

thoden nicht tief genug erläutert werden. Doch das sei hier nicht als Kritik sondern vielmehr als Gebrauchsanweisung im Sinne des Autors selbst gelesen (S. 11).

JÜRGEN SCHIENER, MAINZ

* * * * *



CHRISTIAN FG SCHENDERA, 2010: Clusteranalyse mit SPSS. Mit Faktorenanalyse. Oldenburg Wissenschaftsverlag. München. ISBN 978-3-486-58691-6, 435 Seiten, EUR 49,80.

Die Monographie „Clusteranalyse mit SPSS“ von Schendera reiht sich ein in die Sammlung von Begleitbüchern zur Anwendung von SPSS, für die offenbar mangels SPSS-eigener Publikationen Nachfrage besteht.

Das im Haupttitel und Kapitel 1 behandelte Thema, die Clusteranalyse, ist nicht der einzige Gegenstand des Werks. Im Untertitel „Mit Faktorenanalyse“ (Kapitel 2) wird dem Leser noch ein weiteres Verfahren offeriert und als unangekündigter Überraschungsgast taucht im dritten Kapitel noch die „Diskriminanzanalyse“ auf. Der Autor nimmt in seinem Vorwort die zu erwartende Frage des Lesers vorweg: „Warum die augenscheinlich grundverschiedenen „Verfahren“ Cluster- und Faktorenanalyse in einem Buch?“. Seine „einfache Antwort“ ist: „Beiden Verfahrensgruppen liegt dasselbe Prinzip zugrunde: Klassifikation mit dem Ziel maximaler Homogenität.“ Da auf der Plattform dieses Prinzips noch eine ganze Reihe weiterer statistischer Methoden

untergebracht werden könnte, ergänzt der Autor den prinzipiellen Grund um eine Liste weiterer – m. E. nicht sehr triftiger – Spezialgründe, die im Einzelnen zu behandeln hier nicht der Ort ist.

Weniger ausführlich wird die Einbeziehung der Diskriminanzanalyse begründet – sie wird mehr oder weniger als ein zur Clusteranalyse komplementäres Verfahren vorgestellt: „Die Diskriminanzanalyse kehrt dabei die Vorgehensweise der Clusteranalyse um und versucht aus den clusteranalytisierten Variablen auf die ermittelte Clusterzugehörigkeit zu schließen. Aus dieser Perspektive heraus kann man die Diskriminanzanalyse *auch* als einen statistischen Plausibilitätstest der erzielten Clusterlösung verstehen.“ (S. 299)

Im Schlusskapitel 4 sind dann den Verfahren zugrundeliegende Formeln aufgelistet.

Wer das Buch in die Hand nimmt, um etwas über die thematisch benannten Verfahren zu lernen, der wird enttäuscht sein – es handelt sich im ersten Teil eher um ein „Kochbuch“. Es hat demzufolge auch das Problem, die Darstellung des Verfahrens und seiner Umsetzung in eine SPSS-Syntax mit anschließender Erläuterung der Ausgabe in einen einheitlichen Textfluss zu bringen. Dies gelingt, kurz gesagt, für den Bereich der Clusteranalyse nahezu nie. Bereits das Vorwort ist in Vorwegnahme später dargestellter Inhalte befrachtet mit Fachterminologie und Meinungen zur richtigen Anwendungsweise und Interpretation, die für den „Einsteiger“ kaum verständlich sind. Das Buch erreicht in der Darstellung der Clusteranalyse stellenweise bestenfalls das Niveau einschlägiger Lehrbücher, die aber den unvergleichlichen Vorteil haben, ausschließlich der Sachlogik folgen zu dürfen und einen Leser auch befähigen, SPSS anzuwenden, eventuell mit kleinen Unterstützungen aus dem Hilfe-Menu.

Zunächst beginnt der Teil der Clusteranalyse noch gut mit einem heuristischen Beispiel aus dem Ferienalltag – Gruppierung von

Muscheln am Strand. Leider ist dieses Beispiel nicht konsequent durchgeführt. Der Leser erfährt nicht, für welche Clusterlösung er sich – etwa aufgrund des Dendrogramms – entscheiden könnte. Außerdem sind bei der Durchführung des Beispiels in der Berechnung der Distanzmatrix nach dem zweiten Fusionsschritt hässliche Schönheitsfehler passiert. Das in diesem Schritt neu gebildete Aggregat aus Muschel 2 und 5 erhält nicht die neu zu berechnenden Aggregatdistanzen, es werden vielmehr alte Maßzahlen aus der Matrix nach dem ersten Fusionsschritt übernommen – verwirrend für Einsteiger.

Aber dann schlägt der Text eine Tonlage an, die spekulativ ist und zu dem Verfahren keinen stringenten Bezug besitzt.

Richtig ist der Hinweis (S. 6 unten, S. 7 oben) darauf, dass die Ergebnisse einer Analyse von der Wahl des Distanzmaßes und des „Fusionsalgorithmus“, also der Definition der Distanz von Aggregaten und der Skalierung der Variablen abhängen. So wenig aber wie am Beispiel überhaupt ein „Ergebnis“ festgehalten wird, so wenig werden diese Abhängigkeiten an dem gewählten Beispiel demonstriert (es ist allerdings auch nicht dafür geeignet). Dass das Ergebnis auch noch von „Merkmalen der zu gruppierenden Elemente (u. a. beobachtbare vs. latente Merkmale)“ abhängen soll, ist kaum zu verstehen. Der nachfolgende Text zeigt jedoch, dass es sich dabei um ein Problem der Interpretation handeln soll.

Angeblich liegt der Clusteranalyse „eine unausgesprochene, aber absolut zentrale, und zwar gleich zweifache Annahme zugrunde: a) Dass die noch nicht geclusterten Daten überhaupt Cluster enthalten, und b) dass diese Cluster Sinn machen. Zusammen münden diese beiden Aspekte wieder in die zu prüfende Voraussetzung, ob die Daten überhaupt Strukturen i. S. v. sinnhaltigen Gruppierungen enthalten (können).“ (S. 7) Das ist schlicht irreführend: mit der Anwendung eines explorativen Verfahrens setzt

man nicht die Struktur schon voraus, die man zu finden hofft.

Die anschließenden Ausführungen über manifeste und latente Variable und ihre jeweilige Bedeutung für die Interpretation von Clustern sollen den Leser zur Vorsicht gemahnen, helfen ihm aber im Verständnis der Clusteranalyse nicht weiter.

Das Problem einer Rezension des vorliegenden Buchs ist, dass sie zu viele einzelne Kritikpunkte erfordern würde, wollte man alle Details des Texts angemessen würdigen. Ich greife daher nur einige besonders typische Passagen des Buches in der Rezension auf.

Der Kochbuchcharakter wird sehr deutlich im weiteren Verlauf der Darstellung der hierarchisch-agglomerativen Clusteranalyse in Abschnitt 1.2. Die vielfältigen Möglichkeiten von SPSS für die verschiedenen Typen von Variablen werden hier tabellarisch quasi plakatiert. Der Begleittext kommentiert und gibt Empfehlungen zur Anwendung, teilweise allerdings nur mit minimalen Informationen. Immerhin ist es vorstellbar, dass sich auch ein „Einsteiger“ nach einer solchen Lektüre vor den PC setzen kann und eine Clusteranalyse rechnet. Aber der Text ist eher für ein „Durchwursteln“ geeignet als für den systematischen Erwerb von Kenntnissen.

Die Anwendungen zeigen die Abhängigkeit der Clusterlösungen von den gewählten Verfahren, dann folgt die Eingabemöglichkeit von verschiedenen Datentypen (Fall vs. Variable – Matrix oder Distanzmatrix) und konkrete Beispiele mit Variablen unterschiedlichen Skalenniveaus. Leider werden bei den Anwendungen die didaktischen Möglichkeiten verschenkt, die beispielhafte Analysen bieten würden. Hier müsste ausführlicher darauf eingegangen werden, wie der Anwender die Entscheidungen über die Verfahrensparameter treffen sollte. Der Leser erfährt hier nur etwas über die korrekte SPSS-Syntax, aber nichts über eine auch inhaltlich befriedigende Datenanalyse.

Besonders problematisch ist außerdem der Vorschlag, der in Abschnitt 1.2.9 „Analyse von gemischten Daten II“ entwickelt wird. Bei einem Satz von Clustervariablen, der aus einem Teil binärer und einem Teil metrischer Variablen besteht, soll für jeden Teil eine eigene Clusteranalyse gerechnet werden. Nach Auswahl je einer Clusterlösung sollen dann die Beobachtungen entsprechend den beiden Clusterzugehörigkeitsvariablen kreuztabelliert werden. Jede Zelle der Kreuztabelle sei dann ein eigenes Cluster. Dieses Vorgehen mag in gewissen Spezialfällen zu sinnvoll interpretierbaren Clustern führen, ist aber generell nicht begründbar. Es wird dann scheitern, wenn die Cluster sich durch bestimmte Interaktionen zwischen den metrischen und den binären Variablen auszeichnen.

Das Two-Step-Clusterverfahren ist unter den hierarchisch-agglomerativen Verfahren die Spezialität von SPSS. Bis auf ca. drei Seiten, auf denen das Verfahren heuristisch beschrieben wird, und die sich stark an Bacher et al. und den von SPSS zitierten Originalarbeiten orientieren und auf diese verweisen, ist dieser Teil im Wesentlichen der Gang durch eine Beispielanalyse mit Erläuterungen.

Auch der Abschnitt 1.4 „Partitionierendes Verfahren: Clusterzentrenanalyse (k-means)“ gibt im Wesentlichen eine korrekte Beschreibung. Allerdings fehlt eine genauere Beschreibung des Sortieralgorithmus, der auf der Minimal-Distanzeigenschaft einer optimalen Partition beruht. Da es mehrere solcher Partitionen geben kann, besitzt das k-means Verfahren mitunter keine eindeutige, sondern mehrere startpunktabhängige Lösungen.

Die beiden „Nebenthemen“ des vorliegenden Buchs, Faktorenanalyse und Diskriminanzanalyse, erscheinen mir dagegen insgesamt gut gelungen. Das liegt sicherlich daran, dass beiden Methoden eine geschlossene mathematische Theorie zugrunde liegt, zum anderen aber auch daran, dass in der Dar-

stellung der Theorie durch Schendera das „Kochbuchprinzip“ aufgegeben worden ist. Insbesondere wird nun die Theorie der jeweiligen Methode in eigenständigen Texten beschrieben. Sie sind zwar alles andere als für „Einsteiger“ geeignet, da sie in relativ dichter Form einen Überblick über das jeweilige Gebiet geben und die Theorie nicht im Stil eines Lehrbuchs entwickeln. Für den Leser, der sich hier schon einmal umgesehen hat, geben sie jedoch einen guten Überblick.

Das gilt besonders für die Darstellung der Faktorenanalyse, bei der die Darstellung der Grundformen und Varianten sowie der Grundzüge der Anwendung und der Schätzungen sehr systematisch entwickelt sind. In fünf sorgfältig durchgeführten Beispielanalysen werden Varianten der Faktorenanalyse vorgeführt und die Ergebnisse interpretiert. Abschließend werden noch die „Faktorenanalyse von Fällen“, ein an Clusteranalyse erinnernder Ansatz, und die „Faktorenanalyse eingeleiteter Fälle“ dargestellt. Letzteres Verfahren ist etwa bei ordinalen Variablen anzuwenden, bei denen anstelle der Korrelationskoeffizienten Spearman- oder Kendallkoeffizienten analysiert werden.

Analoges ist von der Darstellung der Diskriminanzanalyse zu sagen. Einer Darstellung der Theorie, die ebenfalls nicht so einfach für „Einsteiger“ geeignet sein dürfte, folgen zwei ebenfalls sorgfältig präsentierte Beispielanalysen.

Man möchte dem Buch wünschen, dass in künftigen Auflagen das Thema Clusteranalyse eine sorgfältige Überarbeitung erfährt, die es auf das Niveau der beiden letzten Teile bringt.

Literatur

Bacher, J., K. Wenzig und M. Vogler, 2004: SPSS Two-Step Cluster – A First Evaluation. Friedrich-Alexander-Universität Erlangen-Nürnberg. Sozialwissenschaftliches Institut. Lehrstuhl für Soziologie. Arbeits- und Diskussionspapiere 2004-2 (2. Auflage).