

Book Reviews

Rainer SCHLITTEGEN (2009). **Multivariate Statistik**. München, Oldenbourg Verlag, XII+559 S., ISBN 978-3-486-58597-7. (€ 49,80).

Multivariate Verfahren spielen in vielen Anwendungsgebieten eine große Rolle. Vor allem wenn es darum geht Abhängigkeiten in mehrdimensionalen Daten zu erkennen, Zusammenhänge zu beschreiben, sinnvolle Gruppierungen durchzuführen, und Strukturgleichungsmodelle mit beobachtbaren und nicht beobachtbaren (latenten) Variablen aufzustellen und zu analysieren. Der vorliegende Band ist zwar ein anwendungsorientiertes Handbuch, verzichtet aber nicht auf die Darstellung von theoretischen Ergebnissen, unterstützt durch geometrische Darstellungen. Breiten Raum nimmt die Diskussion der Voraussetzungen und Anwendbarkeit von Methoden ein.

Das Buch ist in sechs große Teile untergliedert. Teil I diskutiert die Beschreibung mehrdimensionaler Daten, welche kontinuierliche, nominale oder ordinale Merkmale beinhalten können. In Teil II werden die Multinomialverteilung für diskrete Variable, sowie die multivariate Normalverteilung und elliptisch symmetrische Verteilungen für stetige Variable eingeführt. Im Falle stetiger Variabler werden Maximum-Likelihood-Schätzer für Lokation und Kovarianz, robuste Schätzer, Ein- und Zweistichprobentests und die multivariate Varianzanalyse erörtert. Für nominal skalierte Merkmale werden die üblichen Methoden für Kontingenztafeln angeführt. Das Problem der fehlenden Werte wird ebenfalls aufgegriffen und zwei Lösungsvorschläge für die Parameterschätzung vorgestellt: der EM-Algorithmus und das Prinzip der multiplen Imputation.

Teil III ist der Analyse von Abhängigkeiten gewidmet. Für metrische Zielgrößen werden die klassischen Verfahren der linearen Regression, die robuste Regression und die multivariate multiple Regression erörtert. Die noch nicht so etablierten Methoden der Hauptkomponenten- und partial least squares (PLS)-Regression, die speziell bei Vorliegen sehr vieler Regressoren anwendbar sind, werden ebenfalls untersucht. Die logistische Regression (binomialverteilte Zielgrößen) und multinomiale Logitmodelle (multivariat verteilte Zielgrößen) werden eingeführt und finden Anwendung in der so genannten Conjoint-Analyse, bei der Kundenpräferenzen auf Produkteigenschaften abgebildet werden.

Die Hauptkomponentenanalyse in Teil IV (Zusammenhänge) hat das Ziel die Daten durch eine möglichst geringe Anzahl von Linearkombinationen der beobachteten Variablen zu beschreiben. Dazu passt auch der Biplot, der die gemeinsame Darstellung von Objekten und zweidimensionalen Variablen (als Ergebnis der Hauptkomponentenanalyse) zum Ziel hat. Distanzen werden durch die multidimensionale Skalierung auf eine niedrigere Dimension projiziert und graphisch dargestellt. Zusammenhänge zwischen Gruppen von Variablen können über die kanonische Korrelation ausgedrückt werden. Weiters werden log-lineare Modelle für zwei-, drei- und höher-dimensionale Kontingenztafeln analysiert.

Der Bereich der Gruppierungen (Teil V) lässt sich in zwei methodische Bereiche aufspalten: der Diskriminanzanalyse und der Clusteranalyse. Bei der Diskriminanzanalyse werden Gruppen auf Basis relevanter Variabler durch Trennfunktionen getrennt. Eines der Ziele ist es neue Beobachtungen einer Gruppe zuzuordnen, also zu klassifizieren.

Übliche Ansätze sind die lineare und die logistische Diskrimination, sowie Klassifikationsbäume. Bei der Clusteranalyse werden Beobachtungen in wenige homogene Gruppen oder Cluster zusammengefasst. Diese Verfahren werden häufig explorativ und eher selten modellbasiert eingesetzt.

Das komplexe Gebiet der Strukturgleichungsmodelle ist Inhalt von Teil VI. Die allgemeine Grundlage bildet eine Pfadanalyse, welche gerichtete Zusammenhänge zwischen Variablen, die beobachtbar oder latent sein können, herstellt. In der Faktorenanalyse werden beobachtbare Variable auf einige wenige Faktoren (latente Variable) abgebildet. Eine Erweiterung davon ist der LISREL-Ansatz, der die Faktorenanalyse mit der Regressionsanalyse verbindet und auf der Kovarianzstruktur der Modellgleichungen basiert. Im Gegensatz dazu werden bei der PLS-Pfadmodellierung die Scores der latenten Variablen bestimmt und die Modellparameter geschätzt. In Anhang VII werden schließlich die wichtigsten Eigenschaften aus der linearen Algebra zusammengestellt.

Dieses Lehr- und Handbuch zeichnet sich durch eine ausgewogene Darstellung von theoretischem Hintergrund, praktischen Beispielen und kritischer Analyse der benutzten Methodik aus. Es enthält neben den klassischen Verfahren der multivariaten Statistik auch neuere Zugänge, die in anderen Lehrbüchern kaum zu finden sind (robuste Verfahren, fehlende Werte, kategoriale Daten). Besonders hervorzuheben sind die ausführlichen und kritischen Kommentare zur Methodik, die hilfreichen Literaturhinweise und ein Literaturverzeichnis am Ende jedes der insgesamt 19 Kapitel. Die Softwareaspekte beschränken sich im Buch auf Ausgabelisten (SAS und R) und Kommentare bzgl. einzelner Programmroutinen. Datensätze und Beispielprogramme können aber unter www.oldenbourg-wissenschaftsverlag.de (Homepage des Buches) heruntergeladen werden. Etwas störend waren die auf manchen Seiten gehäuft auftretenden Fehler im Text, die durch eine sorgfältigere Durchsicht leicht zu vermeiden gewesen wären.

*Ernst Stadlober
Institut für Statistik
Technische Universität Graz*