· 28. 02. 2023

# Die Bestimmung latenter Variablen: die Hauptachsentransformation (PCA)

U. Mortensen

## Inhaltsverzeichnis

# 1 Ziel und Ansatz

Das Ziel der Hauptkomponentenanalyse (PCA = Principal Component Analysis) ist, Messungen von $n$ korrelierenden Variablen bei $m$ Fällen (Personen oder Objekten) durch $r \leq \min(m, n)$ nicht-korrelierende "latente", d.h. nicht direkt gemessene Variablen zu interpretieren. Die Messungen werden in einer $(m, n)$-Datenmatrix $X$ zusammengefasst; die Spalten repräsentieren die Variablen, die Zeilen die Fälle (Personen oder Objekte, an denen die jeweils $n$ Messungen vorgenommen wurden. Die Elemente $x_{ij}$ von $X$ werden als spaltenzentriert vorausgesetzt, d.h. $x_{ij} = X_{ij} - \bar{x}_j$, $X_{ij}$ der Messwert des $i$-ten Falls bei der $j$-ten Variable. Für den Fall, dass die Variablen in verschiedenen Maßeinheiten gemessen werden ist es ratsam, zu standardisierten Werten $z_i$ überzugehen, wobei $z_{ij} = x_{ij}/s_j$, $s_j$ die Standardabweichung der Messwerte für die $j$-te Variable. Die Matrix $X$ geht dann in die Matrix $Z = (z_{ij})$ über.

Der Ansatz der PCA kann anhand des Falles $n = 2$ illustriert werden. Sind die Messungen $X_1$ und $X_2$ korreliert, so ist die Konfiguration der Fälle "orientiert", d.h. der Regressionsparameter $b$ in der Regressionsgleichung $X_2 = bX_1 + a + e$, wobei $e$ ein Fehlerterm[1] ist, ist ungleich Null, vergl. Abbildung 1 (a). Nun seien $L_1$ und $L_2$ zwei senkrecht aufeinanderstehende Geraden, die ein alternatives Koordinatensystem bilden. Die Projektionen der Punkte, die die Fälle repräsentieren, auf $L_1$ seien $u_{11}, \ldots, u_{m1}$, und die Projektionen der Punkte auf die Achse $L_2$ seien $u_{12}, \ldots, u_{m2}$. Die Orientierung von $L_1$ und damit – wegen der geforderten Rechtwinkligkeit von $L_1$ und $L_2$ – $L_2$ sei so gewählt, dass die Kovarianz und damit die Kovarianz – d.h. das Skalarprodukt der Vektoren $\mathbf{u}_j = (u_{1j}, u_{2j}, \ldots, u_{mj})'$, $j = 1, 2$ – der $u_{i1}$ und $u_{i2}$-Werte gleich Null ist:

$$\mathbf{u}_2 = \mathbf{u}_1^\intercal \mathbf{u}_2 = \mathbf{u}_1' \mathbf{u}_2 = \sum_{i=1}^{m} u_{i1} u_{i2} = 0. \tag{1.1}$$

$L_1$ und $L_2$ repräsentieren dann voneinander unabhängige latente Merkmale[2]. Der Übergang vom $(X_1, X_2$- zum $L_1, L_2)$-System erfolgt durch eine Rotation der Achsen bzw. der Konfiguration. Es kann dann gezeigt[3] werden, dass die Rotation der Konfiguration auf unkorrelierte Achsen impliziert, dass $L_1$ diejenige Achse ist, in Bezug auf die die Punktekonfiguration ihre maximale Ausdehnung hat, und $L_2$ ist die Orientierung der zweitmaximalen Ausdehnung. Diese Eigenschaften

[1] $e$ ist hier nicht die Eulersche Zahl, sondern entspricht dem in englischsprachingen Texten üblichen Bezeichnung für *error*, also Fehler.

[2] Der Ausdruck 'unabhängig' gilt allerdings nur umgangssprachlich und ist mit Vorsicht zu gebrauchen, da Kovarianzen bzw. Korrelationen auch im Falle deterministisch abhängiger Vriablen gleich Null sein können, d.h. 'unabhängig' kann nicht ohne Weiteres mit 'stochastisch unabhängig' gleichgesetzt werden. Stochastische Unabhängigkeit setzt die Wahl einer bestimmten Verteilung aus der Menge aller Verteilungen voraus, etwa der Exponential- oder der Gamma-Verteilung, insbesondere der Gauß-Verteilung, die wegen ihrer speziellen Eigenschaften besonders beliebt ist. Eine Verteilung diesen Typs wird aber im Folgenden nicht vorausgesetzt. Gleichwohl sollte man bei der Gleichsetzung von 'unabhängig' und 'Korreltion gleich Null' vorsichtig sein.

[3] etwa http://www.uwe-mortensen.de/LineareAlgebraNeua.pdf, Einführung in die Vektor- und Matrixrechnung für die multivariate Statistik, Abschnitt 3.1

von $L_1$ und $L_2$ setzen nicht die multivariate Gauß-Verteilung der Daten voraus. Abbildung 1 (b) illustriert den Sachverhalt. Wählt man $L_1$ und $L_2$ als neues Koordinatensystem, so erhält man als Repräsentation der Daten die Abbildung 1 (c). In diesem Koordinatensystem ergäbe sich ein Regressionskoeffizient mit dem Wert Null, da die Konfiguration nun "achsenparallel" ist. Diese Betrachtungen lässt sich leicht auf $n > 2$ gemessene Variablen $X_1, \ldots, X_n$ verallgemeinern. Die Verallgemeinerung gelingt leicht durch Anwendung der Vektor- und Matrixrechnung[4].

Die Geraden $L_1$ und $L_2$ ergeben sich *nicht* durch Anwendung der Regressionsrechnung. Im 2-dimensionalen Fall ergeben sich bekanntlich zwei Regressionsgeraden, während es nur eine Gerade $L_1$ gibt, die die Orientierung der maximalen Ausdehnung der Konfiguration hat. Zu bestimmen sind also die Vektoren $\mathbf{u}_1, \ldots, \mathbf{u}_n$, deren Komponenten die Koordinaten der Punkte der Konfiguration der Fälle auf den $L_k$ repräsentieren, $k = 1, \ldots, n$. Die Frage, ob $r < \min(m, n)$ latente Variaben genügen, um die Daten zu repräsentieren muß gesondert diskutiert werden.

Natürlich ist man nicht nur an einer Darstellung der Fälle in einem latente Variable abbildenden Koordinatensystem interessiert, dass primäre Interesse richtet sich oft auf eine Repräsentation der Variablen in einem entsprechenden Koordinatensystem. Diese Darstellung ergibt sich aus den folgenden Betrachtungen.

Der eben beschriebende Ansatz läßt sich auch anders formulieren: gesucht sind orthogonale "latente Vektoren" $\mathbf{L}_1, \ldots, \mathbf{L}_n$ derart, dass sich die Vektoren $\mathbf{z}_j$, deren Komponenten die zentrierten oder standardisierten Messungen der $j$-ten Variablen bei den $i = 1, \ldots, m$ Fällen sind, als Linearkombination der $\mathbf{L}_k$ darstellen lassen:

$$\mathbf{z}_j = t_{1j}\mathbf{L}_1 + \cdots + t_{nj}\mathbf{L}_n. \tag{1.2}$$

Diese Gleichung entspricht einer Regressionsgleichung ohne Fehlerterm. Die Koeffizienten $t_{1j}, \ldots, t_{nj}$ sind, wie aus der Indizierung hervorgeht, variablenspezifische Größen, die wie die $\mathbf{L}_k$ zunächst unbekannt sind und wie diese aus den Daten zu schätzen sind. Fasst man die $\mathbf{z}_j$ zu einer Matrix $Z = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ zusammen und die $\mathbf{L}_k$ zu einer Matrix $L = [\mathbf{L}_1, \ldots, \mathbf{L}_n]$, so kann man den Ansatz (1.2) zu einer Matrixgleichung

$$Z = LT' \tag{1.3}$$

zusammenfassen; hierin ist $Z$ eine $(m, n)$-Matrix, $L$ ist ebenfalls eine $(m, n)$-Matrix, und $T$ ist eine $(n, n)$-Matrix. Für $\mathbf{z}_j$ kann man dann $\mathbf{z}_j = L\tilde{\mathbf{t}}_j$ schreiben, wobei $\tilde{\mathbf{t}}_j = (t_{1j}, \ldots, t'_{nj}$ der $j$-te Spaltenvektor von $T'$ ist. Allgemein werden die Spaltenvektoren von transponierten Matrizen mit einer Tilde gekennzeichnet. $\tilde{\mathbf{z}}_i$ ist also der $i$-te Spaltenvektor von $Z'$, d.h. $\tilde{\mathbf{z}}'_i$ ist der $i$-Zeilenvektor von $Z$. Analog dazu ist $\tilde{\mathbf{L}}_i$ der $i$-te Spaltenvektor von $L'$.

Um die Matrizen $L$ und $T$ zu bestimmen, müssen die oben in der intuitiven Beschreibung der PCA gemachten Annahmen formalisiert werden. Demnach müssen

[4]Eine Zusammenstellung der benötigten Elemente dieser Rechnung findet man im genannten Skriptum zur Vektor- und Matrixrechnung.

die $\mathbf{L}_k$ orthogonal sein: sie sollen ja die Orientierung der paarweise orthogonalen Geraden $L_1, \ldots, L_n$ haben. Darüber hinaus muß die Matrix $T$ spezifiziert werden. Dazu muß man nur berücksichtigen, dass der Übergang von der Darstellung in Abbildung 1 (a) zu Abbildung 1 (c) in einer Rotation der Punktekonfiguration besteht, d.h. in eines Rotation jeden Vektors $\tilde{\mathbf{z}}_i$ um einen bestimmen Winkel $\theta$. Transponiert man die Gleichung in (1.3), so erhält man die Gleichung

$$Z' = TL', \text{ d.h. } \tilde{\mathbf{z}}_i = T\tilde{\mathbf{L}}_i, \ i = 1, \ldots, m \tag{1.4}$$

Die Gleichung definiert die Transformation der Vektoren $\tilde{\mathbf{L}}_i$ in die korresponierenden Vektoren $\tilde{\mathbf{z}}_i$. Diese Transformation soll also eine Rotation sein. Dies bedeutet, dass $T'T = I$, $I$ die $(n, n)$-Einheitsmatrix sein. Zusammenfassend kann man die Annahmen in der Form

**A1** Die latenten Achsen $L_k$ gehen aus den Achsen $X_k$ durch Rotation hervor, d.h. $V$ ist orthonormal,
**A2** Die $L_k$ und damit die $\mathbf{L}_k$ sind paarweise orthgonal, so dass

$$L'L = \text{diag}(\lambda_1, \ldots, \lambda_n), \ \lambda_k = \mathbf{L}_k'\mathbf{L}_k = \|\mathbf{L}_k\|^2. \tag{1.5}$$

zusammenfassen. Es sei daraufhingewiesen, dass die Rotation zwar durch einen Winkel $\theta$ bestimmt ist, dieser Winkel aber nicht explizit in den Annahmen genannt wird; er wird implizit durch die Annahmen bestimmt und muß nicht explizit berechnet werden. Es zeigt sich ebenfalls, dass die in der oben dargestellen intuitiven Charakterisierung aufgestellten Forderung, dass die Achse $L_1$ die Orientierung der maximalen Ausdehung der Konfiguration der Fälle haben soll, von den beiden Annahmen **A1** und **A2** impliziert wird.

Es werden zunächst die Vektoren $\mathbf{z}_j$ und $\tilde{\mathbf{z}}_i$ betrachtet:

$$\mathbf{z}_j = L\tilde{\mathbf{t}}_j, \quad \|\mathbf{z}_j\|^2 = \mathbf{z}_j'\mathbf{z}_j = \tilde{\mathbf{t}}_j L'L\tilde{\mathbf{t}}_j = \tilde{\mathbf{t}}_j \Lambda \tilde{\mathbf{t}}_j \tag{1.6}$$

$$\tilde{\mathbf{z}}_i = T\tilde{\mathbf{L}}_i, \quad \|\tilde{\mathbf{z}}_i\|^2 = \tilde{\mathbf{z}}_i'\tilde{\mathbf{z}}_i = \tilde{\mathbf{L}}_i' T'T\tilde{\mathbf{L}}_i = \tilde{\mathbf{L}}_i'\tilde{\mathbf{L}}_i = \|\tilde{\mathbf{L}}_i\|^2 \tag{1.7}$$

Die Gleichung (1.7) zeigt, dass $\tilde{\mathbf{z}}_i$ und $\tilde{\mathbf{L}}_i$ dieselbe Länge haben, was der Tatsache entspricht, dass sie sich nur durch eine Drehung voneinander unterscheiden. Für $\mathbf{z}_j$ und $\mathbf{L}_j$ trifft dies nicht zu.

Die Matrix $T$ ergibt sich aus den Annahmen **A1** und **A2** wie folgt. Aus dem Ansatz (1.3) ergibt sich wegen der postulierten Orthonormalität von $T$ durch Multiplikation von rechts mit $T$ die Beziehung

$$ZT = L, \tag{1.8}$$

d.h. die latenten Vektoren $\mathbf{L}_k$ ergeben sich als Linearkombinationen der Spaltenvektoren von $Z$. **A2** liefert

$$T'Z'ZT = L'L = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n), \quad \lambda_k = \|\mathbf{L}_k\|^2, \ 1 \leq k \leq n \tag{1.9}$$

Nach Multiplikation von links mit $T$ ergibt sich

$$Z'ZT = T\Lambda, \quad Z'Z\mathbf{t}_k = \lambda_k \mathbf{t}_k. \tag{1.10}$$

$\mathbf{t}_k$ ist offenbar ein Eigenvektor von $Z'Z$, und $\lambda_k$ ist der zugehörige Eigenvektor. $T$ und $\Lambda$ sind die Matrizen der Eigenvektoren und der entsprechenden Eigenwerte. Die Matrix $T$ kann numerisch bestimmt werden, worauf hier nicht weiter eingegangen werden muß, und wegen (1.8) ist dann auch $L$ gegeben.

Schreibt man die Gleichung (1.9) für einen speziellen Vektor $\mathbf{t}_k$ von $T$, so erhält man

$$\mathbf{t}_k'(Z'Z)\mathbf{t}_k = \lambda_k, \quad Z'Z = T\Lambda T' \tag{1.11}$$

Dies ist ein spezieller Ausdruck für die quadratische Form

$$\mathbf{t}'(Z'Z)\mathbf{t} = \lambda = \text{eine Konstante.} \tag{1.12}$$

Für einen festen Wert von $\lambda$ beschreibt (1.12) ein Ellipsoid $\mathcal{E}_t = \{\mathbf{t}|\mathbf{t}'(Z'Z)\mathbf{t} = \lambda = \text{konstant}\}$. $Q(\mathbf{t}) = \mathbf{t}'(Z'Z)\mathbf{t}$ werde als Funktion des Vektors $\mathbf{t}$ interpretiert. Unter der Nebenbedingung $\|\mathbf{t}\| = 1$ nimmt $Q(\mathbf{t})$ einen maximalen Wert an: $Q(\mathbf{t}_1) = \lambda_1$, $\lambda_1$ der maximale Eigenwert und $\mathbf{t}_1$ der zugehörige Eigenvektor von $Z'Z$ (Satz von Courant-Fischer). Da $\lambda_1 = \mathbf{L}_1'\mathbf{L}_1 = \|\mathbf{L}_1\|^2$ bedeutet dies, dass $\mathbf{t}_1$ die Orientierung mit der maximalen Orientierung der Konfiguration der Fälle angibt; analoge Aussagen folgen für $\mathbf{t}_2$ etc. Die Annahmen **A1** und**A2** liefern also tatsächlich ein Koordinatensystem $L_1, \ldots, L_n$, in dem die Punktekonfiguration nach Maßgabe ihrer maximalen Ausdehungen beschrieben wird.

Aus $Z' = TL'$ folgt $\tilde{\mathbf{z}}_i = T\tilde{\mathbf{L}}_i$ bzw. $T'\tilde{\mathbf{z}}_i = \tilde{\mathbf{L}}_i$. Aus $ZT = L$ folgt insbesondere $\tilde{\mathbf{z}}_i'T = \tilde{\mathbf{L}}_i'$ und nach Transposition $\tilde{\mathbf{L}}_i = T'\tilde{\mathbf{z}}_i$. Multiplikation von rechts mit $\Lambda\tilde{\mathbf{L}}_i$ liefert

$$\tilde{\mathbf{z}}_i'T\Lambda\tilde{\mathbf{L}}_i = \tilde{\mathbf{L}}_i'\Lambda\tilde{\mathbf{L}}_i$$

Aber wie gerade gezeigt ist $\tilde{\mathbf{L}}_i = T'\tilde{\mathbf{z}}_i$, so dass wegen (1.11) $\tilde{\mathbf{z}}_i'T\Lambda T'\tilde{\mathbf{z}}_i = \tilde{\mathbf{z}}_i'Z'Z\tilde{\mathbf{z}}_i$ schließlich

$$\tilde{\mathbf{z}}_i'Z'Z\tilde{\mathbf{z}}_i = \tilde{\mathbf{L}}_i'\Lambda\tilde{\mathbf{L}}_i = k_{0i}, \quad i = 1, \ldots, m \tag{1.13}$$

folgt. Dies bedeutet, dass einerseits jeder Fall, repräsentiert durch den Vektor $\tilde{\mathbf{z}}_i$ auf einem durch $\mathcal{E}_{iz} = \{\mathbf{z}|\mathbf{z}'Z'Z\mathbf{z} = k_{0i}\}$ definierten Ellipsoid liegt, und andererseits auf einem in den latenten Koordinaten $L_k$ definierten, achsenparallelen Ellipsoid $\mathcal{E}_{iL} = \{\mathbf{y}'\Lambda\mathbf{y} = k_{0i}, \mathbf{y} \in \mathbb{R}^n\}$. Die Abbildung 3 illustriert diesen Sachverhalt. Die Ellipsoide sind durch die Matrizen $Z'Z$ bzw. $\Lambda$ definiert und ihre Existenz setzt *nicht* voraus, dass die Punktekonfigurationen ellipsoid sind. Dies wäre der Fall, wären die Daten multivariat normalverteilt. Es ist durchaus möglich, dass die Stichprobe der $m$ Fälle aus Teilstichproben aus verschiedenen Subpopulatiionen zusammengesetzt ist und die entsprechenden Teilkonfigurationen unterschiedliche Orientierungen im $(X_1, \ldots, X_n)$-Koordinatensystem haben. In den Beispielen in Abschnitt 3 wird diese Möglichkeit illustriert. Wenn die Kovarianzen und Korrelationen zwischen den Variablen berechnet werden, indem über alle Fälle in der Gesamtstichprobe gemittelt wird, so beziehen sich die Ellipsoide stets auf die Gesamtstichprobe, weil sie eben durch $X'X$, $Z'Z$ oder $\Lambda$ definiert sind.

Die folgenden Betrachtungen führen zu einer Repräsentation der Variablen durch die latenten Variablen.

Abbildung 1: Konfiguration von Fällen im ursprünglichen Koordinatensystem: Gewicht versus Körpergröße (a). In (b) sind mögliche latente Variable eingezeichnet worden: $L_1$ hat die Orientierung der maximalen Ausdehnung der Konfiguration, $L_2$ ist orthogonal zu $L_1$ und repräsentiert die Orientierung mit im allgemeinen zweitgrößter Ausdehnung der Konfiguration. (c) zeigt die Konfiguration im Koordinatensystem $(L_1, L_2)$; die Koordinaten in diesem System sind die Projektionen der Punkte im ursprünglichen System auf die Achsen $L_1$ und $L_2$. Die Punkte werden durch Zahlen repräsentiert, um die Identifikation der Punkte im rotierten System zu erleichtern. Der Ausreisser 22 wurde bei der Bestimmung von $L_1$ und $L_2$ *nicht* berücksichtigt, weil er wegen seiner *Hebelwirkung* (leverage) die optimale Bestimmung dieser Achsen verhindert hätte.

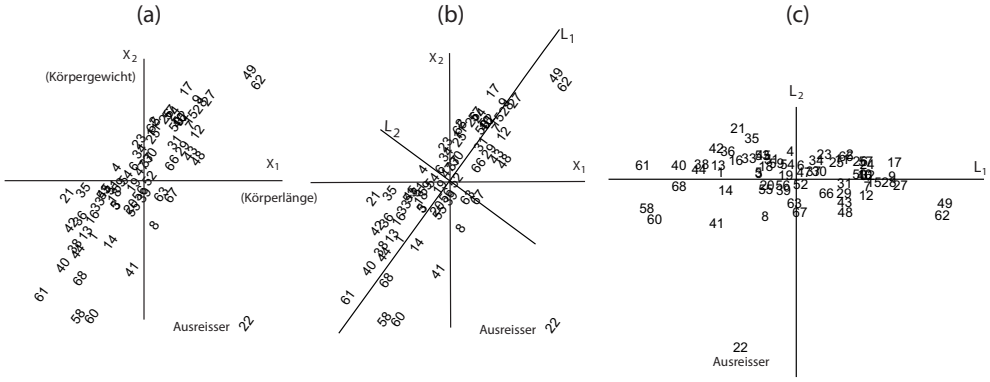
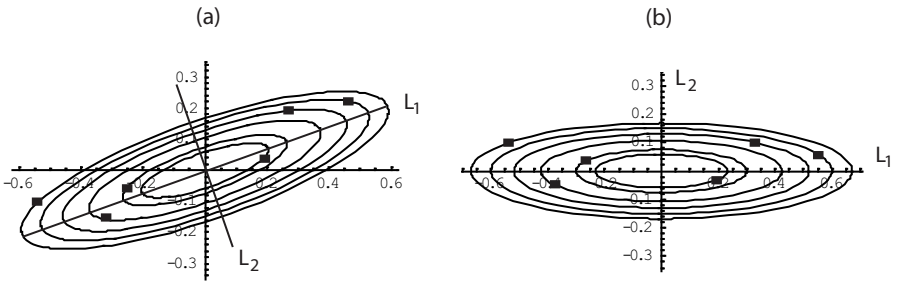



Abbildung 2: Punktekonfiguration und Ellipsen.

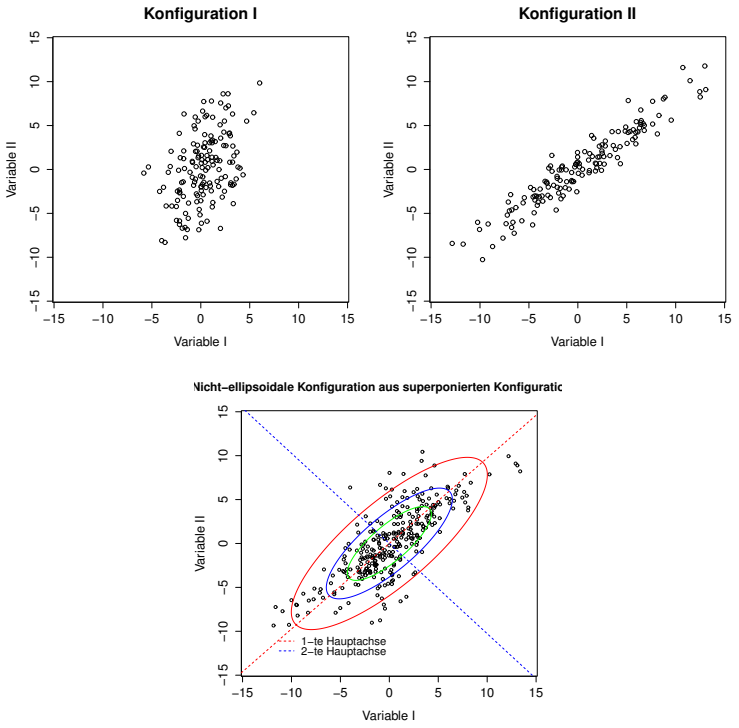



**Singularwertzerlegung** Die Beziehung $Z = LT'$ führt auf die zentrale Gleichung der PCA. Dazu werde $L$ normalisiert. Dazu werden die $\mathbf{L}_L$ mit dem Faktor $1/\|\mathbf{L}_k\|$ multipliziert. Aber $\|\mathbf{L}_k\| = \sqrt{\lambda_k} = \lambda^{1/2}$ Bezeichnet man mit $\Lambda^{-1/2}$ die Diagonalmatrix $\text{diag}(\lambda^{-1/2}, \ldots, \lambda_n^{-1/2})$, so ist

$$Q = L\Lambda^{-1/2} \tag{1.14}$$

die Matrix, deren Spalten die normierten Vektoren $\mathbf{q}_k = \lambda_k^{-1/2}\mathbf{L}_k$ sind. So kommt

Abbildung 3: Superponierte Punktekonfigurationen und Ellipsen




man zur *Singularwertzerlegung* oder SVD[5]

$$Z = Q\Lambda^{1/2}T'. \tag{1.15}$$

So, wie sich die Matrix $T$ als Matrix der Eigenvektoren von $Z'Z$ erwiesen hat, ergibt sich $Q$ als Matrix der Eigenvektoren von $ZZ'$: es ist

$$ZZ' = Q\Lambda^{1/2}T'T\Lambda^{1/2}Q' = Q\Lambda Q'.$$

In $Q$ stehen nicht alle Eigenvektoren von $ZZ'$, sondern nur die, die zu Eigenwerten ungleich Null korrespondieren; die von Null verschiedenen Eigenwerte von $Z'Z$ und $ZZ'$ sind identisch. Die Diagonalelemente $\lambda_k^{1/2}$ von $\Lambda^{1/2}$ heißen auch *Singularwerte*.

[5]= Singular Value Decomposition

# 2 PCA

## 2.1 SVD und PCA

Die Beziehung (1.15) bietet zwei Möglichkeiten, $Z$ zu beschreiben:

$$\begin{aligned} Z &= LT', \quad L = Q\Lambda^{1/2} && (2.1)\\ &= QA', \quad A = T\Lambda^{1/2} && (2.2) \end{aligned}$$

Geht man von der Zerlegung $Z = LT'$ aus, so fokussiert man auf die Struktur der Fälle, da die $\mathbf{L}_k$ der Forderung A3 genügen. Betrachtet man die Zerlegung $Z = QA'$, so fokussiert man auf die Struktur der Variablen, wie im Folgenden spezifiert wird.

Es seien $\mathbf{L}_k$ und $\mathbf{a}_k$ die $k$-ten Spaltenvektoren von $L$ bzw. $A$:

$$\begin{aligned} \mathbf{L}_k &= (\ell_{1k}, \ell_{2k}, \ldots, \ell_{mk})' && (2.3)\\ \mathbf{a}_k &= (a_{1k}, a_{2k}, \ldots, a_{nk})', \quad k = 1, \ldots, n && (2.4) \end{aligned}$$

Die $\ell_{ik}$, $i = 1, \ldots, m$ heißen *Faktorwerte* (Faktor Scores) der Fälle, und die $a_{jk}$ heißen *Faktorladungen* der gemessenen Variablen.

**Faktorwerte:** Es sei $\tilde{\mathbf{z}}_i$ der $i$-e $i$-te Spaltenvektor von $Z'$, (d.h. der $i$-te Zeilenvektor von $Z$); die Komponenten von $\tilde{\mathbf{z}}_i$ sind die (spalten-)standardisierten Messwerte des $i$-ten Falles für die Variablen, und wegen $ZT = L$ folgt[6]

$$\ell_{ik} = \tilde{\mathbf{z}}_i' \mathbf{t}_k = \sum_{j=1}^{n} z_{ij} t_{jk} = \|\tilde{\mathbf{z}}_i\| \|\mathbf{t}_k\| \cos\theta_{ik}. \tag{2.5}$$

Der Faktorwert $\ell_{ik}$ repräsentiert dem $i$-ten Fall auf der $k$-ten latenten Variablen oder Dimension, etwa die Ausprägung des $k$-ten latenten Merkmals beim $i$-ten Fall. Nach (2.5) ist $\ell_{ik}$ das Skalarprodukt des Vektors $\tilde{\mathbf{x}}_i$ und des Vektors $\mathbf{t}_k$, also der Messwerte des $i$-ten Falles in den Variablen und der Repräsentation der Variablen auf der $k$-ten latenten Dimension. $\ell_{ik}$ ist maximal wenn $\tilde{\mathbf{x}}_i$ und $\mathbf{t}_k$ parallel sind; dann ist der Winkel $\theta_{ik} = 0$, die Komponenten von $\tilde{\mathbf{x}}_i$ und $\mathbf{t}_k$ unterscheiden sich nur durch einen gemeinsamen Proportionalitätsfaktor und $\tilde{\mathbf{x}}_i$ liegt auf der $k$-ten latenten Achse, und $\ell_{ik} = 0$ wenn $\tilde{\mathbf{x}}_i$ und $\mathbf{t}_k$ orthogonal sind, wenn also das Profil der Messwerte des $i$-ten Falls mit dem Profil der Repräsentationen der Variablen auf der $k$-ten latenten Dimension gewissermaßen nicht korreliert.

**Faktorladungen:** Aus (2.2) folgt $A = Z'Q$, so dass das Element $a_{jk}$ von $A$ durch

$$a_{jk} = \mathbf{z}_j' \mathbf{q}_k = \sum_{i=1}^{m} z_{ij} q_{ik} = \|\mathbf{z}_j\| \|\mathbf{q}_k\| \cos\varphi_{jk} \tag{2.6}$$

Die Komponenten von $\mathbf{z}_j$ sind die standardisierten Messwerte der Fälle für die $j$-te Variable, und die Komponenten von $\mathbf{q}_k$ sind die Repräsentationen der Fälle

[6] s. http://www.uwe-mortensen.de/VektorenMatrizen2020.pdf, p. 19

auf der $k$-ten latenten Variablen. Man kann die Ladung $a_{jk}$ als Korrelation (bis auf den Faktor $1/m$) zwischen den Messwerten der Fälle für die $j$-te Variable und den Repräsentationen der Fälle auf der $k$-ten latenten Dimension sehen. $a_{jk}$ ist maximal, wenn der Winkel $\varphi_{jk}$ zwischen diesen beiden Vektoren gleich Null ist; dann unterscheiden sich die Komponenten den Vektoren $\mathbf{z}_j$ und $\mathbf{q}_k$ nur durch einen Proportionalitätsfaktor und der Vektor $\mathbf{z}_j$ liegt auf der $k$-ten latenten Achse, und $a_{jk} = 0$, wenn $\mathbf{z}_j$ und $\mathbf{q}_k$ orthogonal sind.

**Messwerte:** Die Messwerte $z_{ij}$ ($i$-ter Fall, $j$-ter Test) sind Skalarprodukte von Vektoren, deren Komponenten durch Werte der latenten Vareiablen definiert sind. Es gilt

$$z_{ij} = \mathbf{q}_i'\tilde{\mathbf{a}}_j = \|\tilde{\mathbf{q}}_i\|\|\tilde{\mathbf{a}}_j\|\cos\phi_{ij} \tag{2.7}$$

$$= \tilde{\mathbf{L}}_i'\tilde{\mathbf{t}}_j = \|\tilde{\mathbf{L}}_i\|\|\tilde{\mathbf{t}}_j\|\cos\phi_{ij} \tag{2.8}$$

$\tilde{\mathbf{q}}_i$ und $\tilde{\mathbf{L}}_i$ sind die $i$-ten Zeilenvektoren von $Q$ bzw. $L$; sie repräsentieren den $i$-ten Fall auf den latenten Dimensionen, und $\tilde{\mathbf{t}}_j$ und $\tilde{\mathbf{a}}_j$ sind die $j$-ten Zeilenvektoren von $T$ bzw. $A$, sie repräsentieren die $j$-te Variable auf den latenten Dimensionen. Die latenten Dimensionen oder Variablen beschreiben also Fälle und Variablen gleichermaßen. $\cos\phi_{ij}$ ist ein Ähnlichkeitsmaß für den $i$-ten Fall und den $j$-ten Test: für $\phi_{ij} = 0$ wird $z_{ij}$ maximal relativ zu den Längen der Vektoren $\tilde{\mathbf{q}}_i$ und $\tilde{\mathbf{a}}_j$, etc. Der Vektor $\tilde{\mathbf{a}}_j$ (oder $\tilde{\mathbf{t}}_j$) definiert eine bestimmte Orientierung und damit eine bestimmte Gerade im Raum der latenten Variablen, auf dem die gemessene $j$-te Variable und $\tilde{\mathbf{q}}_i$ gleichermaßen liegen. Dieser Befund liegt nahe, die Vektoren $\tilde{\mathbf{q}}_i$ oder $\tilde{\mathbf{L}}_i$ einerseits und die $\tilde{\mathbf{t}}_j$ bzw. $\tilde{\mathbf{a}}_j$ andererseits simultan in einer Graphik darzustellen. Das Resultat ist ein *Biplot*, auf den weiter unten zurückgekommen wird.

**Varianzen der Faktorwerte:** Da $L'L = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ folgt

$$\mathbf{L}_k'\mathbf{L}_k = \|\mathbf{L}_k\|^2 = \lambda_k, \quad k = 1, \ldots, n \tag{2.9}$$

Die Standardisierung der Datenmatrix impliziert, dass die Spaltensummen von $Z$ und damit die von $L$ stets gleich Null sind[7]. Deswegen kann man $\frac{1}{m}\|\mathbf{L}_k\|^2 = \frac{1}{m}\lambda_k$ als Varianz der Abbildungen der Fälle auf die $k$-te latente Dimension betrachten.

**Ladungen und Korrelationen:** Es ist $\frac{1}{m}Z'Z = R$ die Matrix der Korrelationen zwischen den Variablen. Aus $Z = QA'$ folgt dann

$$\frac{1}{m}Z'Z = R = \frac{1}{m}AQ'QA' = \frac{1}{m}AA'. \tag{2.10}$$

Das Element in der $u$-ten Zeile und $v$-ten Spalte von $AA'$ ist gleich dem Skalarprodukt der $u$-ten und der $v$-ten Zeile von $A$, d.h. von $\tilde{\mathbf{a}}_u$ und $\tilde{\mathbf{a}}_v$ (die Zeilen von $A$ repräsentieren die gemessenen Variablen, die Spalten von $A$ repräsentieren die latenten Variablen). Wegen $A = T\Lambda^{1/2}$ hat man $a_{uk} = t_{uk}\sqrt{\lambda_k}$ und $a_{vk} = t_{vk}\sqrt{\lambda_k}$, so dass

$$r_{uv} = \frac{1}{m}\tilde{\mathbf{a}}_u'\tilde{\mathbf{a}}_v = \frac{1}{m}\sum_{k=1}^{n}\lambda_k t_{uk}t_{vk} = \frac{1}{m}\|\tilde{\mathbf{a}}_u\|\|\tilde{\mathbf{a}}_v\|\cos\alpha_{uv} \tag{2.11}$$

[7] V& M, Abschn.

$\alpha_{uv}$ der Winkel zwischen den Vektoren $\tilde{\mathbf{a}}_u$ und $\tilde{\mathbf{a}}_v$. Insbesondere erhält man für den Fall $u = v$ die Beziehung

$$r_{uu} = \frac{1}{m}\|\tilde{\mathbf{a}}_u\|^2 = 1 \tag{2.12}$$

so dass (2.11) impliziert, dass

$$r_{uv} = \cos \alpha_{uv} \tag{2.13}$$

**Geometrische Implikationen und Anzahl der latenten Variablen:** Der Befund (2.12) bedeutet, dass die Variablen stets durch Punkte (Endpunkte von Vektoren) auf einer Hyperkugel mit dem Radius 1 repräsentiert werden. Werden die Daten durch nur zwei latente Variable definiert, so ist die Hyperkugel ein Kreis, im Falle von drei latenten Variablen ist die Hyperkugel eben eine Kugel. Besonders für niedrigdimensionale Lösungen hat man damit einen raschen Test zur Hand: liegen die die Variablen repräsentierenden Punkte auf einem Kreis, so kann man von einer 2-dimensionalen Lösung ausgehen, liegen dagegen einige der Variablen deutlich innerhalb des Kreises, so wird man eine höherdimensionale Lösung akzeptieren müssen.

Es werde noch das Kreuzprodukt $A'A$ betrachtet. Dieses Produkt ist gleich der Matrix der Skalarprodukte der Spaltenvektoren von $A$, und da die Spaltenvektoren von $A$ orthogonal sind, hat man

$$\mathbf{a}'_k\mathbf{a}_{k'} = \begin{cases} 0, & k \neq k' \\ \|\mathbf{a}_k\|^2 = \sum_{j=1}^{n} \lambda_k t_{jk}^2 = \lambda_k, & k = k' \end{cases} \tag{2.14}$$

Fasst man dieses Ergbnis mit (2.9) zusammen, so hat man

$$\|\mathbf{L}_k\|^2 = \|\mathbf{a}_k\|^2 = \lambda_k. \tag{2.15}$$

Da die $\mathbf{L}_k$ zentriert sind, ist $\lambda_k$ proportional zur Varianz der Faktorwerte der Fälle auf der $k$-ten latenten Dimension. Dass $\lambda_k$ auch gleich der Quadratsumme der Ladungen der Variablen auf $L_k$ ist, bedeutet aber nicht notwendig, dass $\lambda_k$ auch proportional zur Varianz der Ladungen ist, da die Ladungen nicht zentriert sind.

## 2.2 Interpretationshilfen

### 2.2.1 Beiträge einer latenten Variablen

Es kann hilfreich sein, den Beitrag des $i$-ten Falles zu einer latenten Dimension zu bestimmen. In ausgeschriebener Form gilt nach (2.15)

$$\|\mathbf{L}_k\|^2 = \sum_{i=1}^{m} \ell_{ik}^2 = \lambda_k.$$

Der Beitrag des $i$-ten Falles zur Definition der $k$-ten latenten Dimension wird dann definiert durch

$$B_{ik} = \frac{\ell_{ik}^2}{\lambda_k}. \tag{2.16}$$

Offenbar gilt $0 \leq B_{ik} \leq 1$ und $\sum_i B_{ik} = 1$. Je größer $B_{ik}$, desto mehr wird die $k$-te Dimension durch den $i$-ten Fall bestimmt. Die Fälle mit hohen $B_{ik}$-Werten und hohen negativen $B_{ik}$-Werten definieren dann die Endpunkte der $k$-ten Dimension und können dabei helfen, eine inhaltliche Bedeutung dieser Dimension zu finden.

Ein weiteres Maß ist der *quadrierte Kosinus* des Winkels zwischen dem Vektor, der einen Fall repräsentiert, und der $k$-ten latenten Dimension. Diese Größe definiert die Bedeutung der $k$-ten latenten Dimension für den $i$-ten Fall. Der $i$-te Fall wird durch den Vektor $\tilde{\mathbf{L}}_i$ repräsentiert. $\ell_{ik}$ ist die Abbildung dieses Vektors auf die $k$-te latente Dimension, und der Winkel zwischen $\tilde{\mathbf{L}}_i$ und der $k$-ten Dimension sei $\theta_{ik}$. Dann ist

$$\cos^2 \theta_{ik} = \frac{\ell_{ik}^2}{\|\tilde{\mathbf{L}}_i\|^2}. \tag{2.17}$$

Für $\theta_{ik} = 0$ ist $\cos^2 \theta_{ik} = 1$ und damit $\ell_{ik}^2 = \|\tilde{\mathbf{L}}_k\|^2$; man kann sagen, dass $\tilde{\mathbf{L}}_k$ durch den $i$-ten Fall definiert wird (oder umgekehrt, dass der $i$-te Fall durch die $k$-te latente Dimension bestimmt wird). Für $\theta_{ik} = \pi/2$ ist $\cos\theta_{ik} = 0$ und $\tilde{\mathbf{L}}_i$ steht orthogonal zur $k$-ten latenten Achse, d.h. $\ell_{ik} = 0$, so dass der $i$-te Fall nichts zur Charakterisierung der $k$-ten latenten Dimension beiträgt, und umgekehrt diese nichts zur Charakterisierung des $i$-ten Falles beiträgt.

Abbildung!

### 2.2.2 Abschätzung der Anzahl latenter Dimensionen

Wie schon angemerkt wurde haben Datenmatrizen im Allgemeinen vollen Rang, d.h.numerisch gilt $\mathrm{rg}(Z) = \min(m, n)$. Oft kann man aber vermuten, dass es nur $r < \min(m, n)$ bedeutsame latente Variable gibt und $\min(m, n) - r$ der berechneten latenten Variablen nur "Rauschen", also zufällige Effekte abbilden.

Im Extremfall ist $r = 1$; dann werden die Kovarianzen zwischen den Variablen durch nur eine latente Variable erklärt. Sowohl Fälle wie Variable werden mit nur vernachlässigbaren Abweichungen auf einer Geraden repäsentiert. Müssen zwei latente Variablen angenommen werden, so liegen wegen (2.12) alle Variablen mit nur vernachlässigbaren Abweichungen auf einem Kreis. Die Abweichungen liegen stets innerhalb des Kreises, nie außerhhalb. Stärkere Abweichungen ins Innere des Kreises legen dann nahe, dass zumindest für einige Variable eine oder mehrere latente Variable eine Rolle spielen könnten.

Generell kann man vermuten, dass sich "bedeutsame" latente Variable dadurch auszeichnen, dass sie mehr zwischen Fällen und Variablen unterscheiden als zufällige Effekte. Gleichung (2.15) legt dann nahe, als Maß für die Bedeutsamkeit die Eigenwerte $\lambda_k$ zu wählen: differenziert eine latente Variable Variable relativ stark zwischen den Fällen, so wird $\lambda_k$ entsprechend groß sein. Die Vermutung bzw. Hofffnung ist dann, dass zwischen der Größe der Eigenwerte für "bedeutende" latente Dimensionen und der für "zufällige" latente Variable ein deutlicher Unnterschied besteht. Diese Betrachtung der Eigenwerte ist der Scree-Test.

Abbildung scree-Test

Varianten des Scree-Tests:

### 2.2.3 Statistische Inferenz

Es gibt grundsätzlich zwei Modelle, in Bezug auf die die Resultate einer PCA interpretiert werden können; es sind die aus der ANOVA bekannten Modelle: (i) das *'"Fixed Effect Model"* und (ii) das *"Random Efffect Model"*. Im Fixed Model wird die Stichprobe als die Population von Messungen betrachtet, die von Interesse ist, während beim Random Model die Daten eben als Stichprobe aus einer größeren Stichprobe betrachtet werden, auf die verallgemeinernd geschlossen werden soll. Insbesondere sollen neue Messungen im Rahmen der Resultate für die gegebenen Stichprobe diskutiert werden.

**Das Fixed Effect Model:** Die Matrix $X$ kann auf der Basis der SVD über die dyadischen Produkte der Spaltenvektoren von $Q$ und $T$ ausgedrückt werden, denn die rechte Seite von $X = Q\Lambda^{1/2}T'$ ist äquivalent zu

$$X = \sigma_1 \mathbf{q}_1 \mathbf{t}_1' + \sigma_2 \mathbf{q}_2 \mathbf{t}_2' + \cdots + \sigma_n \mathbf{q}_n \mathbf{t}_n' = \sum_{k=1}^{n} \sigma_k \mathbf{q}_k \mathbf{t}_k' \quad \sigma_k = \sqrt{\lambda_k}. \tag{2.18}$$

(2.18) kann benutzt werden, um den Wert des Ranges $r$ von $X$ abzuschätzen: Terme mit "hinreichend" kleinen $\lambda_k$-Werten können u.U. vernachlässigt werden. Man hat dazu den

**Satz 2.1** *(Satz von Eckart & Young) Die Approximation*

$$X \approx X_r = Q_r \Lambda_r^{1/2} T_r' = \sum_{k=1}^{r} \sqrt{\lambda_k} \boldsymbol{q}_k \boldsymbol{t}_k', \quad r < n \tag{2.19}$$

*approximiert $X$ im Sinne der Methode der Kleinsten Quadrate.*

**Beweis:** Bekannt wurde diese Aussage (samt Beweis) durch die Arbeit von Eckart & Young (1936); eine modernere Version des Beweises wird in http://www.uwe-mortensen.de/VektorenMatrizen2020.pdf, Abschnitt 2.8.5 angeboten. □

Für $r < n$ wird im Allgemeinen $\hat{X}_r \neq X$ gelten. Dazu sei $E$ eine Fehlermatrix derart, dass

$$X = \hat{X}_r + E, \tag{2.20}$$

d.h. $E = X - \hat{X}_r$. Man betrachtet dann die *Residual Sum of Squares* (RESS):

$$\text{RESS} = \|E\|^2 = \|X - \hat{X}_r\|^2 = \text{spur}(E'E) = I_n - \sum_{k=1}^{r} \lambda_k \tag{2.21}$$

(Vergl. V & M, Abschnitt 2.7.5)

**Das Random Effect Model:** Die Frage ist, ob die Ergebnisse für die Stichprobe auf eine Population generalisiert werden können. Eine M Möglichkeit, dies zu tun, besteht in der Annahme einer multivariaten Verteilung für für Daten; üblicherweise wird die multivariate Gauß-Verteilung angenommen. Diese Annahme muß aber keinesfalls gerechtfertigt sein. Ein alternativer Ansatz besteht in der Anwendung der *jack-knife*-Technik. Dazu wird, einer nach dem anderen, ein Fall aus den Daten gestrichen und die Analyse für die restlichen Fälle durchgeführt. Anschließend wird der herausgenommene Fall auf der Basis der Analyse vorhergesagt. Auf diese Weise werden die Daten eines jeden Falls auf der Basis der jeweils übrigen vorausgesagt. Die vorhergesagten Werte werden dann in einer Matrix $\hat{X}$ zusammengefasst.

Die Qualität des PCA Random-Effekt-Modells wird dann durch den Vergleich von $\hat{X}$ mit den Matrizen $\hat{X}_r$ (vergl. (2.19)) bewertet. Die kritische Grösse ist die *Predicted Residual Sum of Squares* (PRESS). Man hat

$$\text{PRESS} = \|X - \hat{X}_r\|^2. \tag{2.22}$$

Die Qualität der PCA-Lösung ist um so besser, je kleiner die PRESS-Größe ist.

### 2.2.4 Rotationen

Bei der Faktorenanalyse können latente Achsen ("Dimensionen") rotiert werden, um neue Achsen zu bestimmen, die besser interpretiert werden können. Eine PCA-Lösung dagegen könne nur rotiert werden, wenn auf die Unkorreliertheit der rotierten Achsen verzichtet wird. Es zeigt sich aber, dass etwa im Variablenraum Rotationen zu interessanten Lösungen führen können, wie das weiter unten betrachtete Beispiel zeigt.

Es sei $Z$ eine spaltenstandardisierte Datenmatrix, und

$$\frac{1}{\sqrt{m}} Z = Q \Lambda^{1/2} T'$$

sei die SVD von $Z$. Es sei $A = T\Lambda^{1/2}$ die Matrix der Ladungen für die Variablen. Die Spaltenvektoren $\mathbf{t}_k$ von $T$ repräsentieren die Orientierungen, also der Hauptachsen des durch die Matrix $\frac{1}{m} Z'Z = R$ definierten Ellipsoids, und die Orthogonalität der $\mathbf{t}_k$ bedeutet die Unkorreliertheit der durch diese Achsen repräsentierten latenten Variablen. Zur Erinnerung: $A$ ist eine $(n, n)$-Matrix, deren *Zeilen* die Variablen, die *Spalten* die Dimensionen repräsentieren. Man möchte nun die Variablenvektoren rotieren. Dazu muß man eine geeignete Rotationsmatrix $S$ derart, dass $\tilde{A}' = SA'$. Damit die Darstellung von $Z$, $Z = Q\Lambda^{1/2}T'$, nach wie vor gilt, muß auch $Q'$ analog transformiert werden; $W' = S'Q'$: jeder Vektor $\tilde{\mathbf{q}}_i$ wird in einen Vektor $\tilde{\mathbf{x}}_i$ rotiert, $\tilde{\mathbf{w}}_i = S'\tilde{\mathbf{q}}_i$, $i = 1, \ldots, m$, $W = QS$. Dies bedeutet

$$\frac{1}{\sqrt{m}} Z = QSS'A' = QS\tilde{A}' = W\tilde{A}', \quad W = QS \tag{2.23}$$

Es ist

$$W'W = S'Q'QS = S'S = I, \tag{2.24}$$

d.h. die Spaltenvektoren von $W$ sind ebenfalls orthonormal. Für die Matrix $\tilde{A}$ der rotierten Achsen ist

$$\tilde{A} = AS'. \tag{2.25}$$

Nach Gleichung (2.10), Seite 9, gilt $Z'Z = AA'$. Für $\tilde{A}$ erhält man

$$\tilde{A}\tilde{A}' = AS'SA' = AA' = Z'Z, \tag{2.26}$$

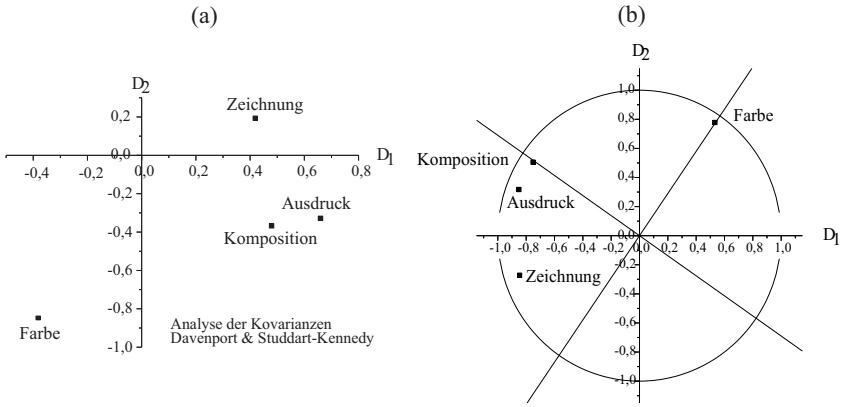
d.h. es gilt wieder $\tilde{\mathbf{a}}'_j\tilde{\mathbf{a}}_k = \mathbf{z}'_j\mathbf{z}_k = r_{jk}$. Während aber $A'A = \Lambda^{1/2}V'V\Lambda^{1/2} = \Lambda$ ist – d.h. die Spaltenvektoren von $A$ sind orthogonal – findet man

$$\tilde{A}'\tilde{A} = SA'AS = S\Lambda S', \tag{2.27}$$

d.h. die Spaltenvektoren von $\tilde{A}$ sind *nicht* mehr orthogonal.

Davenport und Studdert-Kennedy (1972)[8] analysierten die ästhetischen Urteile des Kunstkritikers Roger de Pile über 56 Maler, von Albani, Dürer, Veronese, Holbein, Rembrandt, Rubens, Titian bis Van Dyck, Vanius und den Zuccaros, die de Pile Jahr 1743 notierte; diese Ratings liefern möglicherweise auch Informationen über die Kunstrezeption in der Mitte des 18-ten Jahrhunderts. Monsieur de Pile "ratete" die Maler in bezug auf vier Merkmale: "Komposition", "Zeichnung", "Farbe" und "Ausdruck", d.h. er schätzte die Maler bezüglich dieser Merkmale auf einer Skala von 0 bis 20 ein; die Ratingskala ist also keine Erfindung neuzeitlicher Psychologen[9].

Abbildung 4: Faktorladungen für de Piles Merkmale von Gemälden: (a) von Kovarianzen, (b) von standardisierten Werten




[8]Davenport, M., Studdert-Kennedy, H. (1970) Use of orthogonal factors for selection of variables in a regression equation. Appl. Statist. **21**, 324-333. Dem Titel entsprechend diskutieren die Autoren die Anwendung der Hauptachsentransformation (Principal Component Analysis - PCA) im Rahmen eines Regressionsproblems. Es sollen optimale Prädiktoren für die Ratings gefunden werden.

[9]Eine ausführliche Diskussion dieser Daten findet man unter http://www.uwe-mortensen.de/fakanalysews0506b.pdf.

Tabelle 1: Merkmalskorrelationen und Faktorladungen

| Korrelationen zwischen den Merkmalen | | | | |
|---|---|---|---|---|
| | Kompos. | Zeichn. | Farbe | Ausdruck |
| Kompos. | 1.00 | .415 | -.097 | .656 |
| Zeichn. | .415 | 1.00 | -.517 | .575 |
| Farbe | -.097 | -.517 | 1.00 | -.209 |
| Ausdruck | .656 | .575 | -.208 | 1.00 |
| de Piles Ästhetik: Ladungen bezüglich der Hauptachsen | | | | |
| Merkmal | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_4$ |
| Kompos. | .48 | -.37 | .78 | .10 |
| Zeichn. | .42 | .19 | -.28 | .84 |
| Farbe | -.38 | -.85 | -.21 | .31 |
| Ausdruck | .66 | -.33 | -.31 | -.43 |
| kum. Varianz | 55.95 | 84.48 | 93.59 | 100.00 |

Die Korrelation zwischen den Merkmalen 'Komposition' und 'Farbe' beträgt nach Tabelle 1 $r = .097$, – man kann $r = .00$ annehmen, d.h. die Beurteilungen Maler bezüglich dieser beiden Merkmnale sind unkorreliert. Es liegt dann nahe, diese beiden Merkmale als neue Bezugsmerkmale, also als "latente" Variable zu wählen. Die übrigen Merkmale lassen sich als Linearkombination dieser beiden neuen Variablen darstellen. Dies bedeutet eine Rotation der ursprünglichen latenten Achsen um einen bestimmten Winkel. Die Projektionen der Variablen auf diese neuen Achsen werden nicht mehr unkorreliert sein.

Analoge Betrachtungen gelten, wenn allgemeine Rotationsstrategeien wie die Varimax-Rotation auf die Variablen angewendet werden.

# 3 Beispiele

## 3.1 R.A. Fishers Iris-Daten

R. S. Fisher publizierte 1936 eine Methode zur Klassifikation von Objekten, die *Diskriminanzanalye*, die er an einem mittlerweile berühmten Datensatz aus der Botanik illustrierte: es sind Messungen an der Pflanze *Iris*. Tabelle 2 zeigt zur Illustration einen Ausschnitt aus diesem berühmten Datensatz. Es gibt vier Variablen: Die Kelchblatt (sepal)- sowie die Blütenblatt (petal)-Länge sowie die entprechenden Breiten in cm, und drei Kategorien (Arten: setosa, versicolor und virginica). Für jede dieser Arten gibt es fünzig Fälle, so dass die Tabelle insgesamt 150 Fälle enthält. Die Daten sollen hier einer PCA unterzogen werden. Da die erste Dimension so gewählt wird, dass die Varianz der Abbildungen der Fälle auf die korrespondierende Achse maximal ist, kann untersucht werden, ob diese Erste Achse auch eine Differenzierung zwischen den Irisarten liefert.

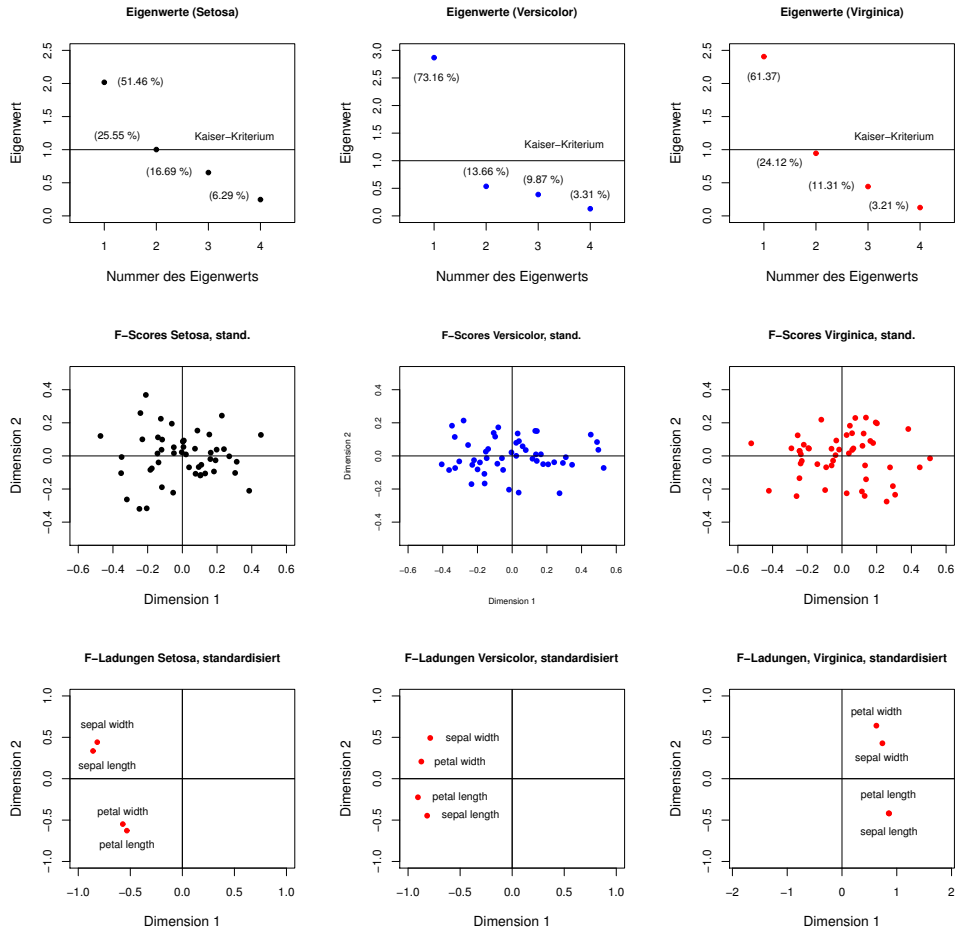
Tabelle 2: Fishers Irisdaten

| | Sepal Length | Sepal Width | Petal Length | Petal Width | Species |
|---|---|---|---|---|---|
| 1 | 5.1 | 3.5 | 1.4 | 0.2 | setosa |
| 2 | 4.9 | 3.0 | 1.4 | 0.2 | setosa |
| 3 | 4.7 | 3.2 | 1.3 | 0.2 | setosa |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 50 | 5.0 | 3.3 | 1.4 | 0.2 | setosa |
| 51 | 7.0 | 3.2 | 4.7 | 1.4 | versicolor |
| 52 | 6.4 | 3.2 | 4.5 | 1.5 | versicolor |
| 53 | 6.9 | 3.1 | 4.9 | 1.5 | versicolor |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 100 | 5.7 | 2.8 | 4.1 | 1.3 | versicolor |
| 101 | 6.3 | 3.3 | 6.0 | 2.5 | virginica |
| 102 | 5.8 | 2.7 | 5.1 | 1.9 | virginica |
| 103 | 7.1 | 3.0 | 5.9 | 2.1 | virginica |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 150 | 5.9 | 3.0 | 5.1 | 1.8 | virginica |

Es wird oft stillschweigend angenommen, dass die Stichprobe der Fälle homogen ist in dem Sinne, dass sich die Ergebnisse der Analyse für Teilstichpoben der Stichprobe nnur zufällig voneinander untereschеiden. In diesem Fall ist aber schon aus der Fischerschen Untersuchung (1936) bekannt, dass sich die drei Arten unterscheiden, d.h. die Punktekonfiguration der Fälle besteht aus Teilkonfigurationen, die sich duch geeignet gewählte Hyperebenen separieren lassen; Fishers Lineare Diskriminanzanalyse (LDA) wird an anderer Stelle vorgestellt. Dieser Befund legt nahe, dass sich die PCAs für diese Teilpopulationen voneinander unterscheiden. Geht man also gewissermaßen naiv an die Daten heran, so findet sich für die Gesamtstichpobe stets ein Ellipsoid, dass die Gesamtstichprobe im Sinne der Abbildung 3 beschreibt. Die Analyse der Gesamtstichprobe kann sich von den Analysen der Teilstichproben unterscheiden. Diese Teilmengen sind zunächst jede für sich mit einer PCA analysiert worden. Die Abbildung 5 zeigt die Resultate der Einzelanalysen.[10] Die Scree-Plots (Eigenwerte versus Rangplatz der Eigenwerte) für die drei Arten legen nahe, dass zwei latente Dimensionen einen gute Approximation an die Daten liefern (die Werte in den Klammern sind die prozentualen Anteile einer Dimension an der Gesamtvarianz). Die Plots der Faktorwerte (F-Scores) zeigen, dass die Orientierungen der Konfigurationen der Fälle jeweils mit der der ersten latenten Dimension übereinstimmen, – dies entspricht dem Ansatz der PCA, als erste latente Dimension die Orientierung mit der maximalen Varianz der Abbildubngen der Fälle zu wählen. Die Struktur der Variablen ist aber, wie die Faktorladungen der Variablen für die drei Klassen von Iris zeigen, für jede der drei Arten spezifisch. Während alle Variablen (petal width, petal length,

[10] Berechnungen in ProbierScriptDiscrim-R

Abbildung 5: PCA Iris-Daten, Einzelanalysen (standardisierte Daten)




etc) auf der ersten latenten Dimension nahezu identische Ladungen zeigen, variieren die Ladungen auf der zweiten latenten Dimension; sogar die Reihenfolge der Abbildungen auf die zweite Achse variiert von einem Iristyp zum anderen.

Fasst man die Daten für die einzelnen Arten zu einer Gesamtstichprobe zusammen, so ergibt sich das in Abbildung 6 gezeigte Bild, falls die Daten nur zentriert werden. Die erste latente Dimension separiert im Wesentlichen die drei Klassen *setosa, versicolor* und *virginica*, und die Faktorladungen für die vier Variablen unterscheiden sich von denen für die einzelnen Klassen, – was nicht verwunderlich ist, die Faktorladungen für die Gesamtstichprobe ergeben sich aus einer Art Mittelung über die Einzelstichproben. Bemerkenswert ist, dass sich die Gruppe der setosa-Pflanzen so deutlich von den beiden übrigen Arten versicolor und virginica unterscheidet, die zwar auch separate, aber gleichwohl dicht beieinander liegende Cluster bilden. Es ist die Trennung von setosa einerseits und versicolor und virginica andererseits, die die erste latente Dimension definiert. Die erste Reihe zeigt die Darstellung der Fälle der drei Gruppen in einem 2-dimensionalen laten-

Abbildung 6: PCA Iris-Daten (zentriert)

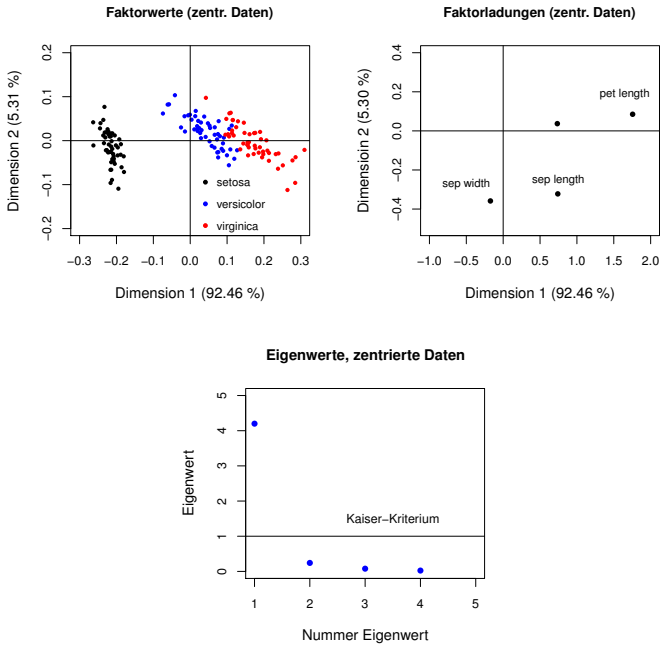
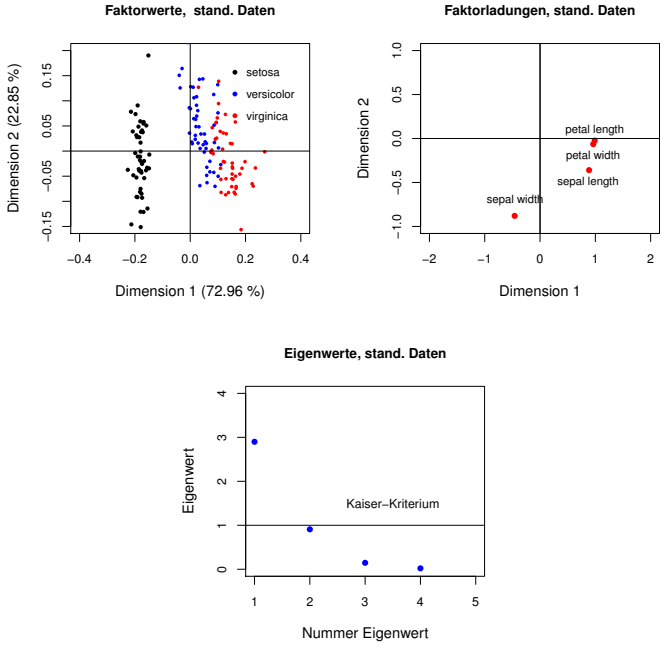



Abbildung 7: PCA der standardisierten Irisdaten



ten Koordinatensystem. Die erste Achse entspricht der größten Ausdehnung der Punktekonfiguration. Die zweite Reihe zeigt die Faktorladungen der vier Variablen sepal und petal length und sepal und petal width. Hier zeigen sich deutliche Unterschiede zwischen den drei Arten Setosa, Versicolor und Virginica. Die letzte Reihe zeigt den Verlauf der Eigenwerte für die drei Arten. Die erste latente Variable "erklärt" demnach den größen Teil der Varianz, während die zweite latente Variable zumindest dem Kaiser-Kriterium zufolge nur noch eine grenzwertige, im Falle Versicolor gar keine erklärende Funktion für die Erklärung der Daten hat.

Abbildung 7 zeigt die Resultate einer PCA, wenn über alle drei Klassen gewissermassen gemittelt wird. Die Faktorwerte zeigen eine deutliche Separation der

Tabelle 3: Matrix der Korrelationen

| | Sep. length | Sep. width | Pet. length | Pet. width |
|---|---|---|---|---|
| sep. length | 1.000 | | | |
| Sep. witdth | -.117 | 1.000 | | |
| Pet. length | .871 | -.428 | 1.000 | |
| pet. width | -.818 | -.366 | .962 | 1.000 |

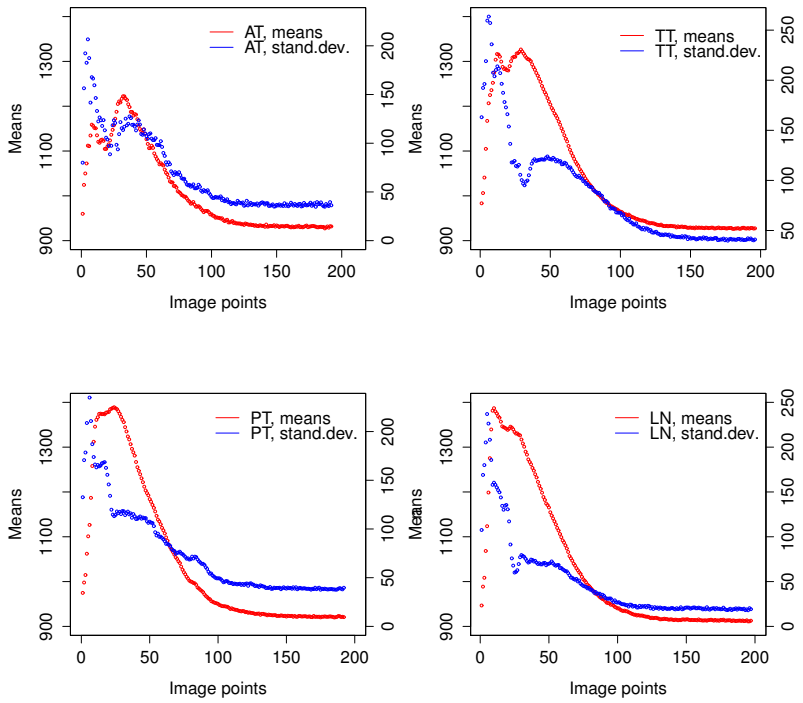
drei Iris-Klassen, auch wenn die Abbildungen der Punkte auf die die erste latente Achse für die Klassen Versicolor und Virginica gewisse Überlappungen zeigen. Deutlich wird die Klasse Setosa von den beiden übrigen Klassen getrennt. Die Eigenwerte zeigen, dass die Lösung maximal 2-dimenional ist; nach dem Kaiser-Kriterium ist die zweite Dimension kaum noch von Bedeutung. Von Interesse sind ebenfalls die Faktor-Ladungen. Die Struktur der Ladungen entspricht keiner der Strukturen in den Einzelanalysen, man kann sagen, dass sie ein Artefakt der Zusammenfassung der Daten ist. Dieser Befund wirft ein Schlaglicht auf die "naive" Interpretation von Analysen, die die Homogenität der Stichprobe von Fällen unterstellen. Allerdings kann die nähere Betrachtung der Konfiguration der Fälle möglicherweise Hinweise auf die Existenz verschiedenere Klassen von Fällen geben.

### 3.2 Die Analyse von Schilddrüsengeweben

In der Medizin müssen vielfach Gewebeproben klassifiziert werden. Für diese Aufgabe können OCT-Bilder (OCT – Optical Coherence Tomography) hilfreich sein. Bei dem hier betrachteten Gewebeproben handelt es sich um Schilddrüsengewebe. Die OCT-Bilder sind 2-dimensional. Es wurden 1-dimensionale Profile der Bilder angefertigt, die den Helligkeitsverlauf der Bilder über 190 bis 200 Pixel beschreiben. Die Frage war, ob diese Profile die für die Klassifikation notwendige Information enthalten. Dementsprechend gibt es 192 Prädiktoren (Pixel), deren Helligkeitswerte als Prädiktorwerte in die Analyse eingingen.

Man kann fragen, wieviele Dimensionen zur Beschreibung der Profile überhaupt benötigt werden, und ob die PCA eine gewisse Trennung der Gewebetypen

Abbildung 8: Mittlere Helligkeiten (Profile) und Standardabweichungen (rechte Skala) von OCT-Bildern (Schilddrüsengewebe)



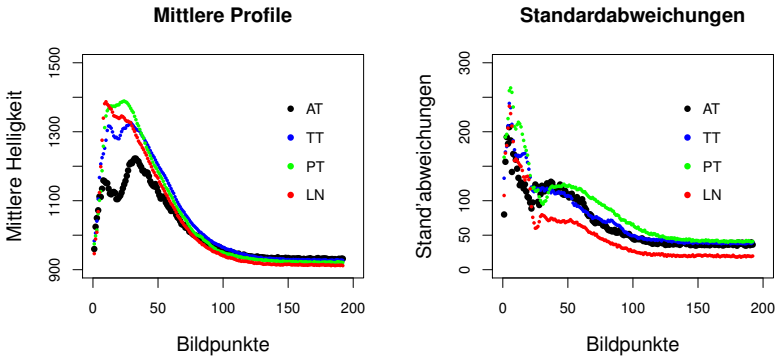

liefert, – schließlich soll die erste latente Dimension maximal zwischen den Fällen trennen, und falls die Fälle dem Gewebetyp entsprechend Cluster bilden, könnte es sein, dass zummindest die erste latente Dimension zwischen den Typen AT, TT, PT und LN diskriminiert. Insgesamt standen 291 "Fälle", d.h. Gewebeproben zur Verfügung: 26 Fälle für die Kategorie AT, 102 für die Kategorie TT, 89 für die Kategorie PT und 74 für die Kategorie LN.

Abbildung 8 zeigt die mittleren Profile und die Standardabweichungen pro Pixel. Die Abbildung 9 zeigt, dass sich die Profile hauptsächlich im Bereich 0 bis 50 Pixel unterscheiden; Beurteiler werden also insbesondere auf diesen Bereich fokussieren.

Die PCA ist einmal für zentrierte, und einmal für standardisierte Daten gerechnet worden. Die zentrierte Lösung macht insofern Sinn, als alle Bewertungen auf demselben Skalentyp abgegeben wurden. Allerdings können die Variablen verschiedene Varianzen haben, was sich auf die Faktorwerte (Gewebeproben) ebenso wie auf die Faktorladungen (Pixel)auswirken kann.

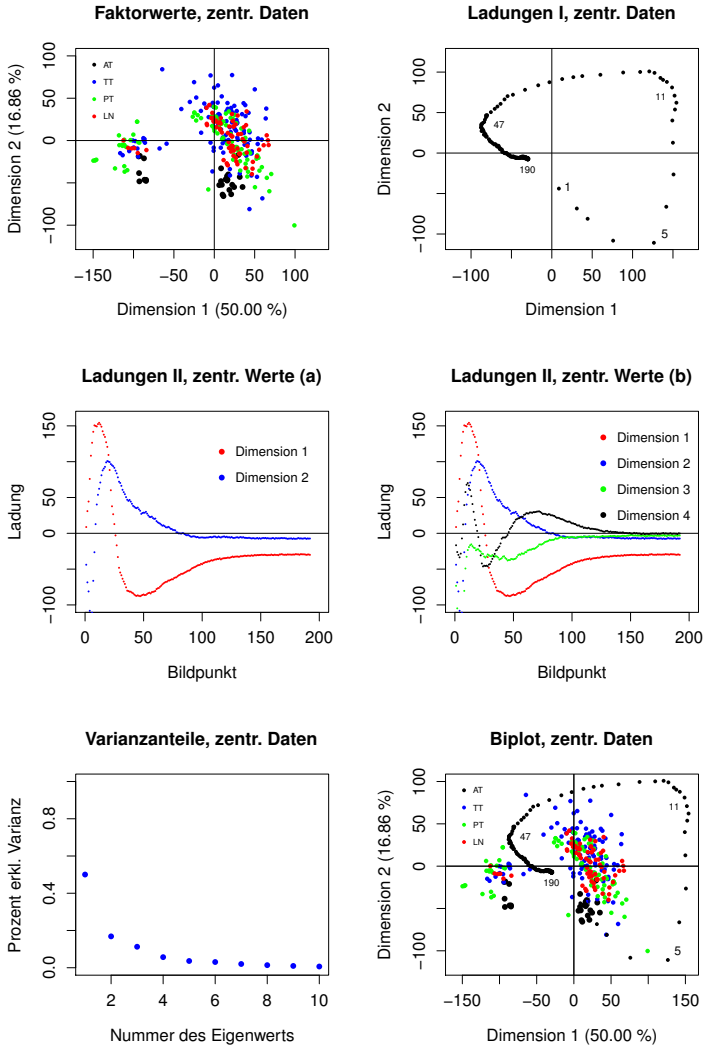
Abbildung 10 zeigt die Ergebnisse für die zentrierten Daten. Der Plot der Faktorwerte zeigt zwei Cluster, die die erste latente Variable definieren. Allerdings sind die Cluster nicht durch die Gewebetypen AT, TT, etc definiert, beide

Abbildung 9: Mittlere Helligkeitsprofile (Schilddrüsengewebe)




Cluster bestehen aus Mischungen dieser Typen. Es ist nicht klar, worin der Unterschied zwischen den Elementen des kleineren Clusters einerseits und des größeren Clusters andererseits ausmacht. Die beiden ersten latenten Variablen erklären überdies nur ca 67% der Gesamtvarianz, und es ist nicht klar, ob die restlichen 33% nur "Rauschen" bedeuten. Interessant ist der Verlauf der Ladungen für die "Variablen", also den Pixeln. Die Ladungen für die ersten 50 Pixel unterschieden sich deutlich voneinander, im Vergleich zu den Ladungen der restlichen 120 Pixel. Die mittleren Helligkeitsverläufe legen nahe, warum dies so ist. Während in Ladungen I die Ladungen der Pixel als Koordinaten der Pixen auf den latenten Dimensionen aufgetragen wurden, werden in Ladungen II die Ladungen auf der ersten und der zweiten latenten Dimension gegen die Pixel selbst aufgetragen. In (b) sind zusätzlich noch die Ladungen für die Dimensionen 3 (grün) und 4 (schwarz) aufgetragen worden. r Beitrag dieser Dimensionen ist offenbar geringer als der der ersten beiden Dimensionen, scheint aber noch hinreichend systematisch zu sein um die Vermutung zu rechtfertigen, dass diese Dimensionen nicht nur Rauschen abbilden. Abbildung 11 zeigt die Ergebnisse für die standardisierten Daten. Die Inspektion der Varianzanteile zeigt, dass die erste Dimension im Vergleich zu den Übrigen eine sehr viel dominantere Rolle zu spielen scheint. Zusammen erklären die beiden ersten Dimensionen ca 81% der Varianz in den Daten. Davon abgesehen zeigt sich die Aufspaltung zwei von den Gewebetypen unabhängige Cluster, wie schon bei den zentrierten Daten. Der Verlauf der Faktorladungen unterscheidet sich allerdings vom Verlauf bei den zentrierten Daten: Die Ladungen für die Dimension 1 streben gegen 1 und nicht, wie bei den zentrierten Daten, gegen Null. Die Darstellung des Biplots ist allerdings nicht ganz korrekt, weil sowohl die Faktor*werte* wie die Faktor*ladungen* eingezeichnet wurden, – es geht bei der Darstellung mehr um das Prinzip des Biplot. Was sich allerdings vermuten läßt, ist, dass das kleine Cluster durch Abweichungen in den höheren Pixelbereichen erzeugt wird.

Abbildung 10: Schilddrüsendaten: erklärte Varianz, zentr. Daten




# 4 PCA und Faktorenanalyse

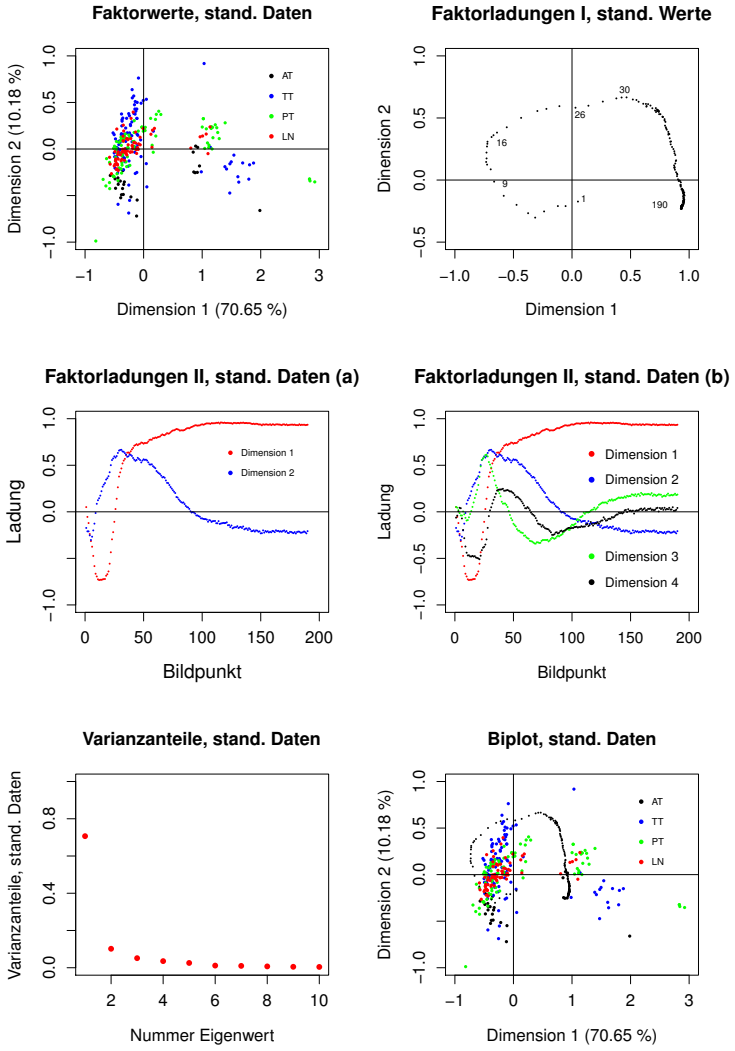
## 4.1 Die Annahmen

Bei der Faktorenanalyse wird angenommen, dass die zentrierten Datenvektoren $\mathbf{x}_j$ sich (i) als Linearkombinationen "gemeinsamer Faktoren" $\mathbf{F}_k$, $k = 1, \ldots, r$, sowie eines jeweilig "spezifischen Faktor"' $\mathbf{e}$ ergeben:

$$\mathbf{x}_j = b_{1j}\mathbf{F}_1 + \cdots + b_{rj}\mathbf{F}_r + \mathbf{e}_j, \quad j = 1, \ldots, n \tag{4.1}$$

wobei die $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_r$ hypothetische, unkorrelierte "Faktoren", also latente Größen sind, und die $b_{jk}$ sind Gewichte, mit denen die Faktoren in die $j$-te gemes-

Abbildung 11: Schilddrüsendaten: erklärte Varianz, stand. Daten




sene Variable eingehen. Die $\mathbf{x}_j$ werden explizit, ebenso die $\mathbf{F}_k$, als Zufallsvektoren aufgefasst. Es sei $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_n]$; dann kann der Ansatz in Matrixform geschrieben werden:

$$X = FB' + \mathbf{e} \tag{4.2}$$

**Annahmen:**

$$\mathbb{E}(\mathbf{e}) = 0, \quad \mathbb{E}(F) = 0, \quad \mathbb{E}(F'F) = I, \quad \mathbb{E}(X) = 0. \tag{4.3}$$

Es werde insbesondere $X = Z$ gesetzt, d.h. es werden standardisierte Messwerte betrachtet. Dann ist $\frac{1}{m}Z'Z = R$ die Matrix der Korrelationen zwischen den Variablen. Der Anstz (4.2) impliziert dann

$$mR = (FBA' + \mathbf{e})'(FB' + \mathbf{e}) = BF'FB' + +BF'\mathbf{e} + \mathbf{e}'FB' + \mathbf{e}'\mathbf{e} \tag{4.4}$$

Die Komponenten von $\mathbf{e}$ werden als statistisch unabhängig angenommen, so dass

$$\mathbf{e}'\mathbf{e} = \mathrm{diag}(e_1^2, \ldots, e_n^2). \tag{4.5}$$

Weiter wird angenommen, dass die Fehler und die Faktoren statistisch unabhängig sind, so dass $AF'\mathbf{e} = 0$ und $\mathbf{e}'FB' = 0$. Dann resultiert

$$\frac{1}{m}\mathbb{E}(Z'Z) = R = B\mathbb{E}(F'F)B' + \mathbf{e}'\mathbf{e} = BB' + \mathbf{e}'\mathbf{e}, \tag{4.6}$$

Diese Gleichung wird gelegentlich als *Fundamentaltheorem der Faktorenanalyse* bezeichnet. Für $r_{jj}$ ergibt sich

$$r_{jj} = \sum_{k=1}^{r} b_{kk}^2 + e_j^2 \tag{4.7}$$

Hierin ist

$$h_j^2 = \sum_{k=1}^{r} b_{kk}^2 \tag{4.8}$$

die *Kommunalität*; die Kommunalität ist derjenige Anteil an $r_{jj}$, der auf die gemeinsamane Faktoren zurückgeführt werden kann, während $e_j^2$ der Anteil ist, der auf den spezifischen Faktor in der $j$-ten Variablen zurückgeht. Wegen $r_{jj} = 1$ folgt

$$h_j^2 + e_j^2 = 1. \tag{4.9}$$

Oft wird die zusätzliche Annahme gemacht, dass die Daten multivariat Gaußverteilt sind. Diese Annahme – falls sie gerechtfertigt ist – erlaubt die Anwendung der Maximum-Likelihood-Methode zur Schätzung der freien Parameter.

## 4.2 Approximation: die Hauptkomponentenanalyse

Vergleicht man den auf der SVD beruhenden Ansatz der PCA mit dem der Faktorenanalyse, so fällt auf, dass der wesentliche Unterschied zwischen dder PCA und der Faktorenanalyse (FA) darin besteht, dass bei der FA die spezifischen Faktoren und damit die Kommunalitäten der Variablen eingeführt werden. Oft wird angeführt, dass die FA ein *Modell*, die PCA dagegen nur eine Beschreibung der Daten sei. Man muß allerdings bedenken, dass auch die PCA auf einer zentralen Annahme beruht, nämlich dass sich die beobachteten Vektoren $\mathbf{x}_j$ als Linearkombinationen der latenten Vektoren ergeben, $X = LT'$; die Annahme der Orthonormalität von $T$ führt dann zu $XT = L$, d.h. die Vektoren von $L$ sind Linearkombinationen der Spaltenvektoren von $X$. Die Linearitätsannahme ist ja nicht trivial: die Frage ist doch, warum sie gelten soll. Ein Antwort auf diese Frage könnte in dem Hinweis bestehen, dass lineare Funktionen oft eine gute Annäherung an die "wahren" nichtlinearen Funktionen bestehen. Man hat es dann bei der PCA ebenfalls mit einem Modell zu tun, dass eben (i) in der Annahme der Linearität als Approximation an möglicherweise existierende Beziehungen und (ii)

in der Annahme, dass die latenten Variablen mit "hinreichend kleinen" Eigenwerten eben "Fehlervariablen" reflektieren. Setzt man $B = A = T\Lambda^{1/2}$ und $F = Q$, so liefert die PCA eine als *Hauptkomponentenanalyse* bekannnte Startlösung für die FA.

# 5 Zusammenfassung

Die PCA ist ein handliches Verfahren, um sich schnell Einen eindruck von der Mehrdimensionalität eines Datensatzes zu verschaffen. Darüber hinaus leistet es gute Dienste im Zusammenhang mit anderen Verfahren wie der multiplen Regression, wenn es Probleme mit korrelierenden Prädiktorvariablen gibt. Die PCA liefert eine erste Appoximation an eine faktorenanalytische Diskussion eines Datensatzes.

Im Zusammenhang mit der FA wird oft angemerkt, dass die FA ein modell der Daten reprräsentiere, während die PCA "nur" ein Beschreibung der Daten liefere. Es darf allerdings nicht übersehen werden, dass es keine Beschreibung ohne theoretische Begriffe gibt, – der Hintergrund dieser Aussage liegt in den Diskussionen der ursprünglichen Annahmen der Philosophen des Wiener kreises, die eine metaphyikfreie, auf elementaren, d.h. theoriefreien Aussagen beruhende Wissenschaft aufbauen wollten. Dieser Anspruch führte bereits innerhalb des Wiener Kreises zur *Protokollsatzdebatte*, deren Resultat die Einsicht war, dass auchh einfache Protokollsätze nicht notwendig frei von jeder Theorie sind, schon Protokollätze sind "theoriegetränkt". Bei der PCA besteht diese nicht weiter hintergehbare Theorie in der Annahme, dass die Variablen linear aufeinander wirken. In http://www.uwe-mortensen.de/fakanalysews0506b.pdf werden nichtlineare Modelle vorgestellt. Ein konzeptuell aufwändigeres Verfahren ist die Kernel-PCA, bei der eine Abbildung in einen Raum höherer Dimension ein additives Modell zuläßt. Die Kernel-PCA muß wegen der vorbereitenden Betrachtungen gesondert dargestellt werden.

Nimmt man bei der PCA noch die Annahme, die Daten seien multivariat normalverteilt hinzu, so lassen sich die orthogonalen latenten Dimensionen als Repräsentationen statistisch unabhängige Merkmale interpretieren. Ein Ansatz, von vorn herein auf statistisch unabhängige latente Variablen zu zielen, ist die *Independent Component Analysis* (ICA). Bei dieser Analyse wird gerade vorausgesetzt, dass die Daten *nicht* multivariat Gauß-verteilt sind. Auch hier wird eine gesonderte Darstellung nötig.

# 6 Anhang

Dazu erinnere man sich an die Forderung, dass die $\mathbf{L}_k$ aus den $\mathbf{z}_j$ erreichnet werden müssen, – es sind ja nur die Daten $Z$ gegeben. Nach (**??**) ist die Beziehung zwischen den $\mathbf{z}_j$ und den $\mathbf{L}_k$ linear, so dass man folgern kann, dass sich die $\mathbf{L}_k$ aus den $\mathbf{z}_j$ ebenfalls als Linearkombination ergeben. Es sei also $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$

eine Matrix mit $n$-dimensionalen Spaltenvektoren $\mathbf{v}_k$ derart, dass

$$ZV = L, \quad Z\mathbf{v}_k = \mathbf{L}_k,\ k = 1, \ldots, n \tag{6.1}$$

gilt. Nach A1 sollen die $\mathbf{L}_k$ orthogonal sein; dementsprechend muß

$$V'(Z'Z)V = L'L = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \tag{6.2}$$

gelten, wobei $\lambda_k = \mathbf{L}_k'\mathbf{L}_k = \|\mathbf{L}_k\|^2$ ist. Nach A2 soll $\lambda_1$ maximal sein. Dazu kann man festellen, dass $\mathbf{v}'Z'Z\mathbf{v}$ eine quadratische Form ist[11], und nach dem Satz von Courant-Fischer gilt[12]

$$Q(\mathbf{v}) = \frac{\mathbf{v}'Z'Z\mathbf{v}}{\mathbf{v}'\mathbf{v}} = \frac{\mathbf{v}'Z'Z\mathbf{v}}{\|\mathbf{v}\|^2} = \frac{\mathbf{v}'}{\|\mathbf{v}\|}Z'Z\frac{\mathbf{v}'}{\|\mathbf{v}\|} = \max \tag{6.3}$$

genau dann, wenn $\mathbf{v}/\|\mathbf{v}\| = \mathbf{t}_1$, wobei $\mathbf{t}_1$ der zum größten Eigenwert $\lambda_1$ korrespondierende Eigenvektor von $Z'Z$ ist. Für $\mathbf{L}_2$ folgt dann ebenfalls aus dem Satz von Courant-Fischer, dass $\mathbf{L}_2 = Z\mathbf{t}_2$ ist $\mathbf{t}_2$ der zum zweitgrößten Eigenwert $\lambda_2$ korrespondierende Eigenvektor von $Z'Z$, etc. Damit wird A2 erfüllt, wenn $V = T$, $T$ die $(n, n)$-Matrix der Eigenvektoren von $Z'Z$ gesetzt wird. Die in A2 genannte Nebenbedigung besteht darin, dass die $\mathbf{v}$ auf die Länge 1 normiert sein muß, und sie wird in (6.3) durch die Betrachtung von $\mathbf{v}/\|\mathbf{v}\|$ erfüllt. Tatsächlich wird für die Eigenvektoren $\mathbf{t}_k$ gefordert, dass sie die Länge 1 haben. Da Eigenvektoren symmetrischer Matrizen außerdem orthogonal sind folgt, dass $T$ orthonormal ist. d.h. es gilt $T'T = I_n$, $I_n$ die $(n, n)$-Einheitsmatrix.

[11] V& M, Abschn. 2.5.2
[12] V& M, Abschn. 2.5.4

## Literatur


[1] Brachinger, HW, Ost F.: Modelle mit latenten Variablen: Faktorenanalyse, Latent-Structure-Analyse und LISREL-Analyse. In: Fahrmeier, L., Hamerle, A., Tutz, G. (Hrsg): Multivariate statistische Verfahren. Walter de Gruyter, Berlin, New York, 1996

[2] Busemeyer, J.R., Jones, L. E. (1983) Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, 93 (3), 549 - 562

[3] Cliff, N. (1988) The Eigenvalues-Greater-Than-One-Rule and the reliability of components. *Psychological Bulletin*, 103, 276-279

[4] Christoffersson, A. (1975) Factor analyis of dichotomized variables. *Psychometrika*, 40, 5-32

[5] Davenport, M., Studdert-Kennedy, H. (1970) Use of orthogonal factors for selection of variables in a regression equation. *Applied Statistics*, 21, 324–333

[6] Eckart, C., Young, G. (1936) The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211–218

[7] Feller, W.: An introduction to probability theory and its applications, Vol. II, New York 1966

[8] Gabriel, K.R. (1971) The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58, 453–467

[9] Gabriel, K.R. (1978) Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Society B*, 40, 186-196

[10] Golub, G.H., Van Loan, C.F.: Matrix Computations. Baltimore 2013.

[11] Gould, J.: Der falsch vermessene Mensch. Frankfurt, 1988

[12] Harman, H.H.: Modern Factor Analysis. Chicago, 1967

[13] Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal Educational Psychology*, 24 (7), 498-520

[14] Gower, C., Hand, D.J.: Biplots. Chapman & Hill, London, 1996

[15] Guttman, L: (1956) Image theory for the structure of quantitative variates, *Psychometrika*, 18, 277–296

[16] Kelly, T.L. (1940) Comment on Wilson and Worcester's "Note on factor analysis". *Psychology*, 5, 117 – 120

[17] Kenny, D. A., Judd, C. M. (1984) Estimating the nonlinear interactive effects of latent variables. *Psychological Bulletin*, 96 (1), 201 – 210

[18] Magidson, J., Vermunt, J. K. (2002) Latent class models for clustering: a comparison wiht K-means. *Canadian Journal of Marketing Research*, 20, 37 – 44

[19] Muthé, B. O. (2002) Beyond SEM: General latent variable modelling. *Behaviormetrika*, 29 (1), 81 – 117

[20] Pearson, K. (1901) On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 6, 557–572

[21] Rist, F., Glöckner-Rist, A., Demmel, R. (2009) The Alcohol Use Disorders Identification Test revisited: Establishing its structure using nonlinear factor analysis and identifying subgroups of respondents using latent factor analysis. *Drug and Alcohol Dependence*, 100, 71 –82

[22] Spearman, C. (1904) General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201 – 293

[23] Thurstone, L.L. (1931) Multiple Factor Analysis, *Psychological Review*, 38, 406 – 427

# Index