

---

# Statistik

---

Udo Kuckartz • Stefan Rädiker  
Thomas Ebert • Julia Schehl

# Statistik

Eine verständliche Einführung

2., überarbeitete Auflage

 Springer VS

Prof. Dr. Udo Kuckartz  
Philipps-Universität Marburg  
Deutschland

Thomas Ebert  
Johann Wolfgang Goethe-Universität  
Frankfurt, Deutschland

Dr. Stefan Rädiker  
Philipps-Universität Marburg  
Deutschland

Julia Schehl  
VERBI Software, Consult,  
Sozialforschung GmbH  
Berlin, Deutschland

ISBN 978-3-531-19889-7  
DOI 10.1007/978-3-531-19890-3

ISBN 978-3-531-19890-3 (eBook)

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

Springer VS  
© Springer Fachmedien Wiesbaden 2010, 2013

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 Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

*Lektorat:* Dr. Cori Mackrodt, Stefanie Loyal

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist eine Marke von Springer DE. Springer DE ist Teil der Fachverlagsgruppe Springer Science+Business Media.  
[www.springer-vs.de](http://www.springer-vs.de)

# Inhalt

<b>Vorwort zur 2. Auflage.....</b>	<b>9</b>
<b>1 Die Aufbereitung der Daten für die statistische Analyse.....</b>	<b>13</b>
1.1 Der Codeplan .....	14
1.2 Fehlende Werte, immer ein Problem .....	20
1.3 Der Entwurf eines Codeplans – ein Beispiel .....	21
1.4 So geht es mit SPSS/MYSTAT.....	26
<b>2 Häufigkeitsverteilungen und ihre grafischen Darstellungen.....</b>	<b>35</b>
2.1 Häufigkeitstabellen.....	36
2.2 Bildung von Kategorien bei intervallskalierten Variablen.....	39
2.3 Grafische Darstellungen einfacher Häufigkeitsverteilungen .....	42
2.4 So geht es mit SPSS/MYSTAT.....	53
<b>3 Mittelwerte und Streuungsmaße.....</b>	<b>61</b>
3.1 Mittelwerte .....	61
3.2 Streuungsmaße.....	68
3.3 Boxplots: Grafische Darstellung von Streuungen .....	74
3.4 Standardisierungsverfahren .....	76
3.5 Mittelwerte und Streuungsmaße in der Forschungsliteratur.....	80
3.6 So geht es mit SPSS/MYSTAT.....	81
<b>4 Kreuztabelle, Chi-Quadrat und Zusammenhangsmaße .....</b>	<b>87</b>
4.1 Das Prinzip der Kreuztabelle .....	87
4.2 Absolute Häufigkeiten, Spaltensummen und Zeilensummen .....	89
4.3 Relative Häufigkeiten, Spaltenprozente und Zeilenprozente .....	90
4.4 Erwartungswerte und die Berechnung von Chi-Quadrat .....	92
4.5 Die Kreuztabelle mit mehrfach gestuften Merkmalen .....	95
4.6 Zusammenhangsmaße für die Kreuztellenanalyse .....	97
4.7 Weitere Variablen in die Analyse einbeziehen .....	100
4.8 Chi-Quadrat-Berechnung für univariate Verteilungen .....	100
4.9 Grafische Darstellung von Kreuztabellen.....	102
4.10 Die Kreuztellenanalyse in der Forschungsliteratur .....	104
4.11 So geht es mit SPSS/MYSTAT.....	107

<b>5</b>	<b>Wahrscheinlichkeit und Wahrscheinlichkeitsverteilungen.....</b>	<b>111</b>
5.1	Was ist Wahrscheinlichkeit und wie berechnet man sie? .....	112
5.2	Irren ist nicht nur menschlich, sondern auch wahrscheinlich .....	119
5.3	Wahrscheinlichkeitsverteilungen.....	121
5.4	Die Binomialverteilung.....	123
5.5	Die Normalverteilung.....	127
5.6	Die t-Verteilung.....	134
5.7	Die Chi-Quadrat-Verteilung.....	135
5.8	Die F-Verteilung.....	136
<b>6</b>	<b>Die Logik des statistischen Schließens.....</b>	<b>137</b>
6.1	Die Verteilung von Stichprobenkennwerten .....	137
6.2	Konfidenzintervalle .....	141
6.3	Die statistische Hypothese.....	144
6.4	Der Hypothesentest.....	147
6.5	Einseitige und zweiseitige Tests.....	149
6.6	Alpha-Fehler und Beta-Fehler .....	151
6.7	Signifikanz – ein Begriff, der in die Irre führen kann.....	153
6.8	Effektgröße.....	154
6.9	So geht es mit SPSS/MYSTAT.....	155
<b>7</b>	<b>t-Test: zwei Mittelwerte vergleichen.....</b>	<b>159</b>
7.1	Mittelwerte von zwei unabhängigen Stichproben vergleichen.....	161
7.2	Mittelwerte von zwei abhängigen Stichproben vergleichen .....	169
7.3	Der Vergleich von zwei Mittelwerten in der Forschungsliteratur.....	173
7.4	So geht es mit SPSS/MYSTAT.....	175
<b>8</b>	<b>Varianzanalyse: mehr als zwei Mittelwerte vergleichen.....</b>	<b>185</b>
8.1	Grundbegriffe der Varianzanalyse.....	185
8.2	Das Prinzip der einfaktoriellen Varianzanalyse.....	186
8.3	Durchführung einer einfaktoriellen Varianzanalyse .....	188
8.4	Die mehrfaktorielle Varianzanalyse .....	196
8.5	Voraussetzungen für die Durchführung einer ANOVA .....	198
8.6	Die Varianzanalyse in der Forschungsliteratur.....	199
8.7	So geht es mit SPSS/MYSTAT.....	200
<b>9</b>	<b>Korrelation: Zusammenhänge identifizieren.....</b>	<b>207</b>
9.1	Zusammenhänge von Variablen grafisch darstellen .....	207
9.2	Die Korrelation von intervallskalierten Variablen.....	210
9.3	Die Korrelation von ordinalskalierten Variablen .....	216
9.4	Die Korrelation von nominalskalierten Variablen .....	219
9.5	Übersicht über die Zusammenhangsmaße .....	222
9.6	Korrelation und Kausalität.....	223

---

9.7 Die Korrelationsanalyse in der Forschungsliteratur .....	226
9.8 So geht es mit SPSS/MYSTAT .....	228
<b>10 Skalenbildung .....</b>	<b>239</b>
10.1 Was ist überhaupt eine Skala? .....	240
10.2 Skalierungsverfahren .....	243
10.3 Konstruktion einer Likert-Skala .....	244
10.4 Die Qualität einer Skala .....	251
10.5 Angaben der Skalengüte in der Forschungsliteratur .....	252
10.6 So geht es mit SPSS/MYSTAT .....	253
<b>11 Regression: komplexe Zusammenhänge analysieren und Vorhersagen treffen .....</b>	<b>259</b>
11.1 Einfache lineare Regression .....	259
11.2 Multiple lineare Regression .....	266
11.3 Logistische Regression .....	273
11.4 Die Regressionsanalyse in der Forschungsliteratur .....	279
11.5 So geht es mit SPSS/MYSTAT .....	282
<b>Glossar .....</b>	<b>287</b>
<b>Literatur .....</b>	<b>295</b>
<b>Anhang .....</b>	<b>299</b>
A. Datensätze von sozial- und erziehungswissenschaftlichen Studien .....	299
B. Tabellen .....	305
<b>Register .....</b>	<b>311</b>

## Vorwort zur 2. Auflage

Dieses Lehrbuch der sozialwissenschaftlichen Statistik ist speziell für Bachelor- und Master-Studiengänge konzipiert: Es ist forschungsorientiert, verständlich und anwendungsorientiert. Was heißt das?

*Forschungsorientiert* bedeutet, dass zu Beginn der Planung des Inhalts dieses Buches die methodenkritische Lektüre der Top-Fachzeitschriften stand. Wir haben zunächst katalogisiert, welche statistischen Verfahren in den forschungsorientierten Artikeln verwendet werden, die bspw. in der Kölner Zeitschrift für Soziologie, der Zeitschrift für Erziehungswissenschaft, der Zeitschrift für Evaluation, der Zeitschrift für Soziologie oder der Zeitschrift für Pädagogik abgedruckt wurden. Verfahren, die dort häufig Verwendung finden, muss man als Fachwissenschaftler/in kennen, allein schon um die Fachbeiträge verstehen und kritisieren zu können. Solche Verfahren haben prioritär Eingang in dieses Buch gefunden, weil diese Zeitschriften den State-of-the-art von Forschung und Forschungsmethodik repräsentieren.

*Verständlich* bedeutet, dass wir darauf verzichtet haben, ein Lehrbuch der Stochastik mit sozialwissenschaftlichen Beispielen zu schreiben. Unsere Devise lautete: Soviel Mathematik und so viele Formeln wie nötig, soviel sozialwissenschaftliche Forschungs- und Interpretationslogik wie möglich. Wir wollen also in diesem Buch darstellen, was man aus der Perspektive der empirischen Sozialforschung an Statistikkenntnissen besitzen muss, und wollen nicht der inneren Logik der statistischen Verfahren und ihrer mathematischen Grundlagen folgen.

*Anwendungsorientiert* bedeutet, dass wir die Leser und Leserinnen in den Stand versetzen wollen, die dargestellten Verfahren auch praktisch im Forschungsalltag einsetzen zu können – und dies geschieht heutzutage mit Statistiksoftware, die preisgünstig und teilweise kostenlos für herkömmliche PCs verfügbar ist. Die Fähigkeit, all dies auch per Hand mit dem eigenen Taschenrechner ausführen zu können, ist dagegen von begrenztem Nutzen. Wir haben deshalb Wert darauf gelegt, jeweils zu erklären, wie die besprochenen Statistikverfahren in die Praxis umsetzbar sind, und zwar zum einen mit dem weit verbreiteten Programm SPSS und zum anderen mit dem Programm SYSTAT, dessen Version für Studierende unter dem Namen MYSTAT kostenlos zur Verfügung gestellt wird.

Statistik-Puristen werden vielleicht vorbringen, dass die Darstellung in diesem Buch zu oberflächlich sei und die Kompetenz preisen, auch eine Varianz-

analyse oder eine multiple Regression mit dem Taschenrechner lösen zu können. Wir halten diese Fähigkeit für entbehrlich. Viel wichtiger ist es unseres Erachtens, die Ergebnisse statistischer Analysen interpretieren zu können und die Realität empirischer Forschung mit kritischem Blick wahrzunehmen. Was nützt bspw. die theoretisch unbestrittene Überlegenheit von Zufallsstichproben als Auswahlverfahren, wenn mehr als die Hälfte der so ausgewählten Menschen – aus was für Motivlagen auch immer – nicht zu einem Interview bereit sind?

Die Kritik, die an den quantitativ-statistischen Methoden und den mit ihnen produzierten Ergebnissen in den letzten Jahrzehnten vorgetragen worden ist, hat nicht zuletzt deshalb gegriffen, weil das Festhalten am standardisierten Vorgehen und mathematisch-statistischen Prinzipien oftmals an den Realitäten der Lebenswelt vorbei ging. Es wäre allerdings falsch, aus der teilweise berechtigten Kritik gleich eine fundamentale Ablehnung zu folgern. Die Betrachtung von zahlenmäßigen Verteilungen, Relationen und Korrelationen kann sehr erhellend sein und stellt ein machtvolles Instrumentarium empirischer Sozialforschung dar. Das ist auch trotz der Ablehnung, auf die der Lerngegenstand Statistik des Öfteren stößt, den Studentinnen und Studenten durchaus bewusst. Im Rahmen unserer Vorlesung „Einführung in die sozialwissenschaftliche Statistik haben wir Studierende im 2. Semester schriftlich gefragt „Was ist der Nutzen von statistischen Erhebungen für die Pädagogik?“ Von den mehreren hundert Antworten auf diese offen gestellte Frage seien hier nur wenige exemplarisch genannt:<sup>1</sup>

- Erfolg oder Misserfolg von bestimmten Maßnahmen/Experimenten überprüfen
- Aktuelles, repräsentatives Bild von der Gesellschaft erhalten
- Bedürfnisse, Wünsche, Meinungen der Klienten und Klientinnen kennenlernen
- Theorien und Hypothesen überprüfen
- Ursachen und Problemfelder erkennen
- Verbesserung der Qualität pädagogischer Arbeit

Schon diese kleine Auswahl zeigt in ihrer verblüffenden Vielschichtigkeit, wie wichtig ein solides Basiswissen in statistischen Methoden sein kann.

Dieses Buch folgt der Logik der statistischen Datenanalyse, wie sie typischerweise in Forschungsprojekten praktiziert wird. Die Kapitel bauen aufeinander auf und sollten deshalb auch der Reihe nach gelesen werden. Der Prozess der Datenerhebung und deren verschiedene Formen wie Interview, Beobachtung, Experiment oder Inhaltsanalyse bleiben in diesem Buch außen vor. Wir starten mit dem ersten Kapitel an dem Punkt, wo die auszuwertenden Daten bereits vorliegen, z.B. in Form ausgefüllter Fragebögen oder Beobachtungs-

---

1 Eine ausführliche Darstellung der aktuellen Befragungsergebnisse zum Nutzen von Statistik findet sich in unserem Blog: <http://methoden-evaluation.de/blog/?p=1095>



bögen. In den folgenden zehn Kapiteln wird ein anwendungsorientiertes statistisches Grundwissen vermittelt.

Die meisten Kapitel enthalten einen Abschnitt, in dem Beispiele aus der aktuellen Forschungsliteratur dargestellt werden, um einen Bezug zur Fachliteratur herzustellen. Alle Kapitel, in denen Berechnungen erläutert werden, beinhalten zudem einen Abschnitt „So geht es mit SPSS/MYSTAT“, in dem die Umsetzung des Beschriebenen mit Statistiksoftware erklärt wird. Neu aufgenommen haben wir in die zweite Auflage Lernfragen am Ende eines jeden Kapitels, die der Wiederholung des Stoffes und der selbstgesteuerten Lernkontrolle dienen sollen. Der Anhang enthält unter anderem ein Glossar der zentralen Begriffe sowie Hinweise auf wichtige, für die Forschung frei zugängliche Datensätze großer sozial- und erziehungswissenschaftlicher Studien.

Das Buch ist als Begleitbuch für einen einsemestrigen Hochschulkurs mit 15 Terminen konzipiert, wobei der erste Termin wie üblich als Einführungstermin und der letzte Termin für die abschließend Klausur vorgesehen ist. Je nach intendierter Schwerpunktsetzung können dann zwei Kapitel, bspw. Kapitel 4 „Kreuztabellenanalyse“ und Kapitel 11 „Regression“, vertiefend an zwei Terminen bearbeitet werden. Es ist empfehlenswert und höchst sinnvoll, parallel zum inhaltlichen Seminar bzw. zur Vorlesung einen Übungstermin oder ein Tutorium vorzusehen.

Wir danken allen Lesern und Leserinnen der ersten Auflage, die uns Verbesserungswünsche übermittelt haben. Wir haben diese soweit wie möglich aufgenommen und bei der Überarbeitung berücksichtigt. Bei allen, die uns bei der Fertigstellung der zweiten Auflage unterstützt haben, möchten wir uns herzlich bedanken. Dies gilt vor allem für die Mitarbeiter und Mitarbeiterinnen der Marburger Arbeitsgruppe für Methoden & Evaluation (magma), insbesondere für Patrick Plettenberg, Mailin Gunkel, Hannah Prömper und Tamara Reisdorf.

*Udo Kuckartz, Stefan Rädiker, Thomas Ebert, Julia Schehl*

Marburg, im Juli 2013

P.S.: Im Internet stellen wir unter [www.statistik-verständlich.de](http://www.statistik-verständlich.de) zusätzliche Informationen (interessante Links, Formelsammlung etc.) bereit.