

# Hauptkomponentenanalyse zur Datenreduktion

Die Hauptkomponentenanalyse wird häufig eingesetzt, um die Menge beobachteter Variablen auf eine geringere Anzahl Faktoren (Hauptkomponenten) so zu reduzieren, dass der Informationsverlust möglichst gering ist. Wesentliche Schritte sind die Bestimmung der Anzahl der Hauptkomponenten und deren Interpretation.

## Faktorladungs- bzw. Komponentenmatrix

Zentrales Ergebnis gleich welchen Verfahrens der Faktorenanalyse ist die Faktorladungsmatrix. Sie zeigt die Zusammenhänge zwischen den in die Analyse eingehenden Variablen und den extrahierten Faktoren auf. In Abbildung 1 ist eine Faktorladungsmatrix (Komponentenmatrix) mit allen möglichen Hauptkomponenten für sieben Merkmale aus einer Studie zur Zufriedenheit von Strandurlaubern dargestellt (Schimmelpfennig 2016).



**Johannes Lüken**, Diplom-Psychologe, ist Leiter des Bereichs Data Sciences bei IfaD.

jlueken@ifad.de



**Prof. Dr. Heiko Schimmelpfennig**, Diplom-Kaufmann, ist Projektleiter für Data Sciences bei IfaD.

hschimmelpfennig@ifad.de

Variable	Komponente						
	1	2	3	4	5	6	7
Wetter	0,47	0,45	0,53	0,49	0,18	-0,14	0,05
Strand	0,68	0,45	-0,06	-0,35	-0,03	-0,11	-0,45
Wasserqualität	0,62	0,54	-0,26	-0,20	0,02	0,23	0,40
Unterhaltungsangebot	0,68	-0,28	0,25	0,13	-0,55	0,27	-0,04
Sportangebot	0,72	-0,31	-0,30	0,07	-0,12	-0,50	0,16
Wellnessangebot	0,64	-0,33	-0,36	0,34	0,38	0,26	-0,17
Shopping-Möglichkeiten	0,49	-0,47	0,46	-0,47	0,31	0,02	0,10
Varianz der Komponente	2,70	1,19	0,85	0,76	0,59	0,48	0,43
Anteil an Gesamtvarianz	39%	17%	12%	11%	8%	7%	6%
kumuliert	39%	56%	68%	79%	87%	94%	100%

Abbildung 1: Beispielhafte Faktorladungsmatrix als Ergebnis einer Hauptkomponentenanalyse

Eine Faktorladung entspricht der Korrelation zwischen einer Variable und einem Faktor. Per Konvention gilt eine Ladung ab 0,5 als hoch. Die quadrierte Faktorladung ist der Anteil der Varianz einer Variablen, der durch einen Faktor erklärt wird. Der durch die Faktoren erklärte Anteil der Varianz einer Variable wird als Kommunalität bezeichnet. Für die Variable Wetter beträgt sie beispielsweise für die ersten beiden Komponenten  $0,47^2 + 0,45^2 = 0,42$ . Das Prinzip der Hauptkomponentenanalyse ist es, dass die Kommunalität jeder Variablen bei Berücksichtigung aller Komponenten gleich eins ist. Das heißt ihre gesamte Varianz kann durch die Komponenten erklärt werden.

Bei standardisierten Variablen entspricht die Summe der quadrierten Faktorladungen über alle Variablen der Varianz einer Hauptkomponente. Die Gesamtvarianz ist dann gleich der Anzahl der Variablen, da jede standardisierte Variable eine Varianz von eins besitzt. Der Quotient ergibt somit den Anteil der durch eine Hauptkomponente erklärten Varianz aller Variablen.

## Anzahl der Hauptkomponenten

Grundsätzlich besteht ein Trade-off zwischen dem Wunsch nach Datenreduktion und möglichst wenig In-

formationsverlust. Die Empfehlung von Kaiser (1960) ist ein gängiges Kriterium, an dem sich bei der Entscheidung orientiert wird, wie viele der möglichen Faktoren tatsächlich genutzt werden: Es sind die Hauptkomponenten zu verwenden, deren Varianz größer als eins und damit größer als die einer einzelnen standardisierten Variable ist. In dem Beispiel trifft dies auf die ersten beiden Komponenten zu. Andere Faustregeln geben vor, dass die Hauptkomponenten zusammen zum Beispiel mindestens 2/3 der Gesamtvarianz erklären sollen. Aber auch die Interpretierbarkeit des Ergebnisses darf nicht außer Acht gelassen werden.

## Interpretation der Hauptkomponenten

Fast alle Variablen in Abbildung 1 „laden hoch auf“ die erste Hauptkomponente. Sie ist eine Art gewichteter Mittelwert der sieben Merkmale, für die die Zufriedenheit erhoben wurde. Das heißt mithilfe der ersten Hauptkomponente können die Befragten anhand ihrer Gesamtzufriedenheit differenziert werden. Die Ladungen der zweiten Hauptkomponente sind positiv und negativ. Ihre Werte sind hoch bei Urlaubern, die zufrieden mit den Bedingungen für Strand- und unzufrieden mit sonstigen Aktivitäten waren, und gering im umgekehrten Fall.

Leichter sind Faktoren zu interpretieren, sofern die Faktorladungsmatrix eine Einfachstruktur aufweist. Wenn Faktoren bei einigen Variablen eine hohe und bei den übrigen Variablen eine geringe Ladung haben, können „einfach“ Gruppen von Variablen identifiziert werden, die zusammenhängen. Um eine Faktorladungsmatrix in eine Matrix zu transformieren, die der Einfachstruktur möglichst nahe kommt, gibt es eine Vielzahl an Verfahren. Das gängigste ist die Varimax-Rotation: Die Varianz der quadrierten Ladungen eines Faktors wird maximiert unter der Bedingung, dass die Kommunalitäten gleich bleiben. Für das Beispiel ergibt sich mit zwei Hauptkomponenten die in Abbildung 2 dargestellte Faktorladungsmatrix.

Die durch die beiden Komponenten zusammen erklärte Varianz in Höhe von 56 % bleibt gleich, sie verteilt sich nur anders auf die Komponenten. Komponente 1 repräsentiert eindeutig die Zufriedenheit mit den sonstigen Aktivitäten in einem Strandurlaub, Komponente 2 die Zufriedenheit mit den Bedingungen für Strandaktivitäten. ■

Variable	Komponente		Kommunalität
	1	2	
Wetter	0,07	<b>0,65</b>	0,42
Strand	0,23	<b>0,78</b>	0,66
Wasserqualität	0,13	<b>0,82</b>	0,68
Unterhaltungsangebot	<b>0,70</b>	0,23	0,54
Sportangebot	<b>0,75</b>	0,22	0,61
Wellnessangebot	<b>0,70</b>	0,16	0,51
Shopping-Möglichkeiten	<b>0,68</b>	-0,04	0,46
Varianz der Komponente	2,07	1,82	
Anteil an Gesamtvarianz	30%	26%	
kumuliert	30%	56%	

Abb. 2: Faktorladungsmatrix mit zwei Komponenten nach Anwendung der Varimax-Rotation



## Literatur

Kaiser, H. F.: *The Application of Electronic Computers to Factor Analysis*. In: Educational and Psychological Measurement, Nr. 1/1960, S. 141-151.

Schimmelpfennig, H.: *Bekannte, aktuelle und neue Anforderungen an Treiberanalysen*. In: Keller, B. et al. (Hrsg.): *Marktforschung der Zukunft – Mensch oder Maschine?*, Wiesbaden, 2016, S. 231-243.