7 Hauptkomponenten- und Faktor-Analyse

7.1 Hauptkomponenten

a **Ziel.** Zusammenhänge zwischen zwei Variablen können wir gut grafisch und auch bedeutungsmässig erfassen. Bei drei Variablen ist dies auch noch machbar, ab vier wird es schwierig. Es ist deshalb wünschenswert, ein Problem, das viele Variable betrifft, auf eines mit zwei bis drei Dimensionen zurückzuführen.

Im Beispiel der NIR-Spektren wurde in der Einleitung (1.2.g) plausibel gemacht, dass sich das Wesentliche des chemischen Prozesses in wenigen Dimensionen „abspielt". In vielen anderen Fällen gibt es keine solche Theorie, und dennoch können wir versuchen, **aus den gegebenen Variablen zwei oder drei neue zu bilden, die „das Wesentliche", das in den Daten steckt, erfassen.**

Diese Aufgabe der **Dimensions-Reduktion** lässt sich gut veranschaulichen, wenn wir zunächst diskutieren, wie man ein zweidimensionales Problem in ein eindimensionales verwandelt – was eigentlich nicht nötig ist.

Daten und Modell. Diesen und den folgenden Abschnitt kann man auch studieren, wenn man das Kapitel über Modelle noch nicht erarbeitet hat. Es gibt jeweils eine Daten- und eine Modell-Version der Konzepte, und man kann die zweite überspringen.

b **Zwei Variable.** Wir wollen zunächst zweidimensionale Daten in sinnvoller Weise in eindimensionale verwandeln. Abbildung 7.1.b veranschaulicht die Idee der **Hauptkomponenten-Analyse**:

> Die „Streuung" in den Daten soll durch Projektion der Datenpunkte auf eine geeignete Richtung so gut wie möglich wiedergegeben werden. Präzisieren wir dies so, dass die Varianz der projizierten Punkte möglichst gross sein soll!

Wie diese Richtung bestimmt wird, wird unten für den Fall von mehr als zwei Variablen angegeben. In der Abbildung ist auch die Richtung senkrecht auf die Richtung der maximalen Streuung eingezeichnet. Man kann zeigen, dass dies die Richtung ist, für die die Projektionen der Punkte die kleinste Varianz haben.

Modell-Version. Wir haben diskutiert, wie die „Streuung" der Daten durch Ellipsen dargestellt werden kann, die „Höhenlinien" gleicher Dichte der entsprechenden bivariaten Normalverteilung darstellen. Die Richtung der längeren Hauptachse der Ellipse maximiert die Varianz der Projektion des Zufallsvektors $\underline{X}$. Die kürzere Hauptachse der Ellipse, die auf der längeren senkrecht steht, gibt die Richtung mit der kleinsten Varianz an. (* Das bedeutet, dass die längere Hauptachse mit der Geraden zusammenfällt, die durch *orthogonale Regression* bestimmt wird.)

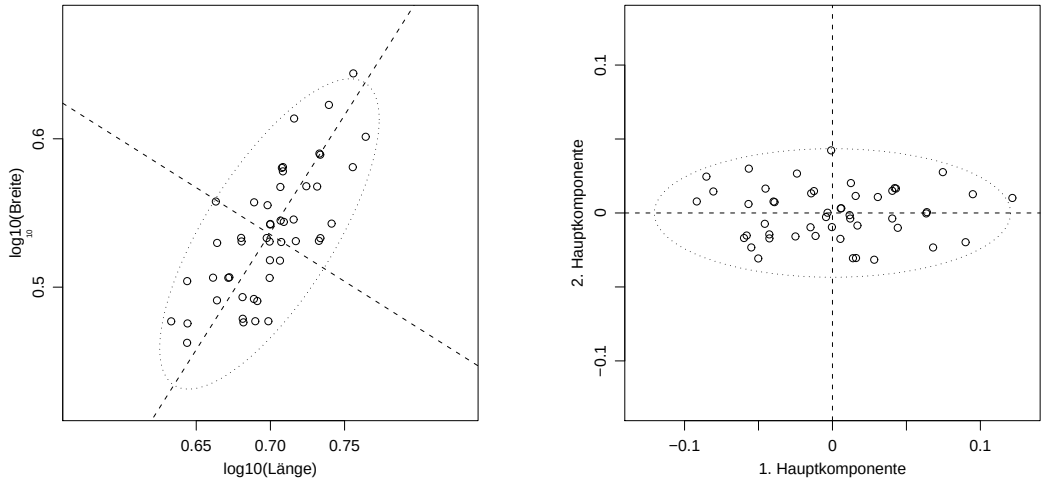

Abbildung 7.1.b: Hauptkomponenten im Beispiel der Iris-Blüten

c **Die Hauptachsen-Transformation.** Die beiden gefundenen Richtungen können als **neues Koordinatensystem** aufgefasst werden. Sie werden auch als **Faktoren** bezeichnet – als Anleihe aus der allgemeineren Methode der **Faktor-Analyse**, siehe unten (7.4.a). Zum Begriff eines Faktors in der Varianzanalyse steht diese Verwendung des Wortes leider im Widerspruch!

Die Umrechnung vom ursprünglichen Koordinatensystem der Merkmale $X^{(j)}$ in Hauptkomponenten-Koordinaten erfolgt für die Daten durch $\underline{z}_i = \widehat{\boldsymbol{B}}\,(\underline{x}_i - \overline{\underline{x}})$ mit einer Matrix $\widehat{\boldsymbol{B}}$, die sich aus der Kovarianzmatrix $\widehat{\boldsymbol{\Sigma}}$ bestimmt. Für die ganze Datenmatrix wird das zu

$$\boldsymbol{z} = \boldsymbol{x}_c \widehat{\boldsymbol{B}}^T \ .$$

Modell-Version. Im Modell wird der Zufallsvektor im neuen Koordiantensystem durch die lineare Transformation

$$\underline{Z} = \boldsymbol{B}\,(\underline{X} - \underline{\mu})$$

bestimmt, wobei $\boldsymbol{B}$ von $\boldsymbol{\Sigma}$ abhängt.

Die Matrizen $\widehat{\boldsymbol{B}}$ und $\boldsymbol{B}$ sollen orthogonal sein, damit die Abstände zwischen den Punkten gleich bleiben, so dass in diesem Sinne an den „Beziehungen zwischen den Beobachtungen“ nichts geändert wird. Wie man sie berechnet, werden wir gleich anführen.

d **Standardisierte Zufallsvariable.** Die Matrix $\boldsymbol{z}$ ist hier keine standardisierte Datenmatrix wie in 2.6.m. Die Spalten $Z^{(1)}$ und $Z^{(2)}$ haben zwar Mittelwert 0 und sind, wie sich zeigen wird, unkorreliert, sie haben aber verschiedene Varianzen. Um eine standardisierte Datenmatrix zu erhalten, müssen wir die einzelnen Spalten noch durch ihre Standardabweichungen dividieren.

Modell-Version. $\underline{Z}$ ist hier kein standardisierter Zufallsvektor wie in 3.2.k. Die Komponenten $Z^{(1)}$ und $Z^{(2)}$ haben zwar Erwartungswert 0 und sind unkorreliert, sie ha-

ben aber verschiedene Varianzen. Um einen standardisierten Zufallsvektor zu erhalten, müssen wir die $Z^{(j)}$ noch durch ihre Standardabweichungen dividieren.

Wir haben früher standardisierte Datenmatrizen (Zufallsvektoren) eingeführt, ohne die Hauptkomponenten-Analyse zu benützen, indem wir die Transformationsmatrix aus der Cholesky-Zerlegung von $\widehat{\mathbf{\Sigma}}$ ($\mathbf{\Sigma}$) bestimmt haben (2.6.m). Das zeigt nochmals konkret, dass die Bestimmung von standardisierten mehrdimensionalen Daten auf verschiedene Weise möglich ist.

e **Mehr als 2 Variable.** Wir verlangten, dass $Z^{(1)}$ eine möglichst grosse Varianz habe, was die erste Hauptkomponente oder die erste Zeile von $\widehat{\boldsymbol{B}}$ ($\boldsymbol{B}$) festlegt. Für $m=2$-dimensionale Daten ist damit die Drehung des Koordinatensystems klar.

Für $m>2$ bleibt die Frage, welche lineare Transformation die zweite Hauptkomponente bilden soll. Die naheliegende Anforderung ist wieder, dass die neue Variable $Z^{(2)}$ möglichst stark streuen soll – unter der Nebenbedingung, dass die gesamte Transformationsmatrix $\widehat{\boldsymbol{B}}$ (respektive $\boldsymbol{B}$) schliesslich orthogonal sein soll. Mit solchen Forderungen können wir fortfahren, bis zur Bestimmung der letzten Variablen $Z^{(m)}$ keine Wahlfreiheit mehr bleibt.

f **Berechnung der Transformationsmatrix.** Die lineare Algebra liefert die Lösung dieses ambitiös tönenden Problems, unter dem Stichwort **Eigenwert-Problem**. Die Matrix $\widehat{\boldsymbol{B}}$ hängt nur von der Kovarianzmatrix $\widehat{\mathbf{\Sigma}}$ ab. Sie umfasst (als Zeilen $\underline{\widehat{b}}_k$) die so genannten **Eigenvektoren** von $\widehat{\mathbf{\Sigma}}$. Die Reihenfolge ist eigentlich willkürlich und wird normalerweise durch die Konvention bestimmt, dass die zugehörigen **Eigenwerte** $\widehat{\lambda}_k$ der Grösse nach sortiert sein sollen, so dass $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq ... \geq \widehat{\lambda}_m$ gilt.

Wenn diese Matrix $\widehat{\boldsymbol{B}}$ für die lineare Tranformation $\underline{z}_i = \widehat{\boldsymbol{B}}(\underline{x}_i - \overline{\underline{x}})$ benützt wird, dann wird

$$\widehat{\mathrm{var}}\langle \underline{Z} \rangle = \widehat{\boldsymbol{B}}\,\widehat{\mathbf{\Sigma}}\,\widehat{\boldsymbol{B}}^T = \widehat{\boldsymbol{D}} = \begin{bmatrix} \widehat{\lambda}_1 & 0 & \dots & 0 \\ 0 & \widehat{\lambda}_2 & \dots & 0 \\ \cdot & \vdots & \ddots & \\ 0 & 0 & \dots & \widehat{\lambda}_m \end{bmatrix} .$$

Die empirischen Varianzen der neuen Koordinaten $z_i^{(k)}$ sind also gleich den Eigenwerten der Matrix $\widehat{\mathbf{\Sigma}}$, und die $\underline{z}^{(k)}$ sind unkorreliert.

Modell-Version. Die genau gleichen Beziehungen gelten für die Zufalls-Vektoren $\underline{X}$ und $\underline{Z}$, wenn man in den Gleichungen die „Hüte“ $\widehat{\ }$ weglässt.

g* Wir wollen skizzenhaft den Zusammenhang zwischen dem gestellten statistischen Problem und dem Eigenwert-Problem der linearen Algebra formulieren, für die Version ohne „Hüte“. Wir suchten in 7.1.b eine orthogonale Transformation, die die Ellipse in Hauptachsenlage bringt. Das bedeutet, dass die Kovarianzmatrix von $\underline{Z}$ diagonal sein soll. Wir suchen also ein orthogonales $\boldsymbol{B}$, so dass die vorhergehende Gleichung gilt. Wenn man diese Gleichung von links mit $\boldsymbol{B}^T$ multipliziert, erhält man $\mathbf{\Sigma}\,\boldsymbol{B}^T = \boldsymbol{B}^T\boldsymbol{D}$. Wenn wir die Spalten dieser Matrix-Gleichung einzeln aufschreiben, wird das zu $\mathbf{\Sigma}\underline{b}_k = \lambda_k \underline{b}_k$. Das ist die Gleichung, die in der linearen Algebra **Eigenwert-Problem** heisst. Wie die Lösung konkret berechnet wird, wollen wir hier nicht diskutieren. (Es gibt maximal m verschiedene Lösungen für λ_k, aber immer mindestens m verschiedene $\underline{b}_k$. Wenn es m verschiedene λ_k gibt, sind die $\underline{b}_k$ eindeutig, sonst ist es komplizierter.)

h Ziel der Übung war die **Dimensions-Reduktion**. Ein zweidimensionaler Datensatz kann auf die erste Hauptachse projiziert werden. Ebenso kann man sich in einem höher-dimensionalen Raum auf die ersten p Hauptkomponenten beschränken, die zusammen einen maximalen Anteil der „Streuung in den Daten“ wiedergeben. Präzise formuliert, transformiert man die Daten in Hauptkomponenten-Koordinaten und lässt jene weg (setzt jene null), die den $m-p$ kleinsten Eigenwerten λ_j entsprechen.

i **Modell.** Diesen einfachen Gedanken kann man auch als Modell formulieren, das einem linearen Regressions-Modell ähnlich sieht. Man löst 7.1.c nach $\boldsymbol{x}_c$ auf und erhält, da $\widehat{\boldsymbol{B}}^{-1} = \widehat{\boldsymbol{B}}^T$ ist,

$$\boldsymbol{x}_c = \boldsymbol{z}\,\boldsymbol{B} \ .$$

Wenn wir jetzt auf der rechten Seite nur die ersten p Hauptkomponenten, also die ersten p Spalten von $\boldsymbol{z}$ „ernst nehmen“ und sie mit $\boldsymbol{S} = \boldsymbol{z}^{[1:p]}$ bezeichnen, erhalten wir als „wesentlichen Teil“ von $\boldsymbol{x}_c$ die Matrix $\widehat{\boldsymbol{x}}_c$,

$$\boldsymbol{x}_c \approx \widehat{\boldsymbol{x}}_c = \boldsymbol{S}\,\boldsymbol{C}^T \ ,$$

wobei $\boldsymbol{C}^T = \boldsymbol{B}_{[1:p]}$ die ersten p Zeilen von $\boldsymbol{B}$ umfasst. Wir spalten deshalb die obige Gleichung auf in einen „wesentlichen Teil“ und einen Rest $\boldsymbol{r}$,

$$\boldsymbol{x}_c = \boldsymbol{S}\,\boldsymbol{C}^T + \boldsymbol{r} \ .$$

Das sieht ähnlich aus wie ein Regressionsmodell in Matrix-Form, $\underline{Y} = \boldsymbol{X}\underline{\beta} + \underline{E}$, und noch ähnlicher wie die multivariate Form davon $\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{E}$, siehe 6.1.b. Es suggeriert die Aufspaltung der Daten in eine „systematische Struktur“ und Abweichungen, die man als zufällig interpretierten kann.

Das wollen wir hier allerdings nicht tun; die Überlegung soll nur als Querverbindung und als Anknüpfungspunkt für später dienen. **Die Hauptkomponenten-Analyse ist im Wesentlichen eine Methode der Beschreibenden Statistik.** Um die obige Gleichung als Modell zu brauchen, müssten wir Verteilungs-Annahmen mindestens über die Abweichungen $\boldsymbol{r}$ treffen. Das werden wir später tun und damit zur **Faktor-Analyse** vorstossen.

j ▷ **I**m **Beispiel der NIR-Spektren** bilden die Absorptionen für $m = 700$ Wellenlängen im nahen Infrarot-Bereich zwischen 1100 und 2500 nm die Variablen, die für $n = 121$ Zeitpunkte im Verlauf einer chemischen Reaktion gemessen wurden. Abbildung 7.1.j stellt den Verlauf im Raum der ersten fünf Hauptkomponenten durch eine Streudiagramm-Matrix dar. In den ersten vier Hauptkomponenten sieht man eine „Kurve“, die dem zeitlichen Verlauf der Reaktion entspricht, und die sicher als „die wesentliche Struktur“ der Daten interpretiert werden kann. ◁

k **Wahl der Dimension.** Wie gross man p wählen soll, ist vom Zweck der Analyse abhängig. In einer Streudiagramm-Matrix der Hauptkomponenten $\boldsymbol{Z}$ beginnt man links oben mit der Betrachtung und hört auf, wenn keine interessante Struktur mehr in den einzelnen Diagrammen erkennbar ist. Im Beispiel zeigt die fünfte Hauptkomponente das typische Bild reiner Zufallsstreuung.

Ein nützliches Hilfsmittel zur Wahl der Dimension bilden die Varianzen der Hauptkomponenten $\mathrm{var}\langle Z^{(k)}\rangle$. Man kann die paar Zahlen grafisch darstellen durch ein Balkendiagramm; das Bild wird als **scree plot** bezeichnet.

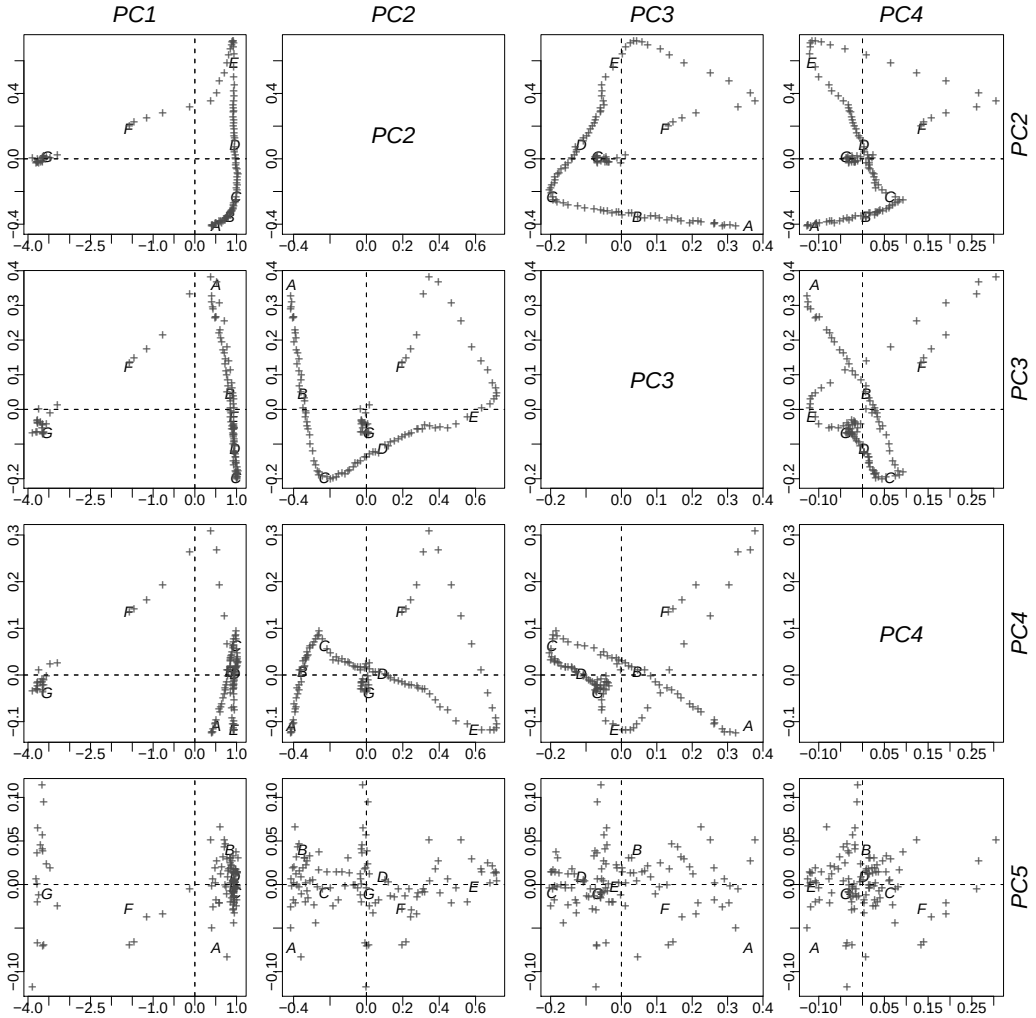

Abbildung 7.1.j: Streudiagramm-Matrix der ersten fünf Hauptkomponenten im Beispiel der NIR-Spektren

Im Beispiel (Abbildung 7.1.k) zeigt sich, dass die erste Dimension den Grossteil der Variabilität der Daten bereits erfasst. Die logarithmierte Darstellung lässt erahnen, dass weitere drei bis vier Komponenten ebenfalls eine Bedeutung haben. Nachher wird die „Kurve“ deutlich flacher, und man kann vermuten, dass diese Komponenten keine wichtige Information mehr beinhalten. Man sucht gerne nach einem solchen „Knie“ im scree plot. Das „Gelenk“ (die Ecke) entspricht dann der ersten Komponente, die man vernachlässigen kann, und p wird um 1 kleiner gewählt.

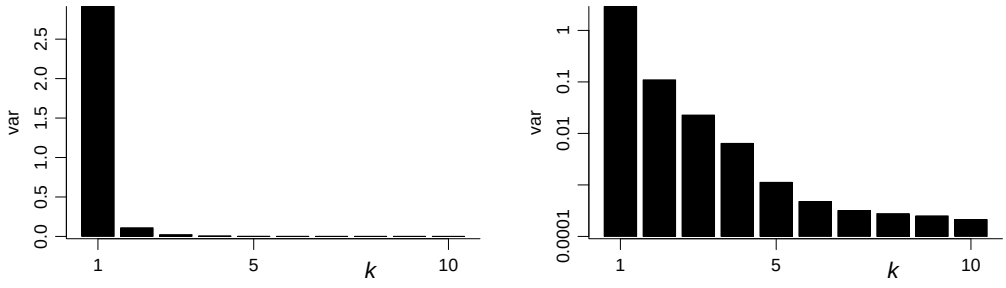

Abbildung 7.1.k: Varianzen der Hauptkomponenten

7.2 Der Biplot

a Die Hauptkomponenten-Analyse liefert die Näherung $\boldsymbol{x}_c \approx \widehat{\boldsymbol{x}}_c = \boldsymbol{S}\,\boldsymbol{C}^T$, bei der $\boldsymbol{S}$ und $\boldsymbol{C}$ nur p Spalten haben. Ist die Näherung für $p = 2$ gut, so stellt das Streudiagramm der ersten beiden Hauptkomponenten „das Wesentliche“ dar. Wenn man darauf achtet, dass die Einheiten auf den beiden Achsen gleich sind, dann sind die Abstände zwischen den Punkten ungefähr gleich den Abständen der Beobachtungs-Vektoren, also stellen sie Unähnlichkeiten zwischen Beobachtungen sinnvoll dar.

b In der gleichen Figur zeichnen wir nun auch die Zeilen $\underline{C}_j$ der Matrix $\boldsymbol{C}$ als Pfeile ein, die die Variablen charakterisiern (Abbildung 7.2.b). (Um sinnvolle Grössen der Pfeile zu erhalten, muss man die Skala anpassen; sie ist am oberen und rechten Rand angegeben.)

Die erwähnte Näherung, geschrieben für den Wert der $j$ten Variablen für die $i$te Beobachtung, lautet

$$(x_c)_i^{(j)} \approx \underline{S}_i^T \underline{C}_j \ .$$

Um die weiteren möglichen Interpretationen zu diskutieren, nehmen wir an, dass $m = 2$ sei. Dann gilt die Gleichung exakt. Für höhere m – erst dann bringt die Hauptkomponenten-Darstellung je etwas – werden die Interpretationen noch näherungsweise richtig sein.

Für $m = 2$ ist $\boldsymbol{C}$ die ganze Matrix $\widehat{\boldsymbol{B}}^{-1}$ der Hauptkomponenten-Analyse, und diese ist orthogonal. Der Vektor $\underline{C}_j$ ist die $j$te Zeile von $\boldsymbol{B}^{-1}$, und deshalb ist seine Länge gleich 1. In diesem Fall wird $\underline{S}_i^T \underline{C}_j$ gleich der Projektion von $\underline{S}_i$ auf $\underline{C}_j$. Also folgt:

Projiziert man einen Punkt i auf einen Pfeil j, so gibt die Distanz des projizierten Punktes zum Nullpunkt approximativ den (zentrierten) Wert der $j$ten Variablen für die $i$te Beobachtung wieder.

▷ Projiziert man den Punkt mit der Bezeichnung 38 (unterhalb des Zentrums) im Beispiel auf den Pfeil ”Nardstri”, so erhält man 1.19 Einheiten. Die standardisierte Anzahl Nardus stricta beträgt für diese Beobachtung 1.13. Für Beobachtung 47 erhält man 1.93 statt 2.00. Die Variable Carelepo (Carex leporina) ist schlechter repräsentiert. Für Beobachtung 42 erhält man -0.70 statt -0.36, Beobachtung 47 ergibt hier 1.54 statt 0.58. ◁

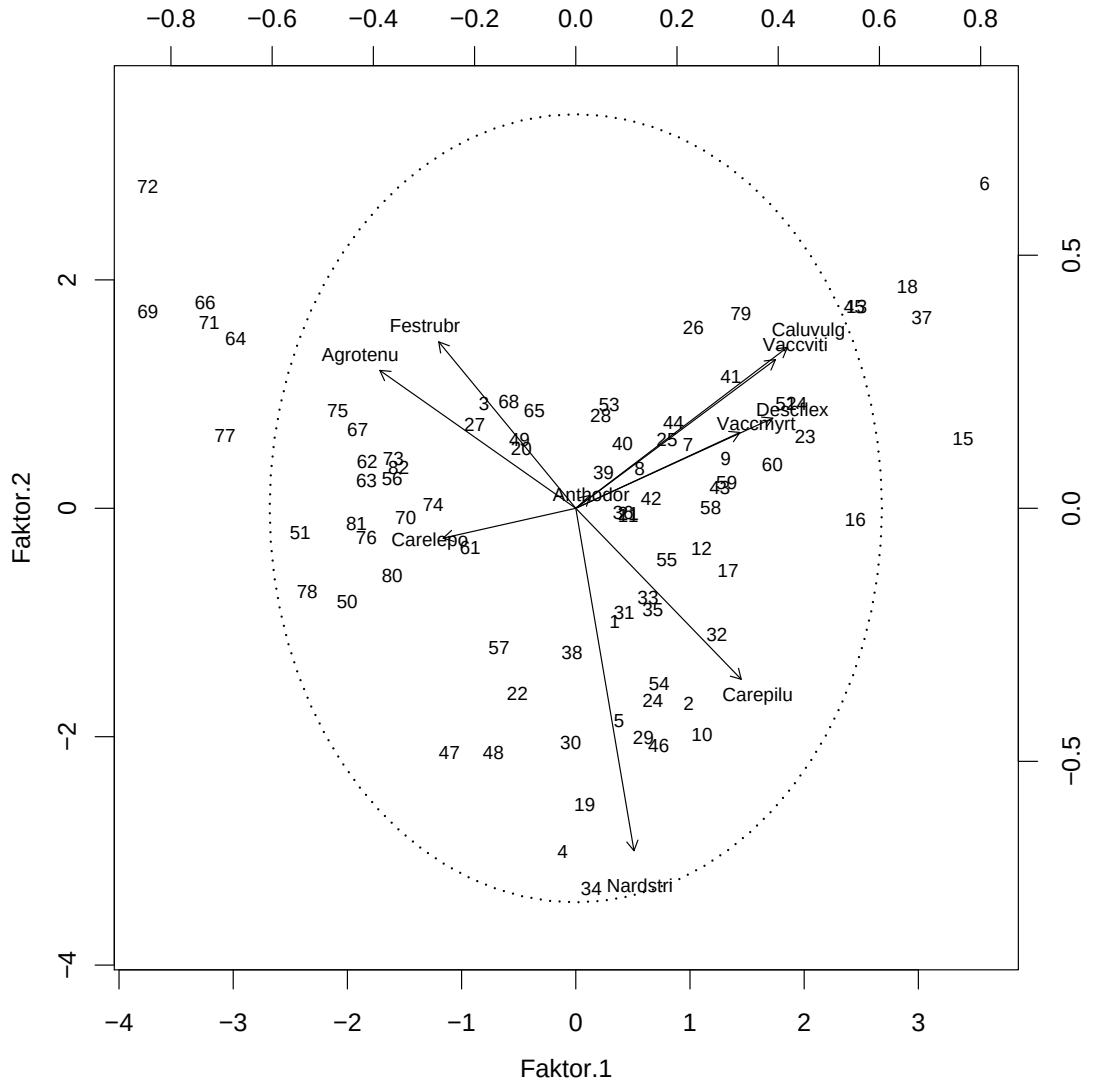

Abbildung 7.2.b: Biplot im Beispiel der Vegetationsstudie, Variante RMP. Die Variablen wurden vorgängig standardisiert.

c Für $m=2$ lässt sich ausrechnen, dass $\widehat{\mathrm{var}}\langle X^{(j)}\rangle = \sum_k \widehat{\lambda}_k \big(C_k^{(j)}\big)^2$ gilt. Wenn man die Variablen $X^{(j)}$ auf Varianz 1 standardisiert hat, dann müssen die Pfeilspitzen auf einer Ellipse liegen, deren Halbachsen $1/\sqrt{\widehat{\lambda}_k}$ messen ($k=1,2$) und in Richtung der Darstellungsachsen zeigen.

Allgemein gilt: Wie präzise die Variablen in der Figur dargestellt sind, lässt sich im Fall von standardisierten X-Variablen durch den Vergleich der Pfeillänge mit der genannten Ellipse ersehen.

Leider wird die Ellipse von den üblichen Programmen nicht eingezeichnet, und damit sind die Pfeile für sich selber genommen schlecht interpretierbar.

d Schliesslich lässt sich auch die Korrelation zwischen Variablen aus der Darstellung ablesen. Denkt man sich die Figur in vertikaler Richtung so gestaucht, dass die Ellipse zum Kreis wird, dann gibt der Winkel zwischen den „gestauchten“ Pfeilen die Korrelation (näherungsweise) wieder.

▷ Im Beispiel bilden gemäss Darstellung die Arten Caluvulg, Vaccviti, Vaccmyrt und Descflex eine Gruppe von hoch korrelierten Variablen. Wenn man die Korrelationen nachrechnet, merkt man allerdings, dass die Korrelationen innerhalb dieser Gruppe von 0.25 bis 0.53 variieren, während die Winkel praktisch 0 und 12^o messen, was Korrelationen von 1 respektive 0.97 entwpräche! Mindestens in diesem Beispiel sind die Winkel also schlecht interpretierbar. ◁

e **Was der Biplot zeigt.** Zusammenfassend zeigen die Punkte $\underline{S}_i$ und Pfeile $\underline{C}_j$ – exakt, falls $m = 2$ ist, und näherungsweise für den Fall $m > 2$ – folgendes:

(a) Die Punkte zeigen die ersten beiden Hauptkomponenten.

(b) Wenn die Variablen auf Varianz 1 standardisiert wurden, entspricht das Verhältnis Pfeillänge/Ellipse dem Verhältnis dargestellte / gesamte Varianz (=1) der Variablen $X^{(j)}$.

(c) Der Cosinus des Zwischenwinkels zwischen zwei Pfeilen zeigt die Korrelation der zugehörigen Variablen.

(d) Die Projektion der Punkte i auf die Richtung eines Pfeils $\underline{C}_j$ gibt die (zentrierten) ursprünglichen Beobachtungen $x_i^{(j)} - \bar{x}^{(j)}$ wieder.

(e) Die Distanzen zwischen Punkten $\|\underline{S}_h - \underline{S}_i\|$ entsprechen den Distanzen zwischen den Beobachtungen, $\|\underline{x}_h - \underline{x}_i\|$.

f Es gibt Varianten des Biplots. Die üblichere stellt die Pfeile so dar, wie sie nach der gerade erwähnten „Stauchung“ herauskommen. Dann sind die Winkel zwischen Pfeilen also direkt mit den Korrelationen verbunden. Damit auch die Projektion von Punkten auf die Pfeilrichtungen noch die gleiche Interpretation zulässt, muss man dann die Punkte „umgekehrt transformieren“, also in vertikaler Richtung strecken. Das bedeutet, dass man statt der Hauptkomponenten $\underline{Z}_i$ die standardisierten Hauptkomponenten für die Zeichnung der Punkte benützt. Die Abstände zwischen Punkten sind dann nicht mehr gleich den Abständen der Beobachtungsvektoren für $m = 2$.

Diese Variante des Biplot läuft unter dem Namen „column metric preserving (CMP)“, während die vorhergehende den Namen „row metric preserving (RMP)“ trägt.

7.S S-Funktionen

a Eine Hauptkomponenten-Analyse wird mit der Funktion

```
> princomp(x, cor = FALSE, scores = TRUE)
```

durchgeführt.

Argument `x`: Daten als Matrix oder als Dataframe.

Argument `cor`: Wenn `TRUE`, wird die Analyse für standardisierte Daten – oder auf Grund der Korrelationsmatrix – durchgeführt, sonst mit der Kovarianzmatrix.
Bemerkung: Die Standardisierung wird mit einer unüblichen Variante der Standardabweichung durchgeführt: Die Quadratsumme wird durch n statt durch $n-1$ dividiert. Dadurch entstehen (meist kleine) Abweichungen bei den ausgewiesenen Eigenwerten gegenüber anderen Programmen.

Das Ergebnis von `princomp`() kann mit `summary`() ausgedruckt werden. Dabei wird für jede Hauptkomponente deren Standardabweichung, der Anteil und der kumulierte Anteil an der totalen Varianz ausgegeben. Wird in `summary` das Argument `loadings=` TRUE gesetzt, werden zusätzlich die so genannten Loadings ausgegeben. Diese kann man aber auch separat mit der Funktion `loadings` erhalten.

b **Grafische Ausgabe.** `plot(t.r)` stellt den so genannten scree plot dar, den man auch über die Funktion `screeplot(t.r)` erhält.

Das Ziel einer Hauptkomponenten-Analyse ist ja oft die grafische Darstellung der ersten paar Komponenten. Die entsprechenden Koordinaten findet man unter `$scores`,

```
> t.r <- princomp(d.abst,cor=TRUE)
> plot(t.r)
> pairs(t.r$scores[,1:3],
  panel=function(x,y) text(x,y,row.names(d.abst)))
```

c Es gibt noch eine ältere Funktion `prcomp` zur Hauptkomponenten-Analyse. Sie berechnet die Hauptkomponenten aus der Kovarianzmatrix. Will man die Korrelationsmatrix verwenden, müssen die Daten zuerst standardisiert werden. Dazu kann man die Funktion `scale` verwenden, welche die übliche Version der Standardabweichung benützt. `prcomp` erlaubt keine Verwendung von Formeln und passt deshalb nicht mehr in die neueren Konzepte der S-Sprache.

d **Faktor-Analyse.** Die entsprechende Funktion heisst `factanal`.

e **Biplot.** Die R-Funktion zur Erzeugung eines Biplot heisst auch so: `biplot`. Sie wird meistens auf das Resultat von `princomp` – oder auch von `factanal` – angewandt.
`> biplot(princomp(iris[1:50,1:4],cor=TRUE))`

Die im Skript erwähnte Funktion, die die Referenz-Ellipse zeichnet und sich in gewissen Details von `biplot` unterscheidet, heisst `g.biplot` und wird auf der Website zur Verfügung gestellt.