

Walter Krämer
Claus Weihs *Hrsg.*

Faszination Statistik

Einblicke in aktuelle
Forschungsfragen und Erkenntnisse



Springer Spektrum

Faszination Statistik

Walter Krämer · Claus Weihs
(Hrsg.)

Faszination Statistik

Einblicke in aktuelle
Forschungsfragen und Erkenntnisse

 Springer Spektrum

Hrsg.
Walter Krämer
Fakultät Statistik
Technische Universität Dortmund
Dortmund, Deutschland

Claus Weihs
Fakultät Statistik
Technische Universität Dortmund
Dortmund, Deutschland

ISBN 978-3-662-60561-5 ISBN 978-3-662-60562-2 (eBook)
<https://doi.org/10.1007/978-3-662-60562-2>

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

Springer Spektrum

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2019

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Iris Ruhmann

Springer Spektrum ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Vorwort

“I keep saying the sexy job in the next ten years will be statisticians ... The ability to take data - to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it is going to be a hugely important skill in the next decades.”
(Hal Varian, Chefökonom von Google; das Zitat ist von 2009¹, die zehn Jahre sind um. Aber wir glauben, jetzt geht es erst richtig los.)

Auch Wissenschaften kennen Konjunktur. Einige, wie die Medizin, erfreuen sich aus offensichtlichen Gründen eines konstant stabilen Interesses. Auch die Philosophie sonnt sich seit Plato und Aristoteles in einem durch weitgehendes Unverständnis befeuerten Wohlwollen der Zeitgenossen gleich welchen Jahrhunderts. Andere Wissenschaften dagegen rücken eher sporadisch in den Fokus der allgemeinen Aufmerksamkeit. So war etwa die erste Hälfte des 20. Jahrhunderts die hohe Zeit der Chemie, der Technik und der Physik: Kunstdünger, Flugzeuge, Farbstoffe, Elektrifizierung, Raketen oder die Kernspaltung wurden auch von Nicht-Experten heftig diskutiert, mit Konrad Röntgen gab es den ersten wissenschaftlichen Pop-Star der Geschichte überhaupt. Nach Sigmund Freud war eine gewisse Zeit sogar die Psychologie in gewissen Kreisen das Partythema Nummer 1. Wer nicht von seinem letzten Besuch bei einem Psychoanalytiker erzählen konnte, war sozusagen sozial deklassiert. Und im Kielwasser der Studentenrevolte Ende der 1960er Jahre galt eine kurze Zeit die Soziologie als die Schlüsselwissenschaft, an welcher die Zukunft der Menschheit abzulesen sei.

Und im Moment ist die Statistik angesagt. Angefeuert durch eine gewaltige Doppelrevolution bei der Datenverarbeitung und der Datengewinnung gleichermaßen eröffnen sich hier neue Chancen und Herausforderungen fast jeden Tag, kommt auch der oberflächlichste Zeitungsläser nicht um die Erkenntnis herum, dass Datenwissenschaft im Allgemeinen und Statistik im Besonderen zu den zentralen Ingredienzien einer erfolgreichen Fortführung der Menschheitsgeschichte zählt. Deswegen reden heute alle von „Big Data“, und die einschlägigen Studiengänge sprießen wie

¹ s. <https://flowingdata.com/2009/02/25/googles-chief-economist-hal-varian-on-statistics-and-data/>, besucht 11.9.2019

Pilze aus dem Boden. Vergessen wird dabei gerne, dass der korrekte Umgang mit Daten, das Ziehen von Stichproben, das Entscheiden unter Unsicherheit, das Trennen von Zufall und System und das Hochrechnen vom Kleinen auf das Große keine neuen Erfindungen sind; seit Jahrzehnten befassen sich damit Tausende von Forscherinnen und Forschern auf der ganzen Welt. Die aktuelle und diesen Forschern und Forscherinnen durchaus nicht unangenehme Aufregung entspringt vor allem der schieren Masse an neuen Daten und Informationen, die nach einer statistischen Auswertung verlangen.

Diese Daten und Informationen gab es aber schon immer und nicht viel weniger als heute. Die Temperatur auf der Nordseite des Matterhorns am Weihnachtsabend existierte solange das Matterhorn existiert. Und dass Frau X am 10. Januar 1998 bei Aldi Süd in Mainz-Bretzenheim zwei Flaschen Rotwein Chateau St. Chinian gekauft hat, war schon seinerzeit ein Fakt. Aber die Verbreitung dieser Information beschränkte sich auf Frau X und möglicherweise die junge Dame an der Kasse. Heute weiß es, sollte Frau X mit Kreditkarte bezahlt haben, im Prinzip die ganze Welt - das heute weltweit diskutierte Phänomen „Big Data“ betrifft nicht die Existenz, sondern die Verfügbarkeit von Daten. Und diese Verfügbarkeit nimmt durch das Zusammenwirken von immer effizienterer Rechner- und Speichertechnik auf der einen und immer ertragreicherer Datenfischerei auf der anderen Seite geradezu gigantische Ausmaße an. Über den größten Teil der Menschheitsgeschichte wussten wenige wenig, heute wissen viele viel und der Zug geht mit Volldampf in Richtung alle wissen - prinzipiell - alles.

Umso wichtiger wird es, mit diesen Daten vernünftig umzugehen. Dazu ist die Statistik heute genauso unverzichtbar wie sie es schon immer war. Natürlich bringt die aktuelle Mengenexplosion auch früher unbekannte logistische Probleme mit, aber die Statistik als Wissenschaft wird dadurch nicht umgewälzt. Die Grundprinzipien des Schätzens und Testens, der Modellwahl oder der Stichprobenerhebung gelten für die Daten von Google genauso wie für die vom Statistischen Bundesamt. Auch wenn die Informatik in den letzten Jahrzehnten neu dazugekommen ist, und ihre Rolle stetig wächst, bleibt die Statistik weiter die zentrale Anlaufstelle.

Der vorliegende Sammelband stellt ausgewählte Ergebnisse vor, die in den letzten Jahren an der Fakultät Statistik der TU Dortmund, der einzigen eigenständigen Statistik-Fakultät im ganzen deutschen Sprachgebiet, sowie darüber hinaus im Rahmen von an diese Fakultät angedockten DFG-Sonderforschungsbereichen entstanden sind, aktuell insbesondere im Sonderforschungsbereich „Statistik nichtlinearer dynamischer Systeme“. Dazu kommen einige zeitlose und unabhängige Einsichten, die uns helfen, mit Daten verschiedenster Art im Alltag besser zurechtzukommen. Die folgenden Seiten zeigen etwa, wie die Statistik bei klinischen Studien die Spreu vom Weizen trennt, wie man mit Statistik auch hörgeschädigten Menschen einen guten Musikgenuss verschafft, wie man Überschwemmungskatastrophen oder Risiken an der Börse statistisch modelliert und damit besser in den Griff bekommt, oder wie man aus qualitativen Information wie zum Beispiel Texten sinnvolle quantitative Daten extrahiert. Andere hier vorgestellte Anwendungen betreffen die Qualitätskontrolle in der Industrie, die statistische Problematik von Mietspiegeln, die Prognosefähigkeiten von Ratingagenturen oder die Vermeidung von Fehlalarmen in

der Intensivmedizin. Und wenn Sie wissen wollen, warum Lotto kein reines Glücksspiel ist, oder warum bei Pferdewetten die Favoriten systematisch zu hoch bewertet werden, sind Sie hier ebenfalls an der richtigen Stelle.

Viele dieser Erkenntnisse sind noch nicht in Lehrbüchern nachzulesen, wir berichten hier sozusagen von der Forschungsfront. Dennoch haben sich die Herausgeber mit ihren mitmachenden Kolleginnen und Kollegen geeinigt, auf Fachjargon möglichst zu verzichten und für ein Publikum zu schreiben, das zwar selbstständig denken, aber nicht die letzten Feinheiten einer formalstatistischen Analyse nachvollziehen will. Wir hoffen deshalb, dass auch diejenigen an diesem Buch Gefallen finden, deren Enthusiasmus für die Statistik im Rahmen eines Studiums der Betriebs- oder Volkswirtschaftslehre, der Psychologie oder Soziologie oder anderer Fächer, wo traditionsgemäß Statistikscheine zu erwerben sind, etwas gelitten hat. Leider wird in der akademischen Ausbildung die intrinsische Schönheit der Statistik durch einen gewissen Formalismus oft verdeckt. Mit diesem Sammelband blasen wir diese formalen Rauchschwaden sozusagen weg, und laden ein zu einem unverdeckten Blick auf eine durch und durch faszinierende Wissenschaft.

Dortmund,
im August 2019

Walter Krämer
Claus Weihs

Danksagung

Diese Arbeit wurde unterstützt durch die Sonderforschungsbereiche „Komplexitätsreduktion in Multivariaten Datenstrukturen“ (SFB 475), „Statistik nichtlinearer dynamischer Systeme“ (SFB 823) sowie „Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung“ (SFB 876) der Deutschen Forschungsgemeinschaft (DFG). Das gilt insbesondere für die Kapitel [3](#), [4](#), [10](#), [13](#), [14](#), [15](#), [16](#), [18](#), [19](#), [20](#), [21](#) und [26](#).

Danken möchten wir außerdem Robert Löser, Lilia Michailov, Jennifer Neuhaus-Stern und Simon Neumärker für ihre Unterstützung beim Korrekturlesen und Editieren der Texte.

Inhaltsverzeichnis

Vorwort	v
Danksagung	ix
Inhaltsverzeichnis	xi
Autoren	xvii
Notation	xxi

Teil I Leben und Sterben **1**

1	Warum leben Novemberkinder länger?	
	<i>Walter Krämer, Katharina Schüller</i>	3
1.1	Was sagt uns das Datum der Geburt?	3
1.2	Daten und Fakten	4
1.3	Auf der Suche nach den Gründen	5
1.4	Die Bedeutung der Daten	8
1.5	Literatur	10
2	Wo wirken Medikamente im Körper? Eine systematische statistische Datenanalyse	
	<i>Claus Weihs</i>	11
2.1	Pharmakokinetik und Vorklinik	11
2.2	Standardvorgehen bei der statistischen Datenanalyse	12
2.3	Die Verteilung von Medikamenten im Körper	13
2.4	Literatur	18
3	Medikamentenstudien: Mit Statistik zur optimalen Dosis	
	<i>Holger Dette, Kirsten Schorning</i>	19
3.1	Die drei klinischen Testphasen	19
3.2	Die Optimierung von Phase 2	20
3.3	Auf den Versuchsplan kommt es an	22
3.4	Auf dem Weg zur praktischen Anwendung	23
3.5	Literatur	24

4	Statistische Alarmsysteme in der Intensivmedizin	
	<i>Roland Fried, Ursula Gather, Michael Imhoff</i>	25
4.1	Alarmer in der medizinischen Akutversorgung	25
4.2	Glättung als Teil der Datenvorverarbeitung	27
4.3	Gemeinsame Analyse der Merkmale	30
4.4	Validierung der Ergebnisse	32
4.5	Literatur	33
5	Personalisierte Medizin: Wie Statistik hilft, nicht in der Datenflut zu ertrinken	
	<i>Jörg Rahnenführer</i>	34
5.1	Genetik	34
5.2	Wirksamkeit und Nebenwirkungen	35
5.3	Genetische Muster	36
5.4	Statistische Kniffe	36
5.5	Medizinische Anwendung	38
5.6	Krankheitsfortschritt	38
5.7	Zusammenfassung	40
5.8	Literatur	40
6	Mit Statistik dem Wirken der Gene auf der Spur	
	<i>Silvia Selinski, Katja Ickstadt, Klaus Golka</i>	41
6.1	Umwelt, Krankheiten und Gene	41
6.2	Epidemiologie und Genetik	42
6.3	Wechselwirkungen zwischen Umwelt und Genetik	46
6.4	Fazit	48
6.5	Literatur	49
7	Statistik und die maximale Dauer eines Menschenlebens	
	<i>Jan Feifel, Markus Pauly</i>	50
7.1	Hintergrund	50
7.2	Über den Durchschnitt zur Extremwerttheorie	51
7.3	Herausforderungen in demographischen Daten	54
7.4	Ergebnisse	55
7.5	Fazit	56
7.6	Literatur	56
	Teil II Sport, Spiel und Freizeit	57
8	Statistik und Fußball	
	<i>Andreas Groll, Gunther Schauburger</i>	59
8.1	Mehr Tore mit Statistik	59
8.2	Ein statistisches Modell für die Tore	59
8.3	Einflussgrößen	62
8.4	Fazit	66
8.5	Literatur	66

9 Die Angst der Spieler beim Elfmeter: Welcher Schütze und welcher Torwart sind die Besten?

Peter Gnändinger, Leo N. Geppert, Katja Ickstadt 67

9.1 Elfmeter im Fußball 67

9.2 Einflussgrößen auf den Erfolg bei einem Elfmeter 69

9.3 Wer ist der Beste? 71

9.4 Fazit und Ausblick 74

9.5 Literatur 74

10 Musikdatenanalyse

Claus Weihs 75

10.1 Was ist das, Musik? 75

10.2 Musikdaten 77

10.3 Erkenntnisse 79

- Tonhöhen 79

- Instrumente 81

- Einsatzzeiten 82

- Automatische Vernetzung 83

- Genres 83

10.4 Literatur 84

11 Statistik und Pferdewetten – Favoriten vs. Außenseiter

Martin Kukuk 85

11.1 Pferdewetten 85

11.2 Wett auszahlungen 86

11.3 Erklärungen für den Außenseitereffekt 87

11.4 Außenseitereffekt durch subjektive Schätzungen 88

11.5 Fazit 90

11.6 Literatur 90

12 Die Statistik des Lottospiels

Walter Krämer 91

12.1 Lotto als Anlagestrategie 91

12.2 Die Optimierung der Quoten 93

12.3 Literatur 94

Teil III Geld und Wirtschaft **95**

13 Mit Statistik an die Börse

Walter Krämer, Tileman Conring 97

13.1 Achtung Abhängigkeiten 97

13.2 Investieren in Aktien 98

13.3 Zeitvariable Abhängigkeiten 99

13.4 Pro und Kontra Normalverteilung 100

13.5 Kointegration 103

13.6 Literatur 104

14	Statistik bei der Risikobewertung von Bankenportfolios	
	<i>Dominik Wied, Robert Löser</i>	105
	14.1 Das Problem	105
	14.2 Expected Shortfall im Vergleich zu Value-at-Risk	105
	14.3 Schätzung	107
	14.4 Validierung	108
	14.5 Literatur	111
15	Statistik in der Ratingindustrie	
	<i>Walter Krämer, Simon Neumärker</i>	112
	15.1 Schulden und Schuldner	112
	15.2 Wie beurteilt man die Qualität von Wahrscheinlichkeitsprognosen?	114
	15.3 Ein Zahlenbeispiel	116
	15.4 Halbordnungen von Wahrscheinlichkeitsprognosen	117
	15.5 Skalarwertige Qualitätskriterien	118
	15.6 Literatur	119
16	Bruttoinlandsprodukt, Treibhausgase und globale Erderwärmung	
	<i>Martin Wagner, Fabian Knorre</i>	120
	16.1 Wirtschaftliche Aktivität und Emissionen	120
	16.2 Statistische Analyse des Zusammenhangs	122
	16.3 Parameterschätzung bei nichtlinearer Kointegration	125
	16.4 Interpretation	127
	16.5 Literatur	127
17	Ein wahres Minenfeld: Die statistische Problematik von Mietpreisspiegeln	
	<i>Walter Krämer</i>	128
	17.1 Zwei statistische Probleme	128
	17.2 Die Datenerfassung	129
	17.3 Die Berechnung der Nettomieten	130
	17.4 Die Bestimmung der Mietspiegelzellen	130
	17.5 Tabellen- versus Regressionsmietspiegel	131
	17.6 Die Problematik der Mietpreisspannen	133
	17.7 Literatur	135
Teil IV Natur und Technik		137
18	Hochwasserstatistik: Nahe am Wasser gebaut?	
	<i>Svenja Fischer, Roland Fried, Andreas Schumann</i>	139
	18.1 Fluten in den Griff bekommen	139
	18.2 Was ist ein Hochwasser?	140
	18.3 Hochwasserrisiko und -wahrscheinlichkeiten	141
	18.4 Robuste Schätzungen	144
	18.5 Hochwassertypen und Änderungen im Zeitverlauf	146
	18.6 Regionalisierung	147
	18.7 Literatur	148

19 Mit Statistik weniger Ausschuss
Claus Weihs, Nadja Bauer 149
 19.1 Ausschuss beim Tiefbohren 149
 19.2 Qualitätsverbesserung: Six Sigma 149
 - Problemdefinition (Define) 151
 - Gemessener Ausschuss (Measure) 151
 - Datenanalyse (Analyze) 153
 - Prozessverbesserung (Improve) 155
 - Prozesskontrolle (Control) 155
 19.3 Literatur 156

20 Statistik und die Zuverlässigkeit von technischen Produkten
Christine H. Müller 157
 20.1 Zuverlässigkeit und Zufall 157
 20.2 Einfache Lebensdauer-Analysen 158
 20.3 Lebensdauer-Analyse bei verschiedenen Belastungen 159
 20.4 Lebensdaueranalyse bei Produkten mit mehreren Komponenten . . . 160
 20.5 Prognoseintervalle 161
 20.6 Ausblick 162
 20.7 Literatur 163

21 Langlebige Maschinenteile: Wie statistische Versuchsplanung Verschleißschutz optimiert
Sonja Kuhnt, Wolfgang Tillmann, Alexander Brinkhoff, Eva-Christina Becker-Emden 164
 21.1 Beschichtungsprozesse 164
 21.2 Optimierung mit statistischer Versuchsplanung 165
 21.3 Herausforderungen im realen Spritzprozess 168
 21.4 Literatur 171

Teil V Messen und Vergleichen **173**

22 Das Unmessbare messen: Statistik, Intelligenz und Bildung
Philipp Doeblér, Gesa Brunn, Fritjof Freise 175
 22.1 Bildungstests und Bildung 175
 22.2 Latente Variablen und ihre Indikatoren 176
 22.3 Ein statistisches Modell für Lernverlaufsdiagnostik 177
 22.4 Von den Daten zu den latenten Variablen 178
 22.5 Literatur 181

23 Peinliche Wahrheiten zutage fördern mit Statistik
Andreas Quatember 182
 23.1 Die Methode der indirekten Befragung 182
 23.2 Eine Erweiterung 183
 23.3 Aufgaben für die Forschung 185
 23.4 Literatur 186

24 Stichproben und fehlende Daten	
<i>Andreas Quatember</i>	187
24.1 Stichproben in Theorie und Praxis	187
24.2 Statistische Reparaturmethoden	188
24.3 Literatur	190
25 Wer soll das alles lesen? Automatische Analyse von Textdaten	
<i>Jörg Rahnenführer, Carsten Jentsch</i>	191
25.1 Große Textsammlungen	191
25.2 Textanalysen in den Sozialwissenschaften	192
25.3 Vorverarbeitung von Textdaten	193
25.4 Thematische Einteilung	194
25.5 Unterschiede finden	195
25.6 Textanalyse von Wahlprogrammen	196
25.7 Zusammenfassung und Ausblick	198
25.8 Literatur	199
Teil VI Wo die Reise hinget	201
26 Ist Data Science mehr als Statistik? Ein Blick über den Tellerrand	
<i>Claus Weihs, Katja Ickstadt</i>	203
26.1 Data Science: Was ist das überhaupt?	203
26.2 Data Science: Schritte	204
- Datenerhebung und -anreicherung	205
- Datenexploration	207
- Statistische Datenanalyse	207
- Modellvalidierung und Modellauswahl	209
- Darstellung und Bericht	209
26.3 Schlussfolgerung	209
26.4 Literatur	210
Sachverzeichnis	211

Autoren

Nadja Bauer [Kapitel 19]

Fachhochschule Dortmund, Fachbereich Informatik,
nadja.bauer@fh-dortmund.de

Eva-Christina Becker-Emden [Kapitel 21]

FH Dortmund, Arbeitsgebiet Mathematische Statistik,
eva-christina.becker-emen@fh-dortmund.de

Alexander Brinkhoff [Kapitel 21]

TU Dortmund, Fakultät Maschinenbau,
alexander.brinkhoff@tu-dortmund.de

Gesa Brunn [Kapitel 22]

TU Dortmund, Statistik in den Sozialwissenschaften,
gesa.brunn@tu-dortmund.de

Tileman Conring [Kapitel 13]

tileman.conring@tu-dortmund.de

Holger Dette [Kapitel 3]

RUB, Lehrstuhl für Stochastik,
holger.dette@ruhr-uni-bochum.de

Philipp Doebler [Kapitel 22]

TU Dortmund, Statistik in den Sozialwissenschaften,
doebler@statistik.tu-dortmund.de

Jan Feifel [Kapitel 7]

Universität Ulm, Institut für Statistik,
jan.feifel@uni-ulm.de

Svenja Fischer [Kapitel 18]

RUB, Ingenieurhydrologie und Wasserwirtschaft,
svenja.fischer@ruhr-uni-bochum.de

Fritjof Freise [Kapitel 22]

TU Dortmund, Statistik in den Sozialwissenschaften,
freise@statistik.uni-dortmund.de

Roland Fried [Kapitel 4, 18]

TU Dortmund, Lehrstuhl Statistik in den Biowissenschaften,
fried@statistik.tu-dortmund.de

Ursula Gather [Kapitel 4]

TU Dortmund, Fakultät Statistik,
gather@statistik.tu-dortmund.de

Leo N. Geppert [Kapitel 9]

TU Dortmund, Lehrstuhl Mathematische Statistik und biometrische Anwendungen,
geppert@statistik.tu-dortmund.de

Peter Gnädinger [Kapitel 9]

peter.gnaendinger@tu-dortmund.de

Klaus Golka [Kapitel 6]

IfADo, Klinische Arbeitsmedizin,
golka@ifado.de

Andreas Groll [Kapitel 8]

TU Dortmund, Datenanalyse und Statistische Algorithmen,
groll@statistik.tu-dortmund.de

Katja Ickstadt [Kapitel 6, 9, 26]

TU Dortmund, Lehrstuhl Mathematische Statistik und biometrische Anwendungen,
ickstadt@statistik.tu-dortmund.de

Michael Imhoff [Kapitel 4]

RUB, Medizinische Fakultät,
mike@imhoff.de

Carsten Jentsch [Kapitel 25]

TU Dortmund, Lehrstuhl für Wirtschafts- und Sozialstatistik,
jentsch@statistik.tu-dortmund.de

Fabian Knorre [Kapitel 16]

TU Dortmund, Lehrstuhl Ökonometrie und Statistik,
knorre@statistik.tu-dortmund.de

Walter Krämer [Herausgeber, Kapitel 1, 12, 13, 15, 17]

TU Dortmund, Lehrstuhl für Wirtschafts- und Sozialstatistik,
walterk@statistik.tu-dortmund.de

Sonja Kuhnt [Kapitel 21]

FH Dortmund, Arbeitsgebiet Mathematische Statistik,
sonja.kuhnt@fh-dortmund.de

Martin Kukuk [Kapitel 11]

Universität Würzburg, Lehrstuhl für Ökonometrie,
martin.kukuk@uni-wuerzburg.de

Robert Löser [Kapitel 14]

TU Dortmund, Lehrstuhl für Wirtschafts- und Sozialstatistik,
robert.loeser@tu-dortmund.de

Christine H. Müller [Kapitel 20]

TU Dortmund, Lehrstuhl Statistik mit Anwendungen im Bereich der Ingenieurwissenschaften,
cmueller@statistik.tu-dortmund.de

Simon Neumärker [Kapitel 15]

TU Dortmund, Lehrstuhl für Wirtschafts- und Sozialstatistik,
simon.neumaerker@tu-dortmund.de

Markus Pauly [Kapitel 7]

TU Dortmund, Lehrstuhl Mathematische Statistik und industrielle Anwendungen,
markus.pauly@tu-dortmund.de

Andreas Quatember [Kapitel 23, 24]

Johannes Kepler Universität Linz, Institut für Angewandte Statistik,
andreas.quatember@jku.at

Jörg Rahnenführer [Kapitel 5, 25]

TU Dortmund, Statistische Methoden in der Genetik und Chemometrie,
rahenfuehrer@statistik.tu-dortmund.de

Gunther Schauburger [Kapitel 8]

TU München, Fakultät für Sport- und Gesundheitswissenschaften,
gunther.schauburger@tum.de

Kirsten Schorning [Kapitel 3]

RUB, Lehrstuhl für Stochastik,
kirsten.schorning@rub.de

Katharina Schüller [Kapitel 1]

Stat-Up Statistical Consulting & Data Science GmbH,
katharina.schueller@stat-up.com

Andreas Schumann [Kapitel 18]

RUB, Ingenieurhydrologie und Wasserwirtschaft,
andreas.schumann@ruhr-uni-bochum.de

Silvia Selinski [Kapitel 6]

IfADo, Toxikologie/Systemtoxikologie,
selinski@ifado.de

Wolfgang Tillmann [Kapitel 21]
TU Dortmund, Fakultät Maschinenbau,
wolfgang.tillmann@udo.edu

Martin Wagner [Kapitel 16]
TU Dortmund, Lehrstuhl Ökonometrie und Statistik,
mwagner@statistik.tu-dortmund.de

Claus Weihs [Herausgeber, Kapitel 2, 10, 19, 26]
TU Dortmund, Lehrstuhl Computergestützte Statistik,
weihs@statistik.tu-dortmund.de

Dominik Wied [Kapitel 14]
Universität zu Köln, Wirtschafts- und Sozialwissenschaftliche Fakultät,
dwied@uni-koeln.de

Notation

Abweichend von Standardrechtschreibregeln haben wir im gesamten Buch

- einen Dezimalpunkt verwendet und kein Dezimalkomma (also z. B. 1.23 statt 1,23) und
- die deutschen Anführungsstriche auch bei englischen Ausdrücken (also z. B. „english“ und nicht ‘english’).

Teil I
Leben und Sterben



Kapitel 1

Warum leben Novemberkinder länger?

Walter Krämer, Katharina Schüller

Wer im November geboren ist, lebt im Durchschnitt ein halbes Jahr länger als jemand mit Geburtstag im Mai. Die letztendliche Ursache muss das Klima sein, denn auf der Südhalbkugel der Erde ist das umgekehrt. Aber über welche konkreten Kanäle das Klima die weitere Lebenserwartung beeinflusst, war lange ungeklärt. Wir bieten eine Erklärung an.

1.1 Was sagt uns das Datum der Geburt?

Seit alters her versuchen Astrologen, aus dem Datum der Geburt auf das weitere Schicksal eines Menschen zurückzuschließen. Dabei hilft ihnen die Tatsache, dass es tatsächlich solche Einflüsse gibt. Die haben allerdings mit Planeten und Sternen nichts zu tun. So hat etwa der Psychologe Peter Jensen einmal die Geburtstage aller kanadischen Eishockeynationalspieler notiert. Kein einziger davon hatte seinen Geburtstag im Mai, Juni, Juli, August, September, Oktober, November oder Dezember. Und in Australien hat man herausgefunden, dass es bei fast allen Profisportarten erheblich mehr erfolgreiche Aktive mit Geburtstag im Januar als mit Geburtstag im Dezember gibt. Das kommt aber nicht von den Sternen, sondern daher, dass Januararkinder in ihren nach Altersjahrgängen gruppierten Trainingsgruppen die ältesten und oft allein schon deshalb auch die Besten sind. Damit erfahren sie aber auch die intensivste Förderung. Das Ganze vom Alter sechs bis achtzehn zwölfmal wiederholt führt dann fast schon automatisch dazu, dass sie auch beim Eintritt in die Profikarriere die besten sind.

Hier beeinflusst der Zeitpunkt der Geburt den späteren Erfolg im Beruf. Seit zwei Jahrzehnten häufen sich Indizien, dass der Geburtszeitpunkt auch einen Einfluss auf die durchschnittliche Länge des nachfolgenden Lebens hat. Abbildung 1.1 zeigt einmal für jeden Monat getrennt das durchschnittliche Sterbealter aller Schweizer Frauen und Männer, die zwischen 1969 und 2010 gestorben sind. Und wie man sieht, leben im Oktober / November geborene Schweizer und Schweizerinnen im Durchschnitt sechs Monate länger als im Mai geborene. (Wir sprechen hier von Novemberkindern wie in der ursprünglichen Literatur.)

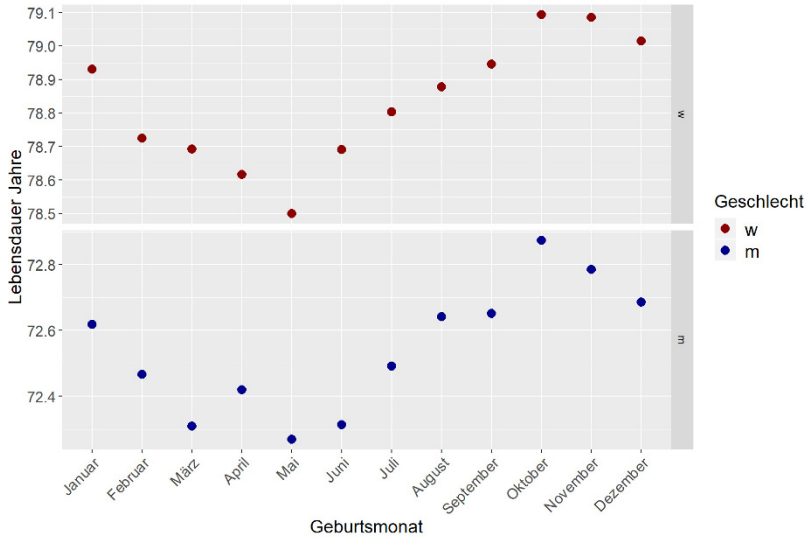


Abb. 1.1: Durchschnittliche Lebensdauer aller Schweizerinnen (oben) und Schweizer (unten), die zwischen 1969 und 2010 gestorben sind. Die Mittelwerte basieren auf tagesgenauen Lebensdauern, die in Jahre umgerechnet wurden.

Dieser Effekt ist kein Produkt des Zufalles. Ähnliche Ergebnisse zeigen andere Untersuchungen auch für Dänemark, Schweden oder Deutschland, wenn auch basierend auf weniger großen Datensätzen als dem, den wir für unsere Analysen zur Verfügung haben. Auf der Südhalbkugel der Erde ist das Ganze umgekehrt: Hier leben Novemberkinder rund sechs Monate kürzer als Kinder, die im Mai geboren sind. Damit ist klar: Das Klima zum Zeitpunkt der Geburt muss die letztendliche Ursache dieser Differenzen sein. So folgt etwa die Lebenserwartung von Australiern, die aus Europa eingewandert sind, dem Muster der Nordhalbkugel dieser Erde. Offen ist allein, durch welche Kanäle das Klima um unsere Geburt herum die Länge unseres Lebens mitbestimmt.

1.2 Daten und Fakten

Basis alles Weiteren ist ein einzigartiger Datensatz, den uns freundlicherweise das Schweizer Bundesamt für Statistik zur Verfügung gestellt hat: alle 2 531 335 Todesfälle von 1969 bis 2010 in der Schweiz mit detailliertem Todesdatum und -ort. Dazu kommt in den meisten Fällen auch das Geburtsdatum inklusive Geburtsort und sozioökonomische und soziale Angaben über Ehepartner, Eltern, Religion, Sprachregion, Beruf und wirtschaftliche Stellung, plus oft auch die Art des Todes: Hier haben wir alle möglichen Ursachen von Unfällen, über Selbstmord, Tötung durch andere, Vergiftung, bis zu verschiedenen Krankheiten, die weiter in Hauptkran-

kung, Folgeerkrankungen, Begleiterkrankungen und Nebenerkrankungen unterteilt sind, unter verschiedenen Gesichtspunkten analysiert.

Abbildung 1.2 etwa zeigt die monatspezifische Lebenserwartung von Schweizerinnen und Schweizern, die an Krebs gestorben sind, und Abbildung 1.3 das gleiche für Herz-Kreislaufkrankheiten.

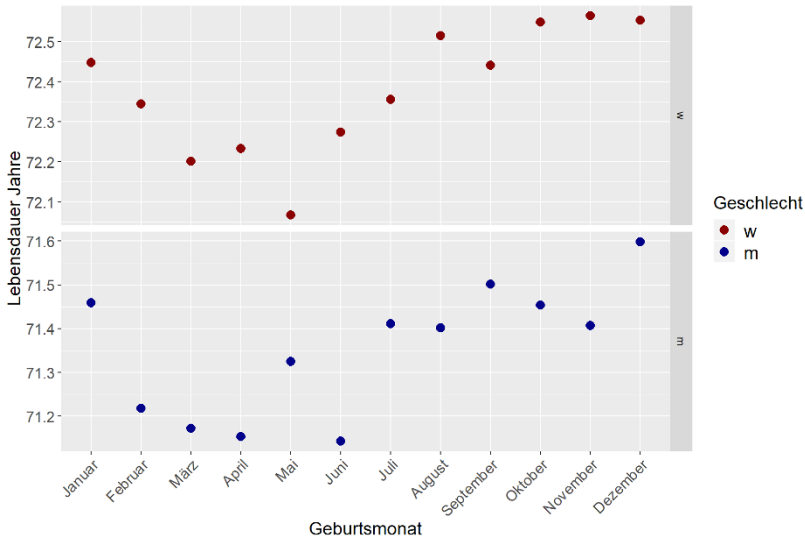


Abb. 1.2: Durchschnittliche Lebensdauer aller an Krebs gestorbenen Schweizerinnen (oben) und Schweizer (unten) in Abhängigkeit des Geburtsmonats.

Abbildung 1.4 dagegen zeigt die monatspezifische Lebenserwartung aller Schweizer, die durch einen Unfall gestorben sind. Zunächst ist sie erheblich kürzer, kein Wunder, und auch eine Abhängigkeit der Lebenserwartung vom Geburtsmonat ist hier nicht zu sehen.

1.3 Auf der Suche nach den Gründen

Lange Zeit hat man die Sonneneinstrahlung für einen wichtigen Faktor gehalten, der unsere Lebensspanne beeinflusst. So haben etwa amerikanische Forscher einmal die Lebensspanne von rund 7000 Abgeordneten des amerikanischen Kongresses ermittelt, die zwischen den Jahren 1750 und 1900 geboren worden waren. Dabei zeigte sich, dass das erreichte mittlere Alter keineswegs monoton im Zeitverlauf gestiegen ist, sondern in einem Zyklus von neun bis zwölf Jahren schwankt. So stieg etwa die Lebenserwartung von Abgeordneten, die im Jahr 1752 und danach geboren worden waren, bis zum Geburtsjahrgang 1763 stetig an und nahm dann ebenso stetig wieder ab. Diese Zyklen entsprechen fast exakt denen von Sonnenflecken und damit