Ordinalskala 4
- 1.4 Daten auf einer Nominalskala 5
- 1.5 Aufbau dieses Buchs 6
- 1.6 Kapitelzusammenfassung 6

Teil 2 Daten auf Intervallskalen

- 2 Beschreibende Statistik** 9
 - 2.1 Zusammenfassung von Datensätzen 9
 - 2.2 Zentrale Lagemaße – der Mittelwert, der Median und der Modalwert 10
 - 2.3 Beschreibung der Spannweite – die Standardabweichung und die relative Standardabweichung 16
 - 2.4 Quartile – eine andere Möglichkeit, Daten zu beschreiben 20
 - 2.5 Verwendung von Software für die beschreibende Statistik 23
 - 2.6 Kapitelzusammenfassung 25
-
- 3 Die Normalverteilung** 27
 - 3.1 Was ist eine Normalverteilung? 27
 - 3.2 Wie erkennt man normalverteilte Daten? 28
 - 3.3 Anteile von Einzelwerten innerhalb von einer oder zwei Standardabweichungen vom Mittelwert 31
 - 3.4 Kapitelzusammenfassung 34

4	Stichproben aus einer Grundgesamtheit und der Standardfehler des Mittelwerts	35
4.1	Stichproben und Grundgesamtheiten	35
4.2	Von der Stichprobe zur Grundgesamtheit	36
4.3	Verschiedene Stichprobenfehler	37
4.4	Welche Faktoren bestimmen die Höhe des zufälligen Stichprobenfehlers?	39
4.5	Abschätzung des wahrscheinlichen Stichprobenfehlers und der Standardfehler	42
4.6	Aufrechnung von Stichprobengröße und Standardabweichung	46
4.7	Kapitelzusammenfassung	47
5	Das 95 %-Konfidenzintervall für den Mittelwert	49
5.1	Was ist ein Konfidenzintervall?	49
5.2	Wie breit sollte das Intervall sein?	50
5.3	Was meinen wir mit „95 %-Konfidenz“?	51
5.4	Berechnung der Intervallbreite	52
5.5	Eine Reihe von Stichproben und 95 %-Konfidenzintervallen	53
5.6	Wie stark hängt die Breite des Konfidenzintervalls von Änderungen der Standardabweichung, des Stichprobenumfangs und des gewünschten Konfidenzniveaus ab?	54
5.7	Zwei Aussagen	57
5.8	Einseitige 95 %-Konfidenzintervalle	57
5.9	Das 95 %-Konfidenzintervall für den Unterschied zweier Behandlungen	60
5.10	Über die Notwendigkeit, dass die Daten einer Normalverteilung folgen und Datentransformation	62
5.11	Kapitelzusammenfassung	65
6	Der doppelte t-Test (1). Einführung in Hypothesentests	67
6.1	Der doppelte t-Test – ein Beispiel für einen Hypothesentest	67
6.2	Signifikanz	74
6.3	Das Risiko eines falsch-positiven Ergebnisses	75
6.4	Von welchen Faktoren hängt es ab, ob wir ein signifikantes oder ein nicht signifikantes Ergebnis erhalten?	76
6.5	Voraussetzungen für einen doppelten t-Test	79
6.6	Kapitelzusammenfassung	80
7	Der doppelte t-Test (2): Der berühmte p-Wert	81
7.1	Wie kann man die Signifikanz eines Ergebnisses beziffern?	81
7.2	p-Werte	81
7.3	Gibt es zwei Arten, Signifikanz zu definieren?	83

- 7.4 Bestimmung des p -Wertes 83
- 7.5 p -Werte oder 95 %-Konfidenzintervalle? 84
- 7.6 Kapitelzusammenfassung 85

- 8 Der doppelte t -Test (3). Falsch-negative Befunde, Güte und notwendige Stichprobenumfänge 87**
 - 8.1 Was könnte sonst noch schief gehen? 87
 - 8.2 Die Güte 89
 - 8.3 Berechnung des notwendigen Stichprobenumfangs 92
 - 8.4 Kapitelzusammenfassung 99

- 9 Der doppelte t -Test (4). Statistische Signifikanz, praktische Bedeutung und Äquivalenz 101**
 - 9.1 Praktische Bedeutung – ist die Differenz so groß, dass sie eine Rolle spielt? 101
 - 9.2 Äquivalenztests 105
 - 9.3 Tests auf Nicht-Unterlegenheit 109
 - 9.4 p -Werte sind weniger aussagekräftig und können förmlich in die Irre führen 111
 - 9.5 Setzen von Äquivalenzgrenzen vor dem eigentlichen Versuch 113
 - 9.6 Kapitelzusammenfassung 114

- 10 Der doppelte t -Test (5). Einseitige Tests 115**
 - 10.1 Suche nach einer Veränderung in einer bestimmten Richtung 115
 - 10.2 Schutz vor falsch-positiven Befunden 118
 - 10.3 Versuchung 119
 - 10.4 Einsatz eines Softwarepakets bei einem einseitigen Test 123
 - 10.5 Sollte man häufiger einseitige Tests einsetzen? 123
 - 10.6 Kapitelzusammenfassung 124

- 11 Was sagt uns ein statistisch signifikantes Ergebnis wirklich? 125**
 - 11.1 Wie interpretiert man statistische Signifikanz? 125
 - 11.2 Am Anfang steht äußerste Skepsis 129
 - 11.3 Kapitelzusammenfassung 130

- 12 Der gepaarte t -Test – Vergleich von zwei zusammenhängenden Datensätzen 131**
 - 12.1 Gepaarte Datensätze 131
 - 12.2 Untersuchung der Daten mithilfe eines doppelten t -Tests 132
 - 12.3 Alternative Anwendung eines gepaarten t -Tests 133

- 12.4 Durchführung eines gepaarten t -Tests 134
- 12.5 Wodurch ist bestimmt, ob ein gepaarter t -Test signifikant ist? 136
- 12.6 Größere Teststärke beim gepaarten t -Test 137
- 12.7 Der gepaarte t -Test ist nur auf natürliche Paare von Daten anwendbar 137
- 12.8 Auswahl des passenden Versuchsaufbaus 138
- 12.9 Voraussetzungen für das Anwenden eines gepaarten t -Tests 139
- 12.10 Stichprobenumfänge, praktische Bedeutung und einseitige Tests 140
- 12.11 Zusammenfassung der Unterschiede zwischen dem gepaarten und dem doppelten t -Test 141

13 Varianzanalyse – über t -Tests hinaus 143

- 13.1 Erweiterung zu komplexen Versuchsdesigns 143
- 13.2 Einfache Varianzanalyse 144
- 13.3 Zweifache Varianzanalyse 153
- 13.4 Multifaktorenversuche 162
- 13.5 Einfache Form – starke Aussage 162
- 13.6 Kapitelzusammenfassung 165

14 Korrelation und Regression – Zusammenhänge zwischen Messwerten 167

- 14.1 Korrelationsanalyse 167
- 14.2 Regressionsanalyse 176
- 14.3 Mehrfache Regression 183
- 14.4 Kapitelzusammenfassung 190

Teil 3 Daten auf Nominalskalen

15 Beschreibung von kategorisierten Daten 195

- 15.1 Beschreibende Statistik 195
- 15.2 Tests, ob der wahre Anteil möglicherweise einen vorbestimmten Wert hat 200
- 15.3 Kapitelzusammenfassung 205

16 Vergleich beobachteter Anteile – der Chi-Quadrat-Kontingenztest 207

- 16.1 Anwendung des Chi-Quadrat-Kontingenztests für den Vergleich von beobachteten Anteilen 207
- 16.2 Ein 95 %-Konfidenzintervall für die Änderung der Ausstoßquote – ist die Änderung von praktischer Bedeutung? 211
- 16.3 Größere Kontingenztafeln – Nutzung der Diabetes-Sprechstunde 212
- 16.4 Planung der Versuchsgröße 215
- 16.5 Kapitelzusammenfassung 217

Teil 4 Daten auf Ordinalskalen

- 17 Ordinalskalierte, nicht normalverteilte Daten. Transformationen und parameterfreie Tests 221**
- 17.1 Transformation auf eine Normalverteilung 222
 - 17.2 Der Mann-Whitney-Test – ein nicht parametrisches Verfahren 226
 - 17.3 Umgang mit Daten auf Ordinalskalen 230
 - 17.4 Andere nicht parametrische Verfahren 233
 - 17.5 Kapitelzusammenfassung 240
 - 17.6 Anhang zu Kapitel 17 241

Teil 5 Reale Herausforderungen

- 18 Mehrfachtests 245**
- 18.1 Was ist ein Mehrfachtest und warum ist er problematisch? 245
 - 18.2 Wo treten Mehrfachtests auf? 246
 - 18.3 Verfahren zur Vermeidung von Falsch-positiven 248
 - 18.4 Die Rolle der wissenschaftlichen Zeitschriften 253
 - 18.5 Kapitelzusammenfassung 254

- 19 Fragebögen 255**
- 19.1 Gibt es Besonderheiten bei Fragebögen? 255
 - 19.2 Arten von Fragen 255
 - 19.3 Entwurf eines Fragebogens 259
 - 19.4 Stichprobenumfang und Rücklaufquoten 261
 - 19.5 Untersuchung der Ergebnisse 263
 - 19.6 Verquickte epidemiologische Daten 265
 - 19.7 Mehrfachtests bei Fragebogendaten 270
 - 19.8 Kapitelzusammenfassung 271

Teil 6 Fazit

- 20 Schlussfolgerungen 275**
- 20.1 Machen Sie sich das Ziel des Versuchs klar 275
 - 20.2 Bauen Sie den Versuch einfach und damit klar und aussagekräftig auf 276
 - 20.3 Planen Sie die statistischen Analysen schon als Teil des Versuchsdesigns und nicht erst auf den letzten Drücker 277
 - 20.4 Untersuchen Sie die Daten visuell, bevor Sie in die statistischen Tests einsteigen 279
 - 20.5 Hüten Sie sich vor Mehrfachtests 280
 - 20.6 Interpretieren Sie sowohl Signifikanz als auch Nicht-Signifikanz mit gebührender Sorgfalt 280

Stichwortverzeichnis 283

Vorwort

An wen richtet sich dieses Buch?

Zielpublikum sind eher die Statistik-Anwender als Statistiker

Statistik als solche ist langweilig, uninteressant und unverständlich. Zumindest, wenn man sie auf traditionelle Art betreibt. Das Fach selbst muss dagegen gar nicht so schlecht sein. Es hat sich nur nicht aus einer selbstgewählten Zeitschleife befreien können. Vor gut dreißig Jahren gab es noch keine allgemein zugänglichen Computer und Softwarepakete, die uns all die Rechnerei abgenommen haben. Daher steckte das Fach bis zum Hals in den „Wie“-Fragen – wie kann ich am einfachsten das Durchschnittsgewicht von 500 Kartoffeln berechnen? Wie findet man heraus, ob die Rübenernte tatsächlich zunimmt, wenn man ein paar Schaufeln Pferdemist auf den Acker bringt? Wie berechnet man, ob es wirklich einen Zusammenhang zwischen der Regenmenge und der Größe unserer Äpfel gibt? Zur damaligen Zeit hatten die Statistiker noch eine echte Entschuldigung dafür, dass ihre Lehrbücher vollgestopft mit detaillierten Anleitungen für Rechenverfahren waren. Doch, so darf man fragen, ist das heute noch immer eine Entschuldigung? Heute würde doch niemand, der bei klarem Verstand ist, einen komplexen statistischen Test von Hand durchrechnen. Und warum wollen sie uns immer noch beibringen, wie man das tut?

Natürlich braucht die Welt auch heute noch echte „Statistiker“, die die Untersuchungsverfahren pflegen und verbessern, und natürlich auch Menschen, die die genaue Funktionsweise solcher Verfahren verstehen. Die meisten von uns sind aber keine Statistiker als solche. Wir wollen nur statistische Verfahren anwenden, um ein bestimmtes Ziel zu erreichen. Wir sollten demnach zwischen „Statistikern“ und „Statistik-Anwendern“ unterscheiden. Das ist so ähnlich wie die Unterscheidung zwischen Kfz-Entwicklern und Autofahrern. Autos zu entwickeln ist ein ehrenwerter und notwendiger Berufsstand, aber die meisten von uns wollen Autos einfach nur fahren. Glücklicherweise berücksichtigt die Literatur auf dem Gebiet der Kfz-Technik diesen Unterschied. Wenn ein Buch für Autofahrer geschrieben ist, steht einfach darin, wer schneller fahren will, muss das Gaspedal durchdrücken. Aber der Autor wird nicht versuchen, uns mit Details zur Ein-

spritzanlage und Gemischaufbereitung zu langweilen. In Statistikbüchern hat diese Logik leider noch nicht Einzug gehalten. Sie versuchen noch immer die Funktionsweise eines Kolmogorow-Smirnow-Tests *en détail* einem Publikum zu erklären, das sich dafür nicht im geringsten interessiert.

Nun, dann gibt es gute Nachrichten: Dieses Buch behandelt die statistischen Rechnungen fröhlich als Blackbox. Es erklärt, was man in die Blackbox hineinsteckt und was am anderen Ende wieder herauskommt. Sie können dabei aber völlig ahnungslos bleiben, was im Inneren der Box abgeht, so wie Sie beim Einparken auch nicht über die Funktionsweise einer Servolenkung nachdenken. Bitte betrachten Sie diesen Ansatz nicht als Faulheit oder als negativ. Aber indem wir all das unnötige Beiwerk abstreifen, können wir uns auf die wirklich interessanten Punkte konzentrieren. Darum wird dieses Buch versuchen, genau die Themen zu behandeln, die Statistik-Anwender verstehen müssen:

- Warum sind statistische Verfahren überhaupt nötig?
- Wie kann die Statistik bei der Versuchsplanung helfen?
- Welche Verfahren sollte ich anwenden, um meine Ergebnisse zu untersuchen?
- Was besagen die statistischen Ergebnisse wirklich, wenn ich sie habe?

Wer sind diese „Statistik-Anwender“?

Der Leserkreis, an den dieses Buch sich richtet, sind die Tausende von Menschen, die statistische Verfahren anwenden müssen, aber keinerlei Ehrgeiz haben, Statistiker zu werden, weil sich ihre wissenschaftliche Vorliebe auf ganz andere Themen richtet. Eine ganze Reihe von Studiengängen – von Pharmakologie und Botanik über Volks- und Betriebswirtschaft bis zur Psychologie – enthält gewisse Elemente der Statistik. Die Studenten dort (und wenn ich „Studenten“ sage, meine ich sowohl männliche als auch weibliche) müssen zumindest die grundlegenden statistischen Verfahren lernen. Viele von uns sind in der Forschung oder in der Industrie beschäftigt. Einige müssen ihre eigenen statistischen Untersuchungen durchführen, andere können auf die Hilfe professioneller Statistiker zurückgreifen. Doch selbst wenn Profis zur Hand gehen, stellt sich das Problem der Kommunikation. Wenn Sie nicht einmal die Bedeutung der Wörter kennen, werden Sie große Schwierigkeiten haben, einem Statistiker genau zu erklären, was Sie eigentlich tun wollen. Mein Ziel ist es, dass alle Personen aus diesen Gruppen dieses Buch nützlich finden.

Wenn Sie Student der Statistik oder gar professioneller Statistiker sind, kann ich Ihnen nur raten, das Buch schnell wieder wegzulegen. Sie könnten die herablassende Haltung gegenüber den mathematischen Grundlagen Ihres Fachs stark irritierend finden. Eigentlich wäre ich sogar ziemlich enttäuscht, wenn sich nicht wenigstens ein „traditioneller Statistiker“ bitterlich über dieses Buch beklagen würde.

Für welche Disziplinen ist dieses Buch relevant?

Alle behandelten statistischen Verfahren und Tests werden mit praktischen Beispielen und Datensätzen illustriert. Diese Beispiele stammen aus den pharmazeutischen Wissenschaften, wie der Buchtitel es ja andeutet. Doch ziemlich alle beschriebenen Verfahren und Prinzipien sind natürlich auch für verwandte wissenschaftliche Disziplinen wie Medizin, Biologie, Biomedizin und Chemie interessant.

Auf welchem Niveau bewegt sich das Buch?

Das Buch richtet sich an Studenten vor dem ersten Abschluss (Bachelor) bzw. vor dem Vordiplom, an ihre Dozenten und an weniger erfahrene Forscher.

Die ersten Kapitel (1–5) sind recht grundlegend. Sie behandeln die Datenbeschreibung (Mittelwert, Median, Modalwert, Standardabweichung und Quartile) und führen das Problem ein, die Unsicherheit aufgrund des Stichprobenfehlers zu beschreiben (SEM und 95 %-Konfidenzintervall für den Mittelwert). In der Theorie sollte viel davon schon aus der Schule vertraut sein, doch nach meiner Erfahrung scheitern viele junge Studenten schon daran, (beispielsweise) den Median eines kleinen Datensatzes zu berechnen. Diese Kapitel sind also für Anfänger besonders interessant, entweder zum Lernen oder zum Wiederholen.

Die Kapitel 6–17 behandeln dann die am häufigsten verwendeten statistischen Testverfahren. In den meisten Studiengängen kommen diese Themen schon in den ersten Semestern. Dabei folgt der Zugang nicht den traditionellen Mustern, bei denen einer breiten Auswahl von Verfahren jeweils gleiches Gewicht eingeräumt wird. Da sich das Buch eher auf verschiedene Gesichtspunkte rund um das statistische Testen als auf die Rechenverfahren konzentriert, werden alle relevanten Aspekte mit einem einzigen Testverfahren – dem doppelten t -Test – illustriert (Kapitel 6–11). Die folgenden Kapitel (12–17) befassen sich dann etwas kürzer gefasst mit weiteren Testverfahren und greifen immer wieder auf die dann eingeführten Grundlagen zurück.

Die Abschlusskapitel (18 und 19) behandeln einige reale Probleme, mit denen die Studenten während ihrer eigenen Forschungsarbeiten bei der Abschlussarbeit in Berührung kommen werden. Es geht dabei um Mehrfachtests (in studentischen Projekten allzu häufig!) sowie das Design und die Auswertung von Fragebögen. Dieser letzte Punkt führt zwar keine grundlegend neuen Konzepte ein, die Verwendung von Fragebögen hat aber derart zugenommen, dass es mir sinnvoll erschien, alle relevanten Punkte an einer einzigen Stelle zusammenzufassen.

Schlüsselbegriffe und Machiavellis Ratschläge

Schlüsselbegriffe

Im ganzen Buch werden Sie immer wieder Kästen mit wichtigen Begriffen oder Konzepten finden. Sie sehen etwa folgendermaßen aus:



Anteile von Einzelwerten innerhalb gegebener Bereiche

Für Daten, die einer Normalverteilung gehorchen, gilt:

- Gut zwei Drittel der Einzeldaten haben einen Wert innerhalb einer Standardabweichung vom Mittelwert (Mittelwert ± 1 SD),
- etwa 95 Prozent der Einzeldaten haben einen Wert innerhalb von zwei Standardabweichungen vom Mittelwert (Mittelwert ± 2 SD).

Die Kästen bringen niemals neue Informationen, sie dienen nur der Zusammenfassung und Betonung.

Machiavellis Rat

Sie finden auch Kästen mit Ratschlägen wie diesen hier:



Machiavelli rät:

Wechseln Sie einfach zu einem einseitigen Test, nachdem Sie die Ergebnisse angeschaut haben

Noch heute handelt es sich um die beste und wahrscheinlich am meisten verbreitete statistische Kungelei. Sehr schlagkräftig, nicht nachzuweisen – also einfach klasse!

Sie haben einen Versuch ausgeführt und die Ergebnisse bereits mit Ihrem üblichen zweiseitigen Test geprüft. Die Ergebnisse verfehlen die Signifikanzschwelle. Doch es gibt eine ganz einfache Lösung, die garantiert jederzeit funktioniert. Wiederholen Sie die Datenauswertung, diesmal aber mit einem einseitigen Test auf eine Änderung in eben die Richtung, die von den Ergebnissen nach ihren Voruntersuchungen schon nahegelegt werden. Wenn p ursprünglich irgendwo in dem Bereich zwischen 0,05 und 0,1 lag, wird dieser Trick den p -Wert genau halbieren, und damit ist die Signifikanz sichergestellt.

Solange die Zeitschriften nicht die Kurve kriegen und endlich darauf bestehen, dass die Autoren ihre Pläne schon im Voraus hinterlegen, gibt es keine Möglichkeit, diese ausgezeichnete Schummerei nachzuweisen. Sie brauchen nur einen vernünftigen Grund dafür, dass Sie „schon immer“ einen einseitigen Test auf diese bestimmte Richtung durchführen wollten – und dann kommen Sie garantiert damit durch.

Die zynisch formulierten Kästen sollen Sie natürlich nicht wirklich zu einem Missbrauch der Statistik ermuntern, sondern sollen Ihnen die Missbrauchsmöglichkeiten aufzeigen, mit denen andere Ihnen vielleicht etwas unterjubeln wollen. Gefahr erkannt – Gefahr gebannt! Wie groß die Gefahr ist, zeigt die Anzahl der Ausrufezeichen am Rand:

! Kleinere Sache. Der Missbrauch ist leicht zu entdecken oder hat nur geringes Täuschungspotenzial.

!! Mittlere Gefahr. Der gut informierte Leser (also Sie als Leser dieses Buchs) kann die versuchte Täuschung sofort aufdecken.

!!! Schwere Täuschung. Ein besonders arglistiger Trick, den selbst der bestinformierte Lester vielleicht vermuten, aber nie beweisen kann.

Fiktive Daten

Alle in diesem Buch beschriebenen Versuche und die dazugehörigen Daten sind frei erfunden. Aufbau und Ergebnisse sind aber immer realistisch. In einigen Fällen können sowohl der Aufbau eines Versuchs als auch die daraus folgende Datenanalyse etwas einfacher sein als die Gegebenheiten, die sich uns in der realen Welt stellen. Hier schien aber Klarheit wichtiger als strikter Realismus.

An einigen Stellen werden Einschätzungen abgegeben, ab welchem Betrag ein bestimmter gemessener Endpunkt bzw. dessen Änderung praktische Bedeutung hat. Diese Werte sind völlig willkürlich nach Kriterien gewählt, die mir als Autor realistisch erschienen. Ich hoffe, dass ich dabei nicht völlig falsch liege. Auf jeden Fall sollten Sie keinen dieser Werte als fachlich begründet ansehen.

Kurzzusammenfassung des Buchs

Dieses Buch richtet sich an Leser, die Statistik anwenden müssen, aber keine Ambitionen haben, zu eigentlichen Statistikern zu werden. Es verzettelt sich nicht in Rechenverfahren und konzentriert sich stattdessen auf wichtige Themen rund um die Datenerzeugung und -untersuchung (Abschätzung des Stichprobenumfangs, Interpretation der statistischen Ergebnisse, Gefahren von Mehrfachtests, möglicher Missbrauch usw.). In Zeiten gut zugänglicher Computer-Statistikpakete sind nämlich genau solche Themen das Problem, nicht mehr die Rechnung als solche.

Die erläuternden Beispiele des Buchs stammen aus dem pharmazeutischen Umfeld. Demzufolge sollten Studenten (und Arbeitskräfte) etwa aus den Bereichen Pharmazie, Pharmakologie und klinischer Pharmazie mit dem Material vertraut sein. Die betrachteten Themen jedoch spielen in den meisten wissenschaftlichen Disziplinen eine Rolle und sollten darum auch von jedermann aus einem Nachbarggebiet verständlich sein, selbst wenn die Beispiele nicht unmittelbar bekannt sind. Der Stoff wird in aufsteigender Schwierigkeit angeordnet:

- Die ersten Kapitel eignen sich für Studenten am Anfang ihrer Ausbildung. Der Stoff ist sehr grundlegend, besonders wird der zufällige Stichprobenfehler betont.
- Die nächsten Abschnitte behandeln Schlüsselkonzepte, mit denen die Studenten im weiteren Lauf ihres Studiums in Berührung kommen.
- Die letzten paar Kapitel werden wahrscheinlich am Schluss des Studiums, während der Forschungen für die Abschlussarbeit, von besonderem Nutzen sein. Dazu gehören der Entwurf und die Auswertung von Fragebögen.

Dieses Buch ist nicht an ein bestimmtes Statistikpaket gebunden. Mit den gegebenen Anweisungen sollte man seine Daten in jedes Statistikpaket eingeben und die wichtigen Teile der Ausgabe finden können. Spezielle Anweisungen für die Ausführung der besprochenen Verfahren mithilfe von Minitab oder SPSS finden sich auf der mit diesem Buch verbundenen (englischsprachigen) Website (www.staff.ljmu.ac.uk/phaprowe/pharmstats.htm).

Teil 1
Datentypen

1

Datentypen

► Dieses Kapitel ...

- legt ein System zur Beschreibung verschiedener Datentypen dar;
- erklärt, warum wir die Datentypen genau bestimmen müssen, mit denen wir zu tun haben.

1.1

Kommt es wirklich darauf an?

Ein Statistikbuch aufzuschlagen und gleich mit einer Diskussion anzufangen, in welcher Weise man Daten in verschiedene Typen einteilen kann, das klingt entsetzlich abgehoben. Dennoch besteht der erste Schritt für die Behandlung von Daten im Allgemeinen in der Bestimmung, mit welchen Datentypen wir überhaupt zu tun haben. Das mag trocken sein, aber es hat reale Auswirkungen.

Wir schauen uns drei Arten von Daten an. Alle laufen in der Fachliteratur unter verschiedenen Bezeichnungen. Ich habe hier Namen ausgewählt, die ich am ehesten für selbsterklärend halte, und strebe kein einheitliches Benennungssystem an. Ich werde folgende drei Begriffe verwenden:

- Intervallskala – Daten aus stetigen Messwerten,
- Ordinalskala – Daten aus verschiedenen geordneten Kategorien,
- Nominalskala – Daten aus verschiedenen Kategorien.

1.2

Daten auf einer Intervallskala

Die ersten zwei Datentypen, die wir hier diskutieren, hängen beide mit der Messung eines bestimmten Merkmals zusammen. Daten auf einer „Intervallskala“ (man findet auch die Bezeichnung „Proportionalskala“, obwohl es da streng genommen noch einen kleinen Unterschied gibt, auf den wir hier aber nicht eingehen müssen) werden durch eine stetige Messung gewonnen. Sie sind die aussage-

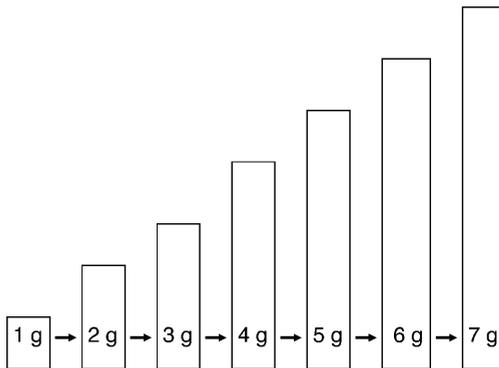


Abbildung 1.1 Daten auf einer Intervallskala – eine Reihe von Massen (1–7 g)

kräftigsten Daten, die man im Labor erzeugen kann, beispielsweise Gewichte, Längen, Zeiten, Konzentrationen, Drücke usw. Stellen Sie sich beispielsweise vor, wir hätten eine Reihe von Objekten mit einer Masse von 1 g, 2 g, 3 g usw. bis 7 g (Abbildung 1.1).

Nun stellen Sie sich die Massendifferenzen vor, die beim Übergang von einem Objekt zum nächsten auftreten. Diese Schritte betragen jeweils eine Einheit auf der Skala und haben folgende Eigenschaften:

- *Die Schritte haben jeweils eine exakt definierte Größe.* Wenn Sie jemandem davon erzählen, dass Sie eine Reihe von Objekten, wie oben beschrieben, haben, weiß Ihr Gegenüber genau, wie groß die Massendifferenzen sind, wenn wir die ganze Messreihe durchgehen.
- *Alle Schritte sind genau gleich groß.* Die Massendifferenz zwischen den Objekten mit 1 g und 2 g ist genauso groß wie der Schritt von 2 zu 3 g oder von 6 zu 7 g usw.

Weil die Messungen Schritte von konstanter Größe zeigen (nämlich Intervalle), spricht man hier von einer Intervallskala. Obwohl die Messwerte in Abbildung 1.1 exakt ganzzahlig sind, könnten sie natürlich genauso gut auch beliebige Werte dazwischen (wie 1,5 g oder 3,175 g) annehmen. Daher nennt man die Maßskala auch „stetig“ oder „kontinuierlich“.

1.3

Daten auf einer Ordinalskala

Auch hier geht es um Messungen, aber die erhobenen Kennzahlen sind meist etwas subjektiver als im vorigen Fall. Es ist schön, wenn man objektive Werte messen kann, wie den Blutdruck oder die Körpertemperatur. Es ist aber ebenso legitim, beispielsweise eine Vorstellung davon zu gewinnen, wie ein Patient seinen Zustand nach einer Behandlung einschätzt. Besonders naheliegend ist es, ein Punkte- oder Notenschema zu verwenden, beispielsweise von -2 bis $+2$ mit den folgenden Einschätzungen:

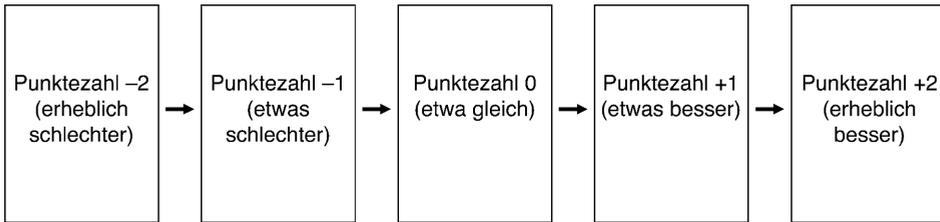


Abbildung 1.2 Daten auf einer Ordinalskala – Punkteschema für die Selbsteinschätzung des Gesundheitszustands von Patienten nach der Behandlung

-2	merklich schlechter
-1	etwas schlechter
0	etwa gleich
+1	etwas besser
+2	erheblich besser

In diesem Fall (Abbildung 1.2) wissen wir nur, dass ein Patient mit einem höheren Wert zufriedener ist mit seiner Behandlung. Wir wissen aber nicht, *um wie viel* zufriedener er ist.

Wir wissen nicht, wie groß die Abstände zwischen den einzelnen Einschätzungen sind; also können wir auch nicht behaupten, sie seien alle gleich groß. Es muss noch nicht einmal so sein, dass die Differenz zwischen den Einschätzungen -2 und 0 größer ist als der zwischen +1 und +2. Keine der Eigenschaften aus einer Intervallskala lässt sich also auf diese Daten übertragen.

Der Begriff „Ordinal“ spiegelt wider, dass die verschiedenen Ergebnisse sich in einer Rangfolge ordnen lassen, von einem Extremwert zum anderen. Daten auf einer Ordinalskala werden daher manchmal auch als „kategorial geordnet“ bezeichnet. In diesem Fall sind die Werte nicht stetig, d. h., die einzelnen Kategorien werden mit -1, +2 usw. bezeichnet, Zwischenwerte gibt es nicht.

1.4

Daten auf einer Nominalskala

In diesem Fall geht es in keinem Fall um die Messung eines Merkmals. Bei diesen Daten verwenden wir eine Einteilung ohne natürliche Rangfolge. Beispielsweise könnte einer der Faktoren, der die Effektivität einer Behandlung beeinflusst, der Hersteller des entsprechenden medizinischen Geräts sein. Die Patienten würde man dann nach den Herstellern „Müller“, „Meyer“ und „Schmidt“ einteilen. Hier gibt es keine natürliche Reihenfolge, es handelt sich nur um drei verschiedene Bezeichnungen.

Bei Ordinaldaten konnten wir wenigstens sagen, dass beispielsweise ein mit +2 bewerteter Fall eher dem Fall mit +1 als einem mit 0 oder -1 ähnelt. Bei Nominaldaten können wir aber nicht davon ausgehen, dass die „Müller“- und die

„Meyer“-Patienten eine irgendwie geartete Ähnlichkeit aufweisen. Die Reihenfolge, in der man sie aufführt, ist völlig beliebig.

Sehr verbreitet sind Einteilungen mit genau zwei Kategorien, etwa männlich/weiblich, lebt/tot oder Erfolg/Misserfolg. In solchen Fällen spricht man von „dichotomen Werten“.



Datentypen

- *Intervallskala* – Messungen mit definierten und stets gleich bleibenden Abständen (Intervallen) zwischen aufeinanderfolgenden Werten. Die Werte sind stetig.
- *Ordinalskala* – Messungen mithilfe einer Klassifikation und einer natürlichen Reihenfolge der Werte (vom niedrigsten bis zum höchsten), aber ohne bestimmte Abstände. Die Werte sind nicht stetig.
- *Nominalskala* – Einteilungen ohne natürliche Reihenfolge

1.5

Aufbau dieses Buchs

Das Buch ist so aufgebaut, dass nacheinander die verschiedenen Datentypen durchgenommen werden. Die Kapitel 2–14 behandeln Daten auf Intervallskalen, die bei stetigen Messungen gewonnen werden. Die Kapitel 15 und 16 befassen sich mit kategorialen Daten (auf Nominalskalen), geordnete Daten (auf Ordinalskalen) werden in Kapitel 17 besprochen.

1.6

Kapitelzusammenfassung

Der unerlässliche erste Schritt bei der Auswahl der passenden statistischen Verfahren ist es, die zu behandelnden Datentypen zu erkennen.

Folgende Fälle können auftreten:

- *Intervallskala* – Messungen auf einer Skala mit definierten und stets gleich bleibenden Abständen. Die Werte sind stetig.
- *Ordinalskala* – Messungen mithilfe einer Klassifikation ohne bestimmte Abstände. Die Werte sind nicht stetig.
- *Nominalskala* – Einteilungen ohne natürliche Reihenfolge.

Teil 2

Daten auf Intervallskalen

2

Beschreibende Statistik

► Dieses Kapitel ...

- legt dar, wie sich mithilfe von *Mittelwert*, *Median* und *Modalwert* beschreiben lässt, wie klein oder wie groß ein Satz von Werten ist, und prüft, welche Beschreibung jeweils am besten geeignet ist;
- beschreibt, wie sich mithilfe der *Standardabweichung* veranschaulichen lässt, wie stark die Werte innerhalb eines Datensatzes schwanken;
- zeigt, wie sich mithilfe von *Quartilen* ähnliche Informationen wie in den ersten beiden Punkten ableiten lassen, selbst wenn es extreme Ausreißer gibt.

2.1

Zusammenfassung von Datensätzen

Experimente und Versuchsreihen erzeugen oft Datensätze, die zu lang sind, als dass man sie einfach erfassen könnte. Wir brauchen ein oder zwei Werte zur Zusammenfassung, die dem Betrachter ein genaues Bild der Gesamtlage vermitteln.

Bei Daten auf Intervallskalen (kontinuierliche Messung) gibt es zwei Gesichtspunkte, die wir mit solchen zusammenfassenden Werten möglichst beschreiben wollen:

- Wie groß sind sie?
- Wie stark schwanken sie?

Zur Antwort auf die erste Frage brauchen wir ein „zentrales Lagemaß“ (zentraler Lageparameter), für die Antwort auf die zweite ein „Streuungsmaß“ (Streuparameter).

In diesem Kapitel schauen wir uns mehrere Möglichkeiten an, wie man das bewerkstelligen kann. Es wäre falsch zu behaupten, dass eine dieser Möglichkeiten immer besser wäre als eine andere. Wir können aber sehr wohl rational begründen, warum wir eine bestimmte dieser Möglichkeiten auswählen, wenn wir die Art der Daten und den beabsichtigten Zweck berücksichtigen.



Beschreibende Statistik

- Zentrale Lagemaße: Wie groß sind die Werte?
- Streuungsmaße: Wie stark schwanken die Werte?

2.2

Zentrale Lagemaße – der Mittelwert, der Median und der Modalwert

Der Begriff „zentrales Lagemaß“ beschreibt eine beliebige statistische Größe, mit der man die „Mitte“ eines Datensatzes angibt, um die herum die anderen Werte sich häufen. Im allgemeinen Gebrauch werden drei solche Indikatoren verwendet – der Mittelwert, der Medianwert und der Modalwert.

2.2.1

Der Mittelwert – zehn Chargen eines Impfstoffs

Der übliche Ansatz, die zentrale Lage von einem Datensatz zu beschreiben, ist es, den Durchschnitt zu bilden. In akademischen Kreisen ist es allerdings unüblich, den allgemein bekannten Begriff „Durchschnitt“ zu verwenden, daher redet man lieber vom „Mittelwert“.

Unser erster Beispieldatensatz beschreibt mehrere Chargen eines Impfstoffs. Jede Charge soll dieselbe Wirkungsstärke haben, allerdings sind Abweichungen bei der Herstellung unvermeidlich. Eine Anzahl von zehn Chargen wurde analysiert, die Ergebnisse sind in Tabelle 2.1 zu sehen.

Die Summe aller Wirkungsstärken (in Einheiten/ml) beträgt 991,5. Dividiert man dies durch die Anzahl der untersuchten Chargen (in diesem Fall zehn), so erhält man einen Durchschnitts- bzw. Mittelwert von 99,15 Einheiten/ml.

Die Rechnung lässt hier keine ernsthaften Zweifel zu, aber wir müssen uns immer klar machen, ob der erhaltene Zahlenwert dem Leser die richtige Botschaft

Tabelle 2.1 Wirkungsstärken von zehn Chargen eines Impfstoffs (in Einheiten/ml)

Wirkungsstärke (Einheiten/ml)
106,6
97,9
102,3
95,6
93,6
95,9
101,8
99,5
94,9
103,4
Mittelwert = 99,15