

Book Reviews

Karl-Ernst BIEBLER und Bernd JÄGER (2008). **Biometrische und epidemiologische Methoden**. München, Oldenbourg Verlag, 384 S., ISBN 978-3-486-58511-7. (broschiert € 34,80)

Dieses Buch ist konzipiert als Lehr- und Handbuch für Wissenschaftler und Studierende verschiedener Fachrichtungen, die sich mit der Auswertung von Daten beschäftigen. Es ist methodenorientiert geschrieben. Die Autoren betrachten neben den Grundlagen der Wahrscheinlichkeitsrechnung und der Statistik weitere Gebiete der Mathematik und ihre Anwendungsmöglichkeiten in der biometrisch-epidemiologischen Methodik. Vorausgesetzt werden hierbei nur Schulkenntnisse in Mathematik. Um ein methodisches Verständnis für statistische Verfahren aufzubauen, wird auf die mathematischen Teilgebiete, wie der n -dimensionale reelle Zahlenraum, Koordinaten und Koordinatentransformationen und den Abstandsbegriff eingegangen. Es wird immer auf die mathematischen Grundlagen der jeweiligen Verfahren hingewiesen.

In der Einleitung (Kapitel 1) wird die historische Entwicklung der biometrischen und epidemiologischen Methoden kurz erläutert.

Im 2. Kapitel werden wahrscheinlichkeitstheoretische, statistische und algebraische Grundlagen erörtert. Es gliedert sich im Wesentlichen in vier Teile: die beschreibende Statistik, Grundelemente der Wahrscheinlichkeitsrechnung, Verfahren der schließenden Statistik und Methoden zur Behandlung höherdimensionaler Datenmengen. Im Abschnitt zur beschreibenden Statistik werden zunächst grundlegende Begriffe, wie z.B. Population, Merkmale, Grundgesamtheit und Skalenniveaus vorgestellt. Unterschiedliche graphische Darstellungsmöglichkeiten und Kennzahlen werden beschrieben, empirische Funktionen werden definiert. Im Anschluss werden Grundlagen der Wahrscheinlichkeitsrechnung und Wahrscheinlichkeitsverteilungen erläutert. Abhängigkeit – Unabhängigkeit, bedingte Wahrscheinlichkeiten, Begriff der Verteilungsfunktion, zweidimensionale Zufallsgrößen, Regression, Grenzwertsätze werden beschrieben. Der dritte Teil dieses Kapitels behandelt Verfahren der schließenden Statistik. Zunächst werden die Begriffe Stichprobe, Stichprobenraum und Stichprobenfunktion erklärt. Es werden einige statistische Tests dargestellt, die konkrete Anwendung der verschiedenen statistischen Testverfahren wird jedoch nicht sehr anschaulich erläutert. Der Hinweis auf Verfahren zum Vergleich von mehr als zwei Gruppen, wie beispielsweise die einfaktorielle Varianzanalyse, fehlt. Jedoch wird das sehr komplexe Gebiet der Analyse von Daten der Lebensdauer angeschnitten. Auf weiterführende statistische Verfahren, wie Generalisierte Lineare Modelle, Faktoren-, Diskriminanz- und Clusteranalyse wird hingewiesen. Es folgt noch ein Punkt zu den Methoden zur Behandlung von höherdimensionalen Datenmengen. In diesem Kontext wird auf Matrizenrechnung und lineare Algebra eingegangen. So wird der n -dimensionale euklidische Raum erläutert, und quadratische Formen und Abstandsbegriffe werden beschrieben. Die konkrete Anwendung dieser Verfahren fällt jedoch sehr kurz aus.

In Kapitel 3 wird eine Einführung in die Epidemiologie, welche in diesem Buch ebenfalls methodenorientiert dargestellt wird, an sehr anschaulichen Beispielen gegeben.

Das 4. Kapitel beschäftigt sich mit spezieller Epidemiologie. Die Möglichkeiten der mathematischen Modellierung populationsbezogener Prozesse werden vorgestellt. Ein deterministischer und ein stochastischer Ansatz zur Modellierung des zeitlichen Verlaufs von Infektionen werden dargestellt. Verschiedene Schätzmethoden werden untersucht. Das Buch endet mit einer Einführung in die genetische Epidemiologie und einer Beschreibung eines stochastischen Prozesses als epidemiologisches Modell.

Kurz zusammengefasst ist dieses Buch empfehlenswert für LeserInnen, die sich gerne mit den mathematischen Hintergründen verschiedener statistische Verfahren auseinandersetzen möchten.

*Petra Ofner-Kopeinig
Institut für Medizinische Informatik, Statistik und Dokumentation
Medizinische Universität Graz*

Christian F. G. SCHENDERA (2007). **Datenqualität mit SPSS**. München, Oldenbourg Verlag, XVII+472 S., ISBN 978-3-486-58214-7. (broschiert € 34,80)

Als ich dieses Buch erstmals in meinen Händen hielt und darin blätterte, fiel mir sogleich auf, dass es sich hierbei keineswegs nur um ein weiteres, einführendes Handbuch über diverse statistische Möglichkeiten einer Datenanalyse mittels SPSS handelt. Bereits beim Durchblättern war zu erkennen, dass darin vor allem Vorschläge für Problemlösungen unter Verwendung der mir aus meiner Studienzeit noch besser bekannten SPSS Syntax angeboten werden. Für eine ausführliche Diskussion der alternativen Handhabung von SPSS mit Maus und auszufüllenden Dialogboxen sei hier beispielsweise auf das Buch von Bühl und Zöfel (2005) verwiesen, das zusätzlich auch eine umfangreiche Information über die in SPSS verfügbaren statistischen Methoden enthält. Im vorliegenden Buch spielt jedoch die statistische Analyse von Daten keine zentrale Rolle, sondern es geht hierbei ausschließlich um Möglichkeiten der Datenmanipulation zur Sicherung der Datenqualität.

Im ersten Kapitel werden dazu die zentralen Qualitätskriterien vorgestellt, die meist aus der Umfrageforschung aber auch aus dem Bereich der Data Warehouses stammen. Dazu zählen die Vollständigkeit und Einheitlichkeit von Datensätzen, das Auffinden und Behandeln von fehlenden Daten (Missings), "Doppelten" oder Ausreißern, sowie die Prüfung der Plausibilität von Variablen oder Werten. Weitere Kriterien bauen auf diese Konzepte auf. All diese Kriterien werden sehr detailliert in den folgenden sechs Kapiteln diskutiert.

Im zweiten Kapitel über Zielsetzung, Konzept und Grundlagen werden unter anderem die klaren Vorteile der Syntax Steuerung gegenüber einer längeren Serie von Mausclicks hergeleitet. Von zentraler Bedeutung ist hierbei die Wiederverwendbarkeit des Codes, wodurch man Effizienz, Flexibilität und Permanenz gewinnt, oder auch die Protokollierung und Dokumentation, um nur einige Aspekte zu nennen. Interessanterweise ist die SPSS Syntax weitgehend plattformunabhängig und aufwärtskompatibel, was jedoch für andere SPSS Features wie Grafiken oder Datendateien nicht immer uneingeschränkt zutrifft. Nicht diskutiert werden ausschlaggebende Gründe für die Verwendung von SPSS selbst, oder was am Einsatz von SPSS attraktiver als beispielsweise von R sein könnte, für das ja auch all diese Punkte zutreffen. Das vorliegende Buch richtet sich daher eher an die Anwender von SPSS und stellt dieser Lesergruppe in SPSS Syntax realisierte Werkzeuge zur Verfügung.

Daten nennt man vollständig, falls deren Anzahl der Summe aller gültigen und fehlenden Angaben entspricht. Das dritte Kapitel beschäftigt sich mit der Prüfbarkeit und den Kontrollmöglichkeiten dieses Kriteriums und bietet auch konkrete Vorschläge zum Füllen von Datenlücken an. Die Einheitlichkeit von Variableninhalten behandelt das vierte Kapitel. Hier wird darauf hingewiesen, wie wichtig es ist, in einem Datensatz auf keine unterschiedlich geschriebenen Personen- oder Ortsnamen, Datumsangaben oder sogar auf falsch geschriebene Variablennamen zu stoßen. Auch die Notwendigkeit einer gleichen Variablenkodierung zählt zu diesem Kriterium. Wie manche Uneinheitlichkeit identifiziert und mittels SPSS wieder vereinheitlicht werden kann, wird anhand von Beispielen demonstriert.

Mit dem Erkennen, Interpretieren und Filtern doppelter Werte oder mehrfacher Datenzeilen beschäftigt sich das fünfte Kapitel. Sehr informativ ist Kapitel 6, das sich mit

fehlenden Daten (Missings) auseinandersetzt. Außer den diversen Ursachen werden auch potentielle statistische Folgen wie Bias oder reduzierte Macht in diesem Kapitel besprochen. Besonders gelungen erscheint mir der anschließende Teil über die Mechanismen von Missings (Missing Completely at Random, Missing at Random, Not Missing at Random) und über die Möglichkeiten zur Rekonstruktion bzw. zum Ersetzen von Missings.

Wie man uni- oder multivariate Ausreißer identifizieren kann (meist mittels Diagrammen) und wie man damit umgeht, wird im Kapitel 7 besprochen. Diese Kriterien und die im folgenden Kapitel 8 verwendeten Plausibilitätsprüfungen werden vom Autor zu den fortgeschrittenen Ansätzen gezählt. Konkrete Beispiele zeigen, wie man die Plausibilität praktisch überprüft und wie dazu die seit SPSS 14 angebotene Funktion DETECTANOMALY eingesetzt werden kann. Im Kapitel 9 wird gezeigt, wie man mausgesteuert mit dem Menüpunkt "Daten validieren" umgeht. Danach wird die etwas gewöhnungsbedürftige Syntax von VALIDATEDATA vorgestellt.

Die beiden anschließenden Kapiteln behandeln das Arbeiten innerhalb eines Datensatzes (Kapitel 10) oder mit mehreren Datensätzen (Kapitel 11). Wie zeit- und datumsbezogene Probleme erkannt und gelöst werden (selbstverständlich auch unter Berücksichtigung des bestens bekannten Y2K Problems) wird im Kapitel 12 diskutiert. Weitere Kriterien, die eher vom Anwender eingesetzt werden sollten und weniger mit der Verwendung von SPSS zu tun haben, erwähnt der Autor im Kapitel 13.

Schließlich folgt noch eine kleine Übungsaufgabe und ein reales Programmbeispiel aus dem Bereich der Marktsegmentierung. Hinweise zum Arbeiten mit Clementine oder für Macintosh User werden noch anschließend angeboten. Sehr interessant fand ich auch den Inhalt des letzten Kapitels, in dem man die wesentlichen Kriterien für die Kommunikation von Qualität hergeleitet findet. Dazu zählen vor allem auch Kriterien für die Qualität der Datenanalyse, die von der verwendeten statistischen Methodik abgeleitet werden müssen.

Noch zu erwähnen bleibt, dass in dem vorliegenden Buch keineswegs in die Syntax von SPSS selbst eingeführt wird. Diese sollte dem Leser bereits gut bekannt sein, um die Befehle zu verstehen und den Abläufen in den Programmen folgen zu können. Leider fehlt auch eine beigelegte CD mit dem gesamten Syntax-Code oder zumindest der Hinweis auf eine Webadresse, wo die im Buch verwendeten Syntax-Listen erhältlich sind. Dieses Buch ist sicherlich weder ein Lehrbuch noch für Studenten verfasst. Es richtet sich an die Gruppe der SPSS Anwender, die größere Datenmengen zu bearbeiten haben und denen es ein verständliches Anliegen ist, dabei natürlich auch auf die Qualität zu achten.

Literatur:

Bühl, A., und Zöfel, P. (2005). **SPSS 12 – Einführung in die moderne Datenanalyse unter Windows (mit CD)**. München, Pearson Studium.

*Herwig Friedl
Institut für Statistik
Technische Universität Graz*