Empirische Wirtschaftsforschung

Grundlagen - Methoden - Beispiele

Bearbeitet von
Prof. Dr. Robert Galata, Prof. Dr. Markus Wessler, Dr. Sandro Scheid, Rita Augustin

1. Auflage 2013 2013. Taschenbuch. 242 S. Paperback
ISBN 978 3 446 43227 7
Format (B x L): 16,7 x 24,2 cm
Gewicht: 465 g

HANSER

Leseprobe

Robert Galata, Markus Wessler, Sandro Scheid, Rita Augustin

Empirische Wirtschaftsforschung

Grundlagen, Methoden, Beispiele

Herausgegeben von Robert Galata, Markus Wessler

ISBN (Buch): 978-3-446-43227-7

ISBN (E-Book): 978-3-446-43783-8

Weitere Informationen oder Bestellungen unter

http://www.hanser-fachbuch.de/978-3-446-43227-7

sowie im Buchhandel.

6 Faktorenanalyse

6.1 Grundlagen

Es kommt vor, dass man Zusammenhänge zwischen vielen beobachteten Variablen auf einige wenige hinter den Variablen vermutete Wahrnehmungsdimensionen, sogenannte *Faktoren*, zurückführen möchte. Man spricht hier von einem *dimensionsreduzierenden Verfahren*, und die erwähnten Zusammenhänge, die eine solche Reduktion erlauben, sind bei den unter dem Namen *Faktorenanalyse* zusammengefassten vielfältigen Verfahren meist linear. Hat man zu Beginn noch keine konkreten Vermutungen über die Anzahl der einem Datensatz zugrunde liegenden Faktoren oder über die Zuordnung der beobachteten Variablen zu den Faktoren, so spricht man von einer *explorativen Faktorenanalyse*, und nur mit solchen Verfahren werden wir uns in diesem Rahmen beschäftigen: Verfahren also, die Zusammenhänge erst entdecken und nicht von vornherein eine Vermutung bestätigen wollen.

Wendet man ein solches Verfahren der Faktorenanalyse, etwa eine *Hauptkomponentenanalyse*, worauf wir uns hier konzentrieren wollen, an, so sucht man nach einer „möglichst einfachen Struktur“, die einen (großen) Satz gegebener Daten „hinreichend genau“ reproduziert. Viel Raum bietet das, viele Interpretationsmöglichkeiten, und man muss sich hier wie auch sonst von dem Gedanken verabschieden, dass es nur einen einzig richtigen Weg gibt.

Ein wesentliches Problem bei einem Dimensionsreduktionsverfahren wie der Hauptkomponentenanalyse besteht auch immer in der Ungewissheit, ob und zu welchem Grad die Faktoren selbst wiederum miteinander korrelieren. Man unterscheidet modelltheoretisch hauptsächlich zwischen zwei Arten von Verfahren: der *Hauptkomponentenanalyse* und der *Hauptachsenanalyse*. Rein rechentechnisch unterscheiden sich dabei die Verfahren tatsächlich nicht; beide konstruieren eine Matrix $\boldsymbol{A}$, die sogenannte *Faktorladungsmatrix*, mit deren Hilfe die Korrelationsmatrix $\boldsymbol{R}$ der standardisierten Daten möglichst gut reproduziert werden kann. Daher reicht es von einem mathematischen Standpunkt aus betrachtet auch wirklich aus, sich beispielsweise auf die Hauptkomponentenanalyse zu konzentrieren. Dies werden wir im Folgenden auch tatsächlich tun und das Verfahren der Hauptkomponentenanalyse ausführlich vorstellen und an Beispielen und Anwendungen erläutern.

Da von einem statistischen Standpunkt aus betrachtet die beiden Verfahren allerdings von völlig unterschiedlichen theoretischen Modellen ausgehen, wollen wir zur Hauptachsenanalyse und ihrer Abgrenzung zur Hauptkomponentenanalyse am Ende dieses Abschnitts wenigstens einige wichtige Punkte zusammenstellen. Eine erste Abgrenzung bereits jetzt:

Die *Hauptkomponentenanalyse* ist ein reines Datenreduktionsverfahren, das davon ausgeht, dass die gesamte Varianz einer Variable durch gemeinsame Faktoren erklärt werden kann bzw. (da mathematisch klar ist, dass das bei weniger Faktoren als Variablen nicht möglich ist) dass der durch die Faktoren nicht erklärbare Varianzanteil als bewusst in Kauf genommener Informationsverlust toleriert wird. Die Faktoren lassen sich als Linearkombinationen aus den Variablen darstellen und, so nimmt man an, korrelieren untereinander nicht. Man geht bei der

Hauptkomponentenanalyse von dem Modell $\boldsymbol{R} = \boldsymbol{A} \cdot \boldsymbol{A'}$ aus, wobei die Matrix $\boldsymbol{A}$ die Faktorladungen enthält.

Bei der *Hauptachsenanalyse* geht man davon aus, dass die Varianz der einzelnen Variablen sich immer in einen durch die Faktoren erklärbaren Anteil (die *Kommunalität*) sowie eine Restvarianz (also eine variablenspezifische Varianz und eine eventuelle Messfehlervarianz) zerlegen lässt. In der Praxis muss daher zu Beginn eine Schätzung der Kommunalitäten stehen. Man geht bei der Hauptachsenanalyse von dem Modell $\boldsymbol{R} = \boldsymbol{A} \cdot \boldsymbol{A'} + \boldsymbol{U}$ aus, wobei die Matrix $\boldsymbol{A}$ die Faktorladungen enthält und U für die Restvarianz steht.

In der Vorgehensweise ist begründet, dass die Faktorladungen bei einer Hauptkomponentenanalyse in der Regel etwas höher sind als bei einer Hauptachsenanalyse. In der Literatur ist die Verwendung der Begriffe *Faktorenanalyse*, *Hauptkomponentenanalyse* und *Hauptachsenanalyse* nicht ganz eindeutig und häufig etwas verwirrend. Manchmal wird *Faktorenanalyse* mit *Hauptkomponentenanalyse* und manchmal mit *Hauptachsenanalyse* identifiziert, was beides nicht ganz korrekt ist.

6.1.1 Einige Zerlegungen

In einer verständlichen Formulierung eines Verfahrens der Faktorenanalyse kommt man ohne Matrizen nur schwer aus, und wir nutzen ihre Sprache im Folgenden sehr intensiv. Bei den Eigenwerten sind Matrizen für das Grundverständnis absolut notwendig, aber auch die bereits formulierten Grundideen zur Faktorenanalyse lassen sich mithilfe von Matrizen sehr bequem und übersichtlich ausdrücken und formalisieren. Besonders zentral dabei sind gewisse Zerlegungen von Matrizen, mit denen wir uns in diesem ersten Unterabschnitt näher beschäftigen wollen.

Der Ausgangspunkt eines dimensionsreduzierenden Verfahrens wie der Faktorenanalyse ist ein Datensatz für eine gewisse (und eben in der Regel große) Anzahl von Variablen, die wir mit $x_1, \ldots, x_n$ bezeichnen. Für jede dieser Variablen x_i liegt nun im Datensatz mit den Werten $x_{i,1}, \ldots, x_{i,m}$ eine Menge von jeweils m erhobenen Ausprägungen vor. Diese Daten können auf die bekannte Weise *standardisiert* werden: Ist $\overline{x_i}$ der Mittelwert der zur Variable x_i erhobenen Werte, also

$$\overline{x_i} = \frac{1}{m} \sum_{j=1}^{m} x_{i,j},$$

und ist $\sigma_i^2 \neq 0$ die Varianz der Variable x_i, so ersetzt man x_i durch die standardisierte Variable

$$z_i = \frac{x_i - \overline{x_i}}{\sigma_i}.$$

Der Mittelwert dieser standardisierten Werte ist dann $\overline{z_i} = 0$ und die Varianz $r_i^2 = 1$. Ordnet man nun den Zeilen die standardisierten Variablen und den Spalten deren Ausprägungen zu, so liegt der Datensatz in Form einer $(n \times m)$-Matrix $\boldsymbol{Z}$ vor, also einer Matrix mit n Zeilen und m Spalten.

Man interessiert sich natürlich auch für die Korrelation der verschiedenen Variablen und fasst die entsprechenden Korrelationskoeffizienten $\varrho_{i,j}$ in der sogenannten *Korrelationsmatrix* zusammen.

Korrelationsmatrix R eines Datensatzes

Zu einem wie oben beschriebenen Datensatz bezeichnen wir mit

$$R = \begin{pmatrix} 1 & \varrho_{1,2} & \cdots & \varrho_{1,n} \\ \varrho_{1,2} & 1 & \cdots & \varrho_{2,n} \\ \vdots & & & \vdots \\ \varrho_{1,n} & \varrho_{2,n} & \cdots & 1 \end{pmatrix} \tag{6.1}$$

die *Korrelationsmatrix*. Es handelt sich bei $\boldsymbol{R}$ bekanntermaßen um eine symmetrische Matrix mit je n Zeilen und Spalten, und die Diagonaleinträge sind gleich eins.

Es gibt nun einen ersten zentralen Zusammenhang zwischen den bisher betrachteten Matrizen; die Korrelationsmatrix $\boldsymbol{R}$ lässt sich mithilfe der Matrix Z der standardisierten Daten in folgender Weise zerlegen:

Zerlegung der Korrelationsmatrix R mithilfe der Matrix Z der standardisierten Daten

$$\boldsymbol{R} = \frac{1}{m} \cdot \boldsymbol{Z} \cdot \boldsymbol{Z}'. \tag{6.2}$$

Beispiel 6.1

In einem Autohaus haben in einem gewissen Zeitraum acht Kunden ($m = 8$) einen Neuwagen bestimmten Typs gekauft. Diese Kunden wurden gebeten, durch Angabe einer Punktzahl zwischen 0 und 10 auszudrücken, wie wichtig ihnen die folgenden vier Merkmale ($n = 4$) beim Kauf eines Autos waren: Anschaffungspreis (x_1), Stauraum (x_2), Betriebskosten (x_3) und Motorleistung (x_4). Die Daten der Befragung liegen in folgender Tabelle vor:

x_1	3	2	5	3	6	6	6	3
x_2	4	8	6	2	5	7	5	4
x_3	2	2	3	3	6	5	5	3
x_4	6	2	2	8	5	5	3	7

Standardisiert man diese Werte wie oben erklärt, so ergibt sich

$$\boldsymbol{Z} = \begin{pmatrix} -0,8006 & -1,4412 & 0,4804 & -0,8006 & 1,1209 & 1,1209 & 1,1209 & -0,8006 \\ -0,6380 & 1,6304 & 0,4962 & -1,7722 & -0,0709 & 1,0633 & -0,0709 & -0,6380 \\ -1,1536 & -1,1536 & -0,4437 & -0,4437 & 1,6860 & 0,9761 & 0,9761 & -0,4437 \\ 0,5934 & -1,3054 & -1,3055 & 1,54282 & 0,1187 & 0,1187 & -0,8308 & 1,0681 \end{pmatrix} \tag{6.3}$$

Aus Platzgründen beschränken wir uns hierbei auf die Angabe von vier Dezimalstellen, führen aber die folgenden Berechnungen, um Rundungsfehler klein zu halten, mit exakteren Werten durch. Berechnet man für die Matrix Z nun das Produkt mit ihrer

$$\begin{array}{cc} \cdot & \begin{pmatrix} f_{1,1} & \cdots & f_{1,j} & \cdots & f_{1,m} \\ f_{2,1} & \cdots & f_{2,j} & \cdots & f_{2,m} \\ \vdots & & \vdots & & \vdots \\ f_{p,1} & \cdots & f_{p,j} & \cdots & f_{p,m} \end{pmatrix} = \boldsymbol{F} \\ \boldsymbol{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,p} \\ \vdots & \vdots & & \vdots \\ a_{i,1} & a_{i,2} & \cdots & a_{i,p} \\ \vdots & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,p} \end{pmatrix} & \begin{pmatrix} & \vdots & \\ & \vdots & \\ \cdots & z_{i,j} = a_{i,1} \cdot f_{1,j} + \ldots + a_{i,p} \cdot f_{p,j} & \cdots \\ & \vdots & \\ & \vdots & \end{pmatrix} = \boldsymbol{Z} \end{array}$$

Bild 6.1 Illustration der Zerlegung (6.4): Die Ausprägung $z_{i,j}$ ist eine *Linearkombination* in den Faktorwerten $f_{k,j}$ mit den *Faktorladungen* $a_{i,k}$ als Gewichten.

Transponierten, so ergibt sich

$$\boldsymbol{Z} \cdot \boldsymbol{Z}' = \begin{pmatrix} 8 & 1,362142064 & 7,161374285 & -1,976386421 \\ 1,362142064 & 8 & 0,553546709 & -6,393782811 \\ 7,161374285 & 0,553546709 & 8 & -0,252743623 \\ -1,976386421 & -6,393782811 & -0,252743623 & 8 \end{pmatrix}.$$

Nach (6.2) – und wie man durch Nachrechnen tatsächlich verifiziert – ergibt die Division durch 8 die Korrelationsmatrix $\boldsymbol{R}$:

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0,170267758 & 0,895171786 & -0,247048303 \\ 0,170267758 & 1 & 0,069193339 & -0,799222851 \\ 0,895171786 & 0,069193339 & 1 & -0,031592953 \\ -0,247048303 & -0,799222851 & -0,031592953 & 1 \end{pmatrix}.$$

■

An der in *Beispiel 6.1* berechneten Korrelationsmatrix $\boldsymbol{R}$ kann man im Hinblick auf die später geplante Faktorenanalyse bereits einige Informationen ablesen: Hohe Korrelationswerte treten (abgesehen von den Selbstkorrelationen auf der Diagonale) zwischen den Variablen x_1 und x_3 (mit $\varrho \approx 0,895$) sowie zwischen den Variablen x_2 und x_4 (mit $\varrho \approx -0,799$) auf. Das passt inhaltlich, denn Preis und Betriebskosten zielen definitiv in die gleiche Richtung, und auch Stauraum und Motorleistung scheinen sich zu beeinflussen – allerdings deutet der negative Korrelationskoeffizient auf einen reziproken Zusammenhang hin: Wem der Stauraum wichtig ist (was auf Kunden mit Familie hindeuten könnte), dem scheint die Motorleistung (und damit die Geschwindigkeit) weniger wichtig zu sein, und umgekehrt. Will man nun die Menge der Variablen, die in diesem ersten Beispiel natürlich sehr übersichtlich ist, verringern und auf die wesentlichen Inhalte reduzieren, so wird man auf die Idee kommen, die Variablenpaare x_1 und x_3 sowie x_2 und x_4 „zusammenzufassen“ und sogenannte *Hauptkomponenten* zu identifizieren: die Grundidee der Hauptkomponentenanalyse.

Wir werden auf dieses übersichtliche Beispiel an einigen Stellen des Kapitels zurückkommen und die im Folgenden entwickelten Ideen daran verdeutlichen. Zunächst wollen wir aber die

Ziele der verschiedenen Verfahren der Faktorenanalyse formulieren. Diese Ziele lassen sich sehr kompakt ebenfalls als Zerlegungen von Matrizen formulieren, so lässt sich etwa in der sogenannten *Fundamentalgleichung der Faktorenanalyse* die Matrix Z und damit auch $\boldsymbol{Z'}$ weiter zerlegen, und zwar in der folgenden Weise:

Fundamentalgleichung der Faktorenanalyse

Die Zerlegung der $(n \times m)$-Matrix Z in

$$\boldsymbol{Z} = \boldsymbol{A} \cdot \boldsymbol{F} \tag{6.4}$$

für eine $(n \times p)$-Matrix $\boldsymbol{A}$ und eine $(p \times m)$-Matrix F heißt (unter gewissen, noch zu spezifizierenden Bedingungen) die *Fundamentalgleichung der Faktorenanalyse*. Man nennt

$$\boldsymbol{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,p} \\ a_{2,1} & a_{2,2} & \dots & a_{2,p} \\ \vdots & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,p} \end{pmatrix} \tag{6.5}$$

die *Faktorladungsmatrix* und

$$\boldsymbol{F} = \begin{pmatrix} f_{1,1} & f_{1,2} & \dots & f_{1,m} \\ f_{2,1} & f_{2,2} & \dots & f_{2,m} \\ \vdots & \vdots & & \vdots \\ f_{p,1} & f_{p,2} & \dots & f_{p,m} \end{pmatrix} \tag{6.6}$$

die *Faktorwertematrix*.

Hierzu gleich eine wichtige Bemerkung: Die Zerlegung (6.4) ist natürlich zunächst in keinster Weise eindeutig. Wenn man so will, besteht genau darin die Schwierigkeit der Zerlegung, dass sie einer gewissen Willkür unterworfen ist. Eine banale Idee wäre etwa, $\boldsymbol{A} = \boldsymbol{Z}$ und $\boldsymbol{F} = \boldsymbol{I}$ zu wählen (wobei $\boldsymbol{I}$ die Einheitsmatrix bezeichnet), was aber wenig zielführend wäre. Später mehr dazu.

Mit den üblichen Doppelindizierungen bedeutet die Gleichung (6.4) übrigens nichts anderes als

$$z_{i,j} = a_{i,1} \cdot f_{1,j} + \ldots + a_{i,p} \cdot f_{p,j}. \tag{6.7}$$

Wie dies auch noch einmal in *Bild 6.1* illustriert ist, lässt sich jede Ausprägung $z_{i,j}$ der standardisierten Variable z_i also als *Linearkombination* in den Faktorwerten $f_{k,j}$ $(k = 1, \dots p)$ ausdrücken, wobei sich die *Faktorladungen* $a_{i,k}$ $(k = 1, \dots p)$ als *Gewichte* auffassen lassen. Die in der Matrix $\boldsymbol{F}$ zusammengefassten Faktorwerte sind, wenn man so will, die „Positionen" der Merkmalsträger auf den Faktoren. Genauer gesagt ist der Faktorwert $f_{k,j}$ ein Maß für die Stärke der Ausprägung des Faktors k (besser: der Ausprägungen der im Faktor k zusammengefassten Variablen) beim Merkmalsträger j.

Die Schwierigkeit bei der Zerlegung besteht darin, dass sie uneindeutig ist: Die bisher erwähnten Forderungen legen die Matrizen $\boldsymbol{A}$ und $\boldsymbol{F}$ in keinster Weise fest. Welche weiteren Bedingun-

gen sollen $\boldsymbol{A}$ und $\boldsymbol{F}$ erfüllen? Darüber muss man sich einigen, um dann Methoden entwickeln zu können, wie man die beiden Matrizen findet.

Die wesentliche Idee hierbei ist es, *formal von der Existenz einer Zerlegung* (6.4) *auszugehen* und Bedingungen an $\boldsymbol{A}$ zu finden. Diese Absicht vor Augen kombiniert man etwa die beiden Gleichungen (6.2) und (6.4) und erhält

$$\boldsymbol{R} = \frac{1}{n} \cdot (\boldsymbol{A} \cdot \boldsymbol{F}) \cdot (\boldsymbol{A} \cdot \boldsymbol{F})' = \frac{1}{n} \cdot \boldsymbol{A} \cdot \boldsymbol{F} \cdot \boldsymbol{F}' \cdot \boldsymbol{A}'.$$

Klammert man hier gemäß dem Assoziativgesetz um, ergibt sich

$$\boldsymbol{R} = \boldsymbol{A} \cdot \left(\frac{1}{m} \cdot \boldsymbol{F} \cdot \boldsymbol{F}' \right) \cdot \boldsymbol{A}'. \tag{6.8}$$

In der Mitte dieses Ausdrucks steht nun mit $\frac{1}{n} \cdot \boldsymbol{F} \cdot \boldsymbol{F}'$ wiederum ein Matrizenprodukt, das zu einer Korrelationsmatrix führt, diesmal zur Korrelationsmatrix der *Faktoren*. Nun kommt ein wesentlicher Gedanke: Man könnte zunächst die weitere formale Annahme treffen, dass die Faktoren als *weitestgehend unkorreliert* aufgefasst werden. In diesem Fall würde gelten

$$\frac{1}{m} \cdot \boldsymbol{F}' \cdot \boldsymbol{F} \approx \boldsymbol{I}, \tag{6.9}$$

und damit vereinfacht sich die Gleichung (6.8) noch einmal erheblich:

$$\boldsymbol{R} \approx \boldsymbol{A} \cdot \boldsymbol{A}'. \tag{6.10}$$

Das ist kurz gefasst die Grundidee der sogenannten *Hauptkomponentenanalyse*: das „Verhältnis“ der Faktoren als möglichst schlicht anzunehmen und sich auf die sehr einfache Form (6.10) der Zerlegung zu konzentrieren. Es gibt andere Varianten: Man könnte die Art und Weise wie gut die Approximation (6.10) ist, in einen Fehlerterm mit einbeziehen, in den auch eventuelle Kenntnisse über Verteilungsfunktionen mit einfließen. Solche Untersuchungen gehen dann eher in Richtung der sogenannten *Hauptachsenanalyse* – auf die Abgrenzung gehen wir später noch etwas näher ein.

Aus (6.10) jedenfalls – wie auch schon aus der Zerlegung (6.4) – wird klar, dass $\boldsymbol{A}$ so viele Zeilen haben muss, wie es Variablen gibt, also n. Da es sich bei (6.10) um eine Zerlegung analog zu (6.2) handelt, könnte man (wie bereits erwähnt) leicht auf die Idee kommen, einfach (bis auf einen Faktor) $\boldsymbol{A} = \boldsymbol{Z}$ zu wählen. Damit allerdings wäre nichts gewonnen: Der wesentliche Punkt ist, dass $\boldsymbol{A}$ die Informationen in gewisser Weise „verdichtet“ tragen soll, mit anderen Worten: dass $\boldsymbol{A}$ „schmal“ ist, also wenige Spalten hat. So lässt sich auch das Ziel der Hauptkomponentenanalyse formulieren:

Ziel der Hauptkomponentenanalyse

Gesucht ist in der Hauptkomponentenanalyse eine Matrix $\boldsymbol{A}$ mit „möglichst wenigen“ Spalten, die die Gleichung

$$\boldsymbol{R} = \boldsymbol{A} \cdot \boldsymbol{A}' \tag{6.11}$$

„möglichst exakt“ erfüllt.

Dabei steckt die Schwierigkeit genau in der Erfüllung der beiden „möglichst“-Formulierungen. Es sei noch einmal deutlich gesagt, dass wir hier, entgegen einer in der Literatur sonst häufig üblichen Vorgehensweise, einen weniger über Fehlerterm und Verteilungsannahmen, sondern mehr über Matrizenzerlegung und Eigenwerte motivierten Zugang wählen.

Im folgenden sehr einfachen Beispiel ist die Zerlegung (6.11) exakt erfüllt, und man kann die Matrizen $\boldsymbol{A}$ und $\boldsymbol{F}$ ohne viel Rechnung gleich angeben.

Beispiel 6.2

Gegeben sei der Datensatz

x_1	1	2	3	4	5	6
x_2	2	4	6	8	10	12
x_3	3	6	9	12	15	18

.

An der Struktur der Daten sieht man sofort, dass die Korrelationsmatrix durch

$$\boldsymbol{R} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

gegeben ist. Die Standardisierung der Daten ergibt

$$\boldsymbol{Z} = \begin{pmatrix} -1,46385 & -0,87831 & -0,29277 & 0,29277 & 0,87831 & 1,46385 \\ -1,46385 & -0,87831 & -0,29277 & 0,29277 & 0,87831 & 1,46385 \\ -1,46385 & -0,87831 & -0,29277 & 0,29277 & 0,87831 & 1,46385 \end{pmatrix}.$$

Die sehr spezielle Struktur von $\boldsymbol{R}$ legt den Gedanken nahe, dass $\boldsymbol{A}$ hier „sehr schmal“ ist, und der nahe liegende Versuch

$$\boldsymbol{A} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

führt auch tatsächlich zum Ziel: In der Tat ist damit die Gleichung $\boldsymbol{R} = \boldsymbol{A} \cdot \boldsymbol{A}'$ bereits exakt erfüllt; es reicht zur Reproduktion der Korrelationsmatrix ein einziger Faktor. Als Matrix der Faktorwerte ergibt sich durch Überlegung

$$\boldsymbol{F} = \begin{pmatrix} -1,46385 & -0,87831 & -0,29277 & 0,29277 & 0,87831 & 1,46385 \end{pmatrix}.$$

Dass damit schließlich die Fundamentalgleichung $\boldsymbol{Z} = \boldsymbol{A} \cdot \boldsymbol{F}$ erfüllt ist, ist auch offensichtlich; die *Bild 6.1* entsprechende Darstellung ist:

	$\cdot$	$\begin{pmatrix} -1,46385 & -0,87831 & -0,29277 & 0,29277 & 0,87831 & 1,46385 \end{pmatrix} = \boldsymbol{F}$
$\boldsymbol{A} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$		$\begin{pmatrix} -1,46385 & -0,87831 & -0,29277 & 0,29277 & 0,87831 & 1,46385 \\ -1,46385 & -0,87831 & -0,29277 & 0,29277 & 0,87831 & 1,46385 \\ -1,46385 & -0,87831 & -0,29277 & 0,29277 & 0,87831 & 1,46385 \end{pmatrix} = \boldsymbol{Z}$

■

So einfach wie in *Beispiel 6.2* wird es in der Realität natürlich niemals sein. Man muss Wege finden (und wir werden dies in den folgenden Abschnitten tun), um die Zerlegung (6.11) systematisch und möglichst genau zu finden. Außerdem sollten idealerweise aus der Matrix $\boldsymbol{A}$ nicht nur die (möglichst kleine) Anzahl p der Faktoren ablesbar sein, sondern auch, anhand der Einträge $a_{i,k}$, welche Variablen wie stark auf welchen Faktor „laden“ – das bedeutet, wie stark die Gewichte $a_{i,k}$ ausfallen. Diese sogenannten *Faktorladungen* werden sich, wie wir sehen werden, zwischen -1 und 1 bewegen; betragsmäßig große Werte sprechen hier wieder für einen großen Zusammenhang. Allerdings kann man auch bei einer recht guten Näherung der Form (6.11) aus $\boldsymbol{A}$ selbst in der Regel noch nicht unmittelbar ablesen, welche Variablen auf welchen Faktor laden – hierzu sind Methoden der *Rotation* erforderlich, auf die wir auch noch zu sprechen kommen werden.

Ist eine solche Matrix $\boldsymbol{A}$ schließlich gefunden, dann lassen sich schließlich auch die Faktorwerte, also die Einträge der Matrix $\boldsymbol{F}$ in der Zerlegung (6.4), angeben. Multipliziert man nämlich die Gleichung (6.4), also

$$\boldsymbol{A} \cdot \boldsymbol{F} = \boldsymbol{Z},$$

von links mit $\boldsymbol{A'}$, so ergibt sich

$$\boldsymbol{A'} \cdot \boldsymbol{A} \cdot \boldsymbol{F} = \boldsymbol{A'} \cdot \boldsymbol{Z},$$

und da wir davon ausgehen können, dass im Allgemeinen die quadratische Matrix $\boldsymbol{A'} \cdot \boldsymbol{A}$ invertierbar ist (in *Beispiel 6.2* war sie dies allerdings nicht!), können wir durch Multiplikation mit der Inversen nach F „auflösen“ und erhalten:

Berechnung der Faktorwertematrix $\boldsymbol{F}$

Hat man zu gegebener Matrix $\boldsymbol{Z}$ der standardisierten Daten eine zufriedenstellende Faktorladungsmatrix $\boldsymbol{A}$ gefunden, und ist die Matrix $\boldsymbol{A'} \cdot \boldsymbol{A}$ invertierbar, so erhält man die entsprechende Faktorwertematrix durch

$$\boldsymbol{F} = (\boldsymbol{A'} \cdot \boldsymbol{A})^{-1} \cdot \boldsymbol{A'} \cdot \boldsymbol{Z}. \tag{6.12}$$

Für die Methoden zur systematischen Konstruktion von $\boldsymbol{A}$ sind aber gewisse theoretische Kenntnisse über Eigenwerte erforderlich, die wir am Ende dieses Abschnitts zusammenstellen werden. Zunächst aber einige kurze Anmerkungen über die Unterscheidung von Hauptkomponentenanalyse und Hauptachsenanalyse.

6.1.2 Abgrenzung Hauptkomponentenanalyse und Hauptachsenanalyse

Nur kurz wollen wir uns in diesem Kapitel der Hauptachsenanalyse zuwenden, die sich, wie eingangs bereits erwähnt, als mathematisches Modell – und insbesondere bei dem hier gewählten Zugang über die Eigenwerte, den wir im Anschluss erklären, – wenig von der Hauptkomponentenanalyse unterscheidet. Wie dort ist es auch bei der *Hauptachsenanalyse* das erklärte Ziel, die Dimension zu reduzieren, und zwar wiederum mithilfe sogenannter verdeckter