Clusteranalyse

8

Inhaltsverzeichnis

K. Backhaus et al., *Multivariate Analysemethoden*,
https://doi.org/10.1007/978-3-658-47929-9_8

8.1 Problemstellung

Empirische Untersuchungen erheben oft die Einschätzungen einer Vielzahl von Personen gegenüber bestimmten Sachverhalten. Häufig ergeben sich dabei große Unterschiede in den Einschätzungen der einzelnen Personen, sodass die Daten in einer Erhebung stark voneinander abweichen. Es besteht eine hohe *Heterogenität.* Werden stark heterogene Daten, z. B. durch einen Mittelwert beschrieben, so ist dieser mit einer entsprechend großen Varianz bzw. Standardabweichung verbunden. Diese deutet statistisch bereits darauf hin, dass der Mittelwert als charakteristisches Maß für die Gesamterhebung eine nur geringe Aussagekraft besitzt. Je geringer hingegen die Heterogenität, desto verlässlicher ist die Aussage des Mittelwertes und desto kleiner die zugehörige Standardabweichung.

Eine Möglichkeit, das Problem der Heterogenität von Daten zu lösen, liegt in der Zusammenfassung von Personen (oder Objekten) zu Gruppen, die in ihren Einschätzungen relativ vergleichbar sind. Die ursprüngliche Erhebungsgesamtheit wird so in Gruppen zerlegt, bei denen die Personen (oder Objekte) in einer Gruppe eine hohe Homogenität aufweisen *(Intragruppen-Homogenität).* Zwischen den Gruppen hingegen besteht eine hohe Heterogenität *(Intergruppen-Heterogenität).* Auf diese Weise können statistische Analysen pro Gruppe vorgenommen werden, die dann pro Gruppe deutlich verlässlichere Ergebnisse liefern.

Die Clusteranalyse ist das methodische Instrument, um heterogene Erhebungsergebnisse in homogene Gruppen zu zerlegen. Sie ist in vielen Disziplinen wie Medizin, Soziologie, Biologie oder Ökonomie anwendbar und wird hier verwendet, um Ähnlichkeiten z. B. von Patienten, Käufern, Pflanzenarten, Unternehmen oder Produkten zu bestimmen. Tab. 8.1 zeigt ausgewählte Fragestellungen in unterschiedlichen Anwendungsfeldern, die alle auf die Bildung von Gruppen von Personen bzw. Objekten abzielen.

Die Fragen machen deutlich, dass die Clusteranalyse zu den explorativen Datenanalyseverfahren gehört, da sie im *Ergebnis* zu Vorschlägen für eine Gruppierung erhobener Untersuchungsobjekte führt und damit „neue Erkenntnisse“ generiert bzw. Strukturen in Datensätzen entdeckt.