

Ordination

Objekttyp: **Chapter**

Zeitschrift: **Veröffentlichungen des Geobotanischen Institutes der Eidg. Tech. Hochschule, Stiftung Rübel, in Zürich**

Band (Jahr): **90 (1986)**

PDF erstellt am: **25.08.2024**

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

7. Ordination

7.1 Vorbemerkungen

Die Schule Braun-Blanquet kennt die Ordination als Analysemethode nicht. Sie bedient sich direkter Schlussfolgerungen im Felde und der sogenannten Tabellenarbeit (ELLENBERG 1956, MUELLER-DOMBOIS und ELLENBERG 1974). Letztere ist ein Klassifikationsprozess innerhalb der Arten und innerhalb der Aufnahmen. Die auf reinen Vegetationsdaten beruhende Ordination ("indirekte Ordination", WHITTAKER 1967) wurde unseres Wissens durch GOODALL (1954) eingeführt. Er bediente sich der Hauptkomponentenanalyse, einer für damalige Verhältnisse sehr aufwendigen Methode. BRAY und CURTIS (1957) publizierten wenig später ein neu entwickeltes Ordinationsverfahren, das in der Folge unter dem Namen Polare Ordination weite Verbreitung fand. Zahlreiche konzeptionelle Fehler führten dazu, dass bald Revisionen vorgeschlagen wurden, welche dessen grösste Schwächen beheben sollten (ORLOCI 1966, SWAN et al. 1969, VAN DER MAAREL 1969). Leider ist jede dieser Verbesserungen auch mit einem zusätzlichen Rechenaufwand verbunden. Bei der Konzeption des vorliegenden Kapitels wurde geprüft, ob sich die Methode Bray und Curtis als grundlegendes, leicht diskutierbares Näherungsverfahren verwenden liesse. Zahl und Schwere der dabei in Kauf zu nehmenden Fehler sind aber nicht akzeptabel, sodass davon Abstand genommen werden muss. Stattdessen wird zuerst die Hauptkomponentenanalyse erörtert. Deren Durchführung ist auch bei kleineren Datensätzen rechenintensiv. Auf Beweisführungen für die Lösungswege wird erneut verzichtet und auf die einschlägige Literatur verwiesen. Der Nachvollzug des Rechenganges und die Interpretation der Ergebnisse dürften es einem Pflanzenökologen ermöglichen, Nutzen und Grenzen der Ordination unter den in der Vegetationskunde gegebenen Umständen zu erkennen.

Eine weitere Bemerkung gilt der Vielfalt der Methoden. Sie ist, sofern man auf die publizierten Namen abstellt, recht gross (ORLOCI 1978). Wie später zu zeigen ist, bleiben die Funktionen aller Ordinationsmethoden stets dieselben: Es

geht um die Abbildung einer vieldimensionalen Aehnlichkeitsstruktur in möglichst wenig Dimensionen. Im Gegensatz etwa zu den Gruppierungsverfahren findet man, dass Ordinationsmethoden, welche sich nur durch das verwendete Aehnlichkeitsmass unterscheiden, öfters unterschiedliche Namen tragen.

Gute Ordinationsmethoden sind mit einem hohen Rechenaufwand verbunden. Wohl bei keiner andern Gruppe von Methoden erscheinen in der pflanzensoziologischen Literatur so viele Arbeiten, welche sich anhand echter oder konstruierter Daten dem Vergleich und der Bewertung verschiedener Algorithmen annehmen (DEL MORAL 1980, GAUCH 1982). Dabei wird das Resultat in der Regel an den Erwartungen des Experimentators gemessen. Diese sind leider oft unrealistisch, wie z.B. FEOLI und FEOLI-CHIAPELLA (1980) treffend nachweisen.

Schliesslich muss betont werden, dass gute Ordinationen mehr leisten als nur eine grafische Darstellung von Beziehungen zwischen Arten oder Aufnahmen. Eines der fundamentalsten Probleme der Analyse ökologischer Daten besteht darin, dass zahlreiche gleichzeitig messbare Faktoren in unterschiedlichem Masse miteinander korreliert sind. Damit ist ihre separate Interpretation nicht sinnvoll. Methoden, wie z.B. die Hauptkomponentenanalyse, liefern eine Lösung zur Entkopplung simultan wirkender Faktoren. Eine pflanzenökologische Analyse, die diesen Schritt umgeht, ist deshalb kaum denkbar.

In der pflanzenökologischen Literatur wird zwischen direkter und indirekter Ordination unterschieden (WHITTAKER 1967, GAUCH 1982). Bei der direkten Ordination geht es darum, Aufnahmen in einem zweidimensionalen Koordinatensystem so darzustellen, dass ihre Aehnlichkeiten bezüglich zweier Standortfaktoren ersichtlich werden. Als Achsen dienen deshalb Messwerte wie Höhenlage, pH, Bodenfeuchte usw. Auch ein Oekogramm ist in diesem Sinne eine Ordination (vgl. ELLENBERG 1978), worin statt einzelner Aufnahmen die Standortbereiche ganzer Pflanzengesellschaften aufgezeichnet werden. Das vorliegende Kapitel beschäftigt sich jedoch

ausschliesslich mit der indirekten Ordination. Deren Prinzip wird im folgenden erläutert.

7.2 Prinzip der Ordinationsmethoden

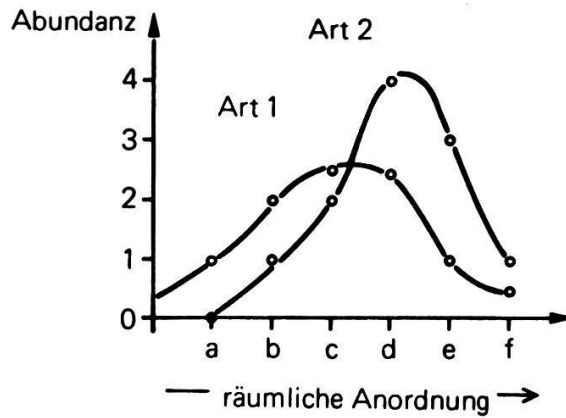
Das Prinzip der Ordination kann an einem Beispiel veranschaulicht werden, das einen Vegetationsgradienten darstellt. Abb. 7.1, A, zeigt zunächst die konstruierten Daten als Vegetationstabelle, der Einfachheit halber mit nur zwei Arten. Nehmen wir an, dass die Aufnahmen in gleichen Abständen entlang eines Transsektes erhoben wurden, so lassen sich die Ergebnisse wie in Abb. 7.1, B, darstellen. Daraus wird klar ersichtlich, dass es sich um einen einfachen, unverzweigten Vegetationsgradienten handelt. Eine solche Gradientenstruktur eignet sich naturgemäss gut zur Darstellung als Ordination. Letztere ist in Abb. 7.1, C, zu sehen. Als Achsen werden die beiden Pflanzenarten, als Koordinaten deren Artmächtigkeiten verwendet. Dabei zeigt es sich, dass die Abfolge der Aufnahmen genau derjenigen im Felde entspricht, die Anordnung jedoch fast kreisförmig ist. Dieses Muster ist für Vegetationsgradienten typisch (FEOLI und FEOLI-CHIAPELLA 1980). Es muss betont werden, dass Abb. 7.1, C, sich durch einfache Darstellung der Daten in Abb. 7.1, A, und ohne Umrechnung ergibt. Oval- und kreisförmige Aehnlichkeitsmuster werden fälschlicherweise immer wieder als durch die Hauptkomponentenanalyse verursacht gedeutet. Letztere ist deshalb - zu unrecht - vielerorts wenig beliebt (VAN DER MAAREL 1980, GAUCH 1982). Dass sie die Aehnlichkeitsverhältnisse perfekt abzubilden erlaubt, wurde jedoch mehrfach nachgewiesen und wird auch im folgenden an einem Beispiel gezeigt.

Die Ordination in Abb. 7.1 ist leicht interpretierbar und bedarf für praktische Anwendungen kaum weiterer Bearbeitung. Geht man jedoch bei komplizierten Daten genau gleich vor, so ergeben sich zwei Probleme, für welche z.B. die Hauptkomponentenanalyse eine Lösung bietet. Das erste Problem ergibt sich aus der Anzahl der Dimensionen. Eine unverzerrte Darstellung der Aehnlichkeitsverhältnisse setzt voraus, dass ebensoviele Achsen berücksichtigt werden, wie der Datensatz

A

		Aufnahme					
		a	b	c	d	e	f
1		1	2	2,5	2,5	1	0,5
2		0	1	2	4	3	1

B



C

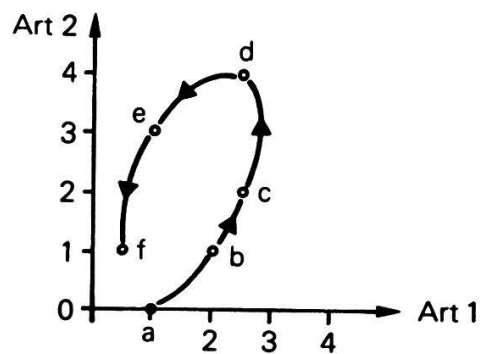


Abb. 7.1 Ein Vegetationsgradient a bis f, der von zwei Arten gebildet wird. Er ist dargestellt als Vegetationstabelle (A), als Transsekte (B) und als Ordination (C).

Arten enthält. Da aber mehr als drei Dimensionen grafisch nur schwer gleichzeitig darstellbar sind, ergibt sich die Forderung nach einer Verminderung der Achsen. Es ist klar, dass nicht jede Art gleichermassen guten Einblick in die Datenstruktur gewährt. Die meisten Ordinationsverfahren bestimmen die Achsen so, dass die Aehnlichkeitsverhältnisse möglichst unverändert wiedergegeben werden. Dabei fallen optimale Achsen selten genau mit einer einzelnen Art zusammen, sondern liegen irgendwo zwischen mehreren, aussagekräftigen Arten.

Das zweite Problem liegt in der Beziehung zwischen den Achsen. Die Punkte in Abb. 7.1, C sind deutlich korreliert. Die zweidimensionale, rechtwinklige Darstellung ist einigermaßen irreführend. Art 2 enthält teilweise die gleiche Information wie Art 1 und ist somit als unabhängige Dimension ungeeignet. Die Hauptkomponentenverfahren bestimmen z.B. die Achsen so, dass sie völlig unkorreliert sind. Die resultierenden Koordinaten eignen sich daher besser zur Konstruktion rechtwinkliger Ordinationen.

7.3 Die Hauptkomponentenanalyse

Die Aehnlichkeitsverhältnisse zwischen Aufnahmen sind durch die Daten, eventuelle Transformationen und das verwendete Aehnlichkeitsmass abschliessend festgelegt. Aufgabe der Hauptkomponentenanalyse ist es, von den ursprünglichen Variablen (Arten) ausgehend neue, unkorrelierte Achsen zu finden. Der Weg dazu führt ausschliesslich über eine Koordinatentransformation, die einigen speziellen Bedingungen zu genügen hat.

Zum besseren Verständnis der Vorgänge betrachten wir im folgenden zuerst das Ergebnis einer Koordinatentransformation. Davon ausgehend wird der Lösungsweg erörtert. In Abb. 7.2 ist das allgemeine geometrische Problem dargestellt.

Das ursprüngliche, rechtwinklige System (mit "Achsen") ist um einen bestimmten Winkel, φ , gekippt. Ein gegebener Punkt P soll seine Lage im Raum nicht verändern. Im alten System

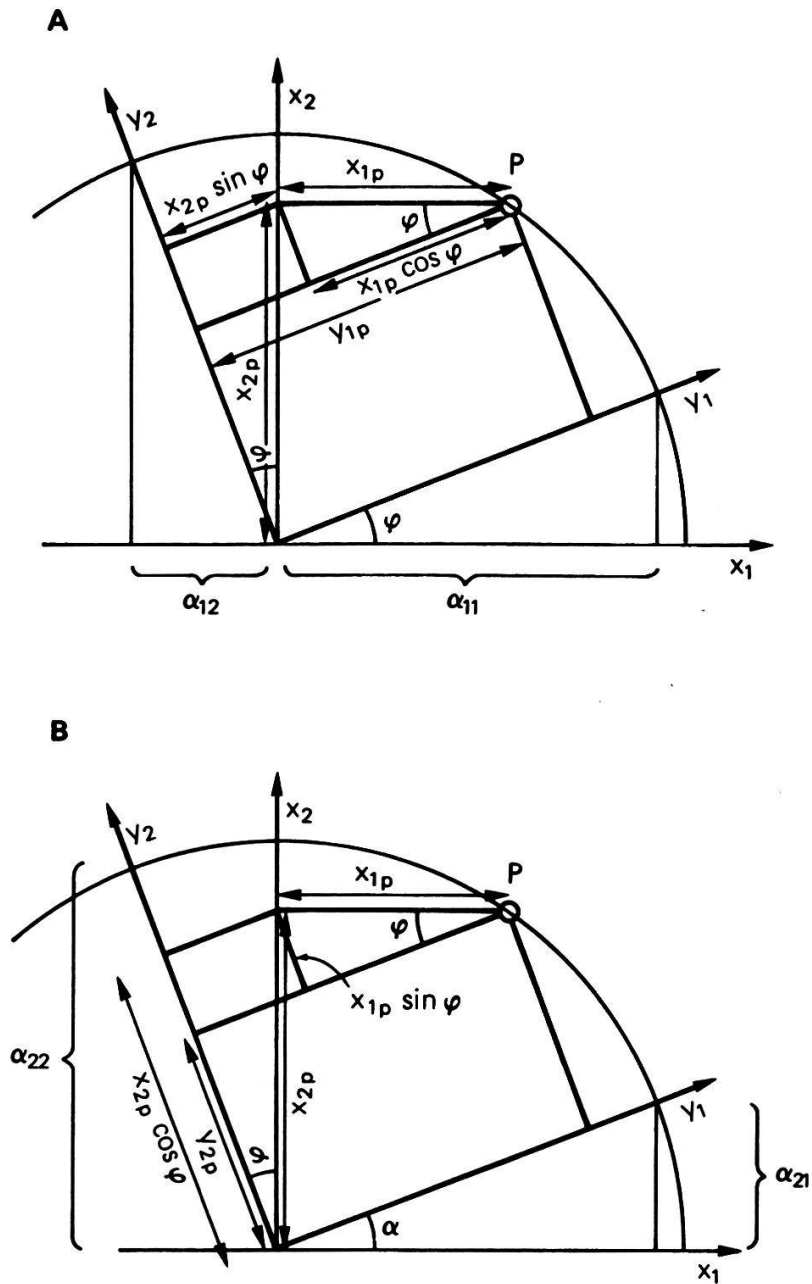


Abb. 7.2 Drehung eines Koordinatensystems mit den Achsen x_1 , x_2 . In A wird mit y_{1p} die erste neue, in B mit y_{2p} die zweite neue Koordinate des Punktes P bestimmt.

hat er die Koordinaten x_{1p} , x_{2p} . Diese sind im neuen System (mit "Faktoren") nicht mehr gültig, sondern sind durch y_{1p} , y_{2p} zu ersetzen. In Abb. 7.2, A ist dargestellt, wie die neuen Koordinaten aus den alten hergeleitet werden. Die Berechnung von y_{1p} zeigt Abb. 7.2, A:

$$y_{1p} = x_{1p} \cos \varphi + x_{2p} \sin \varphi$$

Die neue Koordinate setzt sich also aus zwei Teilstrecken (Summanden) zusammen. Dabei spielt natürlich auch der Drehwinkel φ eine Rolle. Um y_{1p} interpretieren zu können, muss die Bedeutung von $\cos \varphi$ und $\sin \varphi$ verstanden werden. In Abb. 7.2, A sind diese mit α_{11} und α_{12} bezeichnet. α_{11} erreicht den Maximalwert 1, wenn Achse 1 und Faktor 1 zusammenfallen. Der Minimalwert von 0 ergibt sich, wenn Achse 1 und Faktor 1 rechtwinklig zueinander stehen, also unabhängig sind. Die Vermutung, dass α_{11} ein Korrelationskoeffizient ist, lässt sich trigonometrisch leicht nachweisen (BATSCHLET 1982). Es ist auch leicht einzusehen, dass α_{12} die Korrelation von Achse 1 mit Faktor 2 ist. Die Berechnung der neuen Koordinate y_{1p} lässt sich deshalb folgendermassen formulieren:

$$y_{1p} = x_{1p} * (\text{Korrelation 1. Faktor mit 1. Achse}) + x_{2p} * (\text{Korrelation 2. Faktor mit 1. Achse})$$

Analog gestaltet sich nun die Berechnung der zweiten Koordinaten (Abb. 7.2, B). In obiger Schreibweise ergibt sich

$$y_{2p} = x_{1p} * (\text{Korrelation 1. Faktor mit 2. Achse}) + x_{2p} * (\text{Korrelation 2. Faktor mit 2. Achse})$$

Die a -Werte spielen bei der Interpretation der Hauptkomponentenanalyse eine zentrale Rolle. Im Falle vegetationskundlicher Daten entsprechen die ursprünglichen Achsen den Pflanzenarten. Die Matrix α , Matrix der Faktorenladungen genannt, hat folgenden Aufbau:

	Faktor 1	Faktor 2	Faktor 3...	Faktor p
Art 1	α_{11}	α_{12}	α_{13}	$\dots \alpha_{1p}$
Art 2	α_{21}	α_{22}	α_{23}	$\dots \alpha_{2p}$
Art 3	α_{31}	α_{32}	α_{33}	$\dots \alpha_{3p}$
.
Art p	α_{p1}	α_{p2}	α_{p3}	$\dots \alpha_{pp}$

Darin ist z.B. α_{32} die Korrelation der Art 3 mit der neuen Ordinationsachse 2. Es geht nun darum, die α -Matrix zu berechnen. Der Rechengang soll anhand des Beispiels in Abb. 7.1 nachvollzogen werden. Dieses umfasst 6 Aufnahmen und nur 2 Arten, sodass der Arbeitsaufwand bescheiden bleibt. Wird dagegen mit drei, vier oder mehr Dimensionen gearbeitet, so empfiehlt sich die Verwendung eines Computerprogrammes.

Gegeben ist also die Vegetationstabelle A:

Art	Aufnahme						Σx	\bar{x}
	1	2	3	4	5	6		
1	1	2	2.5	2.5	1	0.5	9.5	1.58
2	0	1	2	4	3	1	11	1.83

Diese Daten müssen für eine normale Hauptkomponentenanalyse bezüglich der Arten zentriert werden (Es gibt auch eine Variante der Hauptkomponentenanalyse, die auf die Zentrierung verzichtet!). Dazu ist von jeder Artmächtigkeit der Mittelwert des jeweiligen Artvektors zu subtrahieren. Die neue Tabelle, X, sieht wie folgt aus:

Art	Aufnahme						Σx^2	\bar{x}
	1	2	3	4	5	6		
1	-0.58	0.42	0.92	0.92	-0.58	-1.08	3.71	0
2	-1.83	-0.83	0.17	2.17	1.17	-0.83	10.83	0

Abb. 7.3 A,B, vergleicht die Ordinationen der Vegetationsabellen A und X. Es ergibt sich wie erwartet nur eine

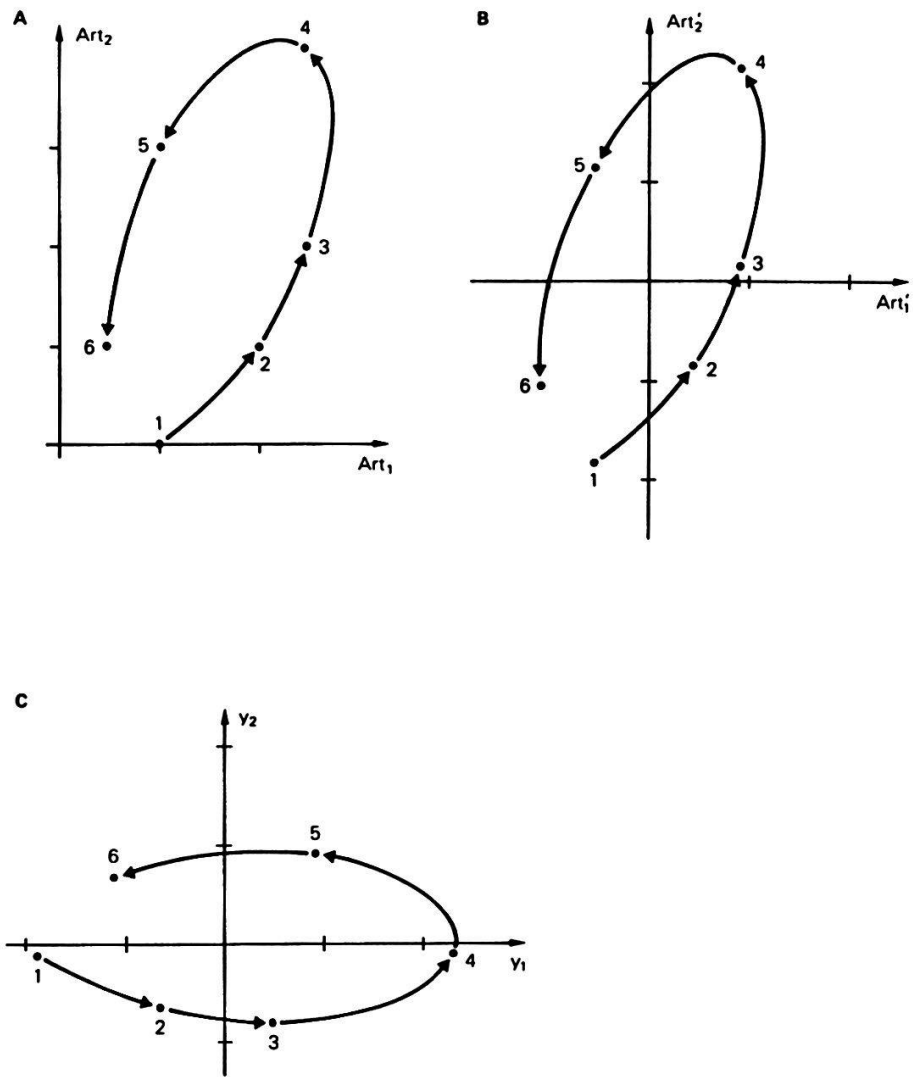


Abb. 7.3 Ordination von 6 Aufnahmen mit den Ausgangsdaten (A), den zentrierten Daten (B) und den Faktorenwerten einer Hauptkomponentenanalyse (C).

Verschiebung des Achsenkreuzes, ohne dass an den Aehnlichkeitsverhältnissen irgend etwas verändert würde. Als nächstes wird die Aehnlichkeitsmatrix der Arten der Vegetationstabelle A berechnet. Als Aehnlichkeitsmass dient das Skalarprodukt, doch könnten ebensogut die Kovarianz oder der Korrelationskoeffizient verwendet werden. Wir erhalten

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} 3.71 & 3.08 \\ 3.08 & 10.80 \end{bmatrix} .$$

Darin ist 3.71 das Skalarprodukt (die n-1 fache Varianz) der Art 1, 10.8 dasjenige der Art 2. Die beiden Arten besitzen auch einen deutlich positiven Zusammenhang von 3.08.

Für die weiteren Betrachtungen machen wir einen Sprung und studieren zunächst die Streuungsverhältnisse des Endresultates. Nach Beendigung der Hauptkomponentenanalyse werden die Arten durch Faktoren ersetzt sein. Die neue Tabelle Y wird so aussehen:

Faktor	Aufnahme					
	1	2	3	4	5	6
1	-1.92	-0.635	0.476	2.35	0.889	-1.16
2	-0.09	-0.68	-0.800	-0.10	0.95	0.72

In Abb. 7.3, C, ist die daraus konstruierbare Ordination dargestellt. Beim genauen Vergleich mit Abb. 7.3, A und B ergibt sich, dass die Beziehungen zwischen den Aufnahmen gleich geblieben sind. Es ist nur eine Veränderung festzustellen: Das Achsensystem wurde gedreht (oder, was dasselbe ist, der Punkteschwarm wurde gedreht). Zweck dieser Drehung ist es, die Achse 1 (Faktor 1) so zu legen, dass sie den Punkteschwarm möglichst gut auflöst. Ferner sind jetzt die Achsen 1 und 2 unkorreliert, was sich in der Vegetationstabelle Y leicht nachrechnen lässt. Die Faktorenladungen, welche zu diesem Resultat führten, sollen nun berechnet werden. Die Lösung erfolgt in zwei Schritten:

1. Berechnung der Skalarprodukte der neuen Koordinaten Y (Eigenwertbestimmung).
2. Berechnung der Faktorenladungen α (Eigenvektorbestimmung).

Aufgrund der Zielsetzung der Hauptkomponentenanalyse kann schon jetzt gesagt werden, dass die Aehnlichkeitsmatrix der neuen Ordinationskoordinaten wie folgt aussieht:

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Darin steht λ_1 für das Skalarprodukt der ersten Achse, λ_2 für dasjenige aller Werte der zweiten Achse. λ_1 wird als erster, λ_2 als zweiter Eigenwert bezeichnet (als Symbol dafür wird in der Literatur meist das griechische Lambda verwendet). λ_1 soll dabei so gross wie möglich sein, um die Auflösung des 1. Faktors maximal zu halten. Alle Elemente ausserhalb der Diagonalen sind gleich null, weil ja die Faktoren unkorreliert sein müssen. Eine weitere Einsicht hilft bei der Bestimmung der λ -Werte (Eigenwerte): Da die Aehnlichkeitsstruktur unverändert bleibt, ändert sich auch die Summe der Skalarprodukte nicht. Es gilt also

$$\lambda_1 + \lambda_2 = s_{11} + s_{22} = 3.71 + 10.8 = 14.51$$

Um die gesuchten Werte für λ zu erhalten, bedient man sich der Differentialrechnung (vgl. BATSCHELET 1982). Man findet, dass die sogenannte Determinante der Aehnlichkeitsmatrix null wird, wenn von jedem Diagonalelement die gesuchten Eigenwerte subtrahiert werden:

$$\begin{vmatrix} 3.71-\lambda & 3.08 \\ 3.08 & 10.8-\lambda \end{vmatrix} = 0$$

Der Vollständigkeit halber sei auch noch die Matrixschreibweise dieser Gleichung angeführt. Sie erscheint in allen einschlägigen Lehrbüchern und lautet $S - \lambda I = 0$. Darin ist I eine Einheitsmatrix, d.h. eine Matrix, deren Dia-

gonale mit 1, die übrigen Positionen mit 0 besetzt sind. Diese Gleichung ist nun nach λ zu lösen. Nach den Regeln der Matrizenrechnung kann sie wie folgt ausmultipliziert werden (BATSCHLET 1982):

$$(3.71-\lambda)(10.8-\lambda) - (3.08)(3.08) = 0 \text{ oder}$$

$$\lambda^2 - 6.2 \lambda + 9.2 = 0$$

Die quadratische Gleichung wird nach bekannter Vorschrift

$$\lambda = [-b \pm (b^2 - 4ac)^{1/2}] / 2a$$

gelöst. Wir erhalten

$$\begin{aligned} \lambda_1 &= (14.51 + (14.51^2 - (4)(30.58))^{1/2}) / 2 = 11.95 \\ \lambda_2 &= (14.51 - (14.51^2 - (4)(30.58))^{1/2}) / 2 = 2.56 \end{aligned}$$

Es kann gezeigt werden, dass 11.95 tatsächlich der maximal mögliche Wert für das Skalarprodukt des ersten Faktors ist, während λ_2 das Skalarprodukt des zweiten Faktors darstellt (BATSCHLET 1982). Auch ergibt sich mit $\lambda_1 + \lambda_2 = 14.51$ das Skalarprodukt der Ausgangsmatrix S.

Nun schreiten wir zur Bestimmung der Faktorenladungen α . Wir beginnen mit dem ersten Faktor und suchen α_{11} und α_{21} . Diese sind aus der alten und der neuen Kovarianzmatrix direkt bestimmbar (BATSCHLET 1982). Es gilt nämlich folgende Beziehung:

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \lambda_1 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} .$$

Setzen wir die schon bekannten Werte ein, so erhalten wir:

$$\begin{bmatrix} 3.71 & 3.08 \\ 3.08 & 10.8 \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = 11.95 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} .$$

Nach den Regeln der Matrixrechnung ausmultipliziert ergeben sich zwei Gleichungen mit den Unbekannten α_{11} und α_{21} :

$$\begin{aligned} 3.71 \alpha_{11} + 3.08 \alpha_{21} &= 11.95 \alpha_{11} \\ 3.08 \alpha_{11} + 10.8 \alpha_{21} &= 11.95 \alpha_{21} \end{aligned}$$

Die Auflösung ist nicht ganz trivial. Zunächst finden wir:

$$\begin{aligned} -8.24 \alpha_{11} + 3.08 \alpha_{21} &= 0 \\ 3.08 \alpha_{11} - 1.15 \alpha_{21} &= 0 \end{aligned}$$

Durch Addition der Gleichungen ergibt sich

$$-5.16 \alpha_{11} + 1.93 \alpha_{21} = 0 .$$

Daraus berechnet sich das Verhältnis von α_{11} zu α_{21} :

$$\alpha_{11} = \frac{1.93}{5.16} \alpha_{21} = 0.374 \alpha_{21}$$

Substituiert man z.B. α_{11} in einer der obigen Gleichungen, so erhält man nur die triviale Lösung $\alpha_{11} = \alpha_{21} = 0$. Aus unendlich vielen Möglichkeiten wählen wir eine aus, welche die Bedingung $\alpha_{11} = 0.374 \alpha_{21}$ erfüllt. Wir setzen kurzerhand und willkürlich

$$\begin{aligned} \alpha_{21} &= 1 \\ \alpha_{11} &= \frac{1.93}{5.16} \alpha_{21} = \frac{1.93}{5.16} = 0.374 \end{aligned}$$

Die a-Werte haben die besondere Eigenschaft, dass ihre absolute Grösse gar nicht eindeutig zu bestimmen ist. Festgelegt ist nur das Grössenverhältnis α_{11}/α_{21} . Es ist üblich, den Vektor

$$\begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \begin{bmatrix} 0.374 \\ 1.0 \end{bmatrix}$$

zu normalisieren. Dazu ist, wie früher gezeigt, jedes Element durch die euklidische Länge des Vektors zu dividieren:

$$\begin{bmatrix} \frac{0.374}{((0.374)^2 + 1^2)^{1/2}} \\ \frac{1}{((0.374)^2 + 1^2)^{1/2}} \end{bmatrix} = \begin{bmatrix} 0.350 \\ 0.937 \end{bmatrix}$$

Nun setzt man die Gleichungen für die Bestimmung von α_{12} und α_{22} auf:

$$\begin{aligned} 3.71 \alpha_{12} + 3.08 \alpha_{22} &= 2.56 \alpha_{12} \\ 3.08 \alpha_{12} + 10.8 \alpha_{22} &= 2.56 \alpha_{22} \end{aligned}$$

Es bleibt dem Leser überlassen, die Rechnung durchzuführen. Die normalisierten Ergebnisse lauten:

$$\begin{aligned} \alpha_{12} &= 0.937 \\ \alpha_{22} &= 0.350 \end{aligned}$$

Nachdem nun alle Faktorenladungen bekannt sind, schreiten wir zur Berechnung der neuen Koordinaten Y. Sie folgt der bereits untersuchten Regel

$$y_{1p} = x_{1p} \alpha_{11} + x_{2p} \alpha_{21}$$

Die aktuellen Werte sind:

$$\begin{aligned} y_{11} &= (-0.58)(0.35) + (-1.83)(0.937) = -1.92 \\ y_{12} &= (0.42)(0.35) + (-0.83)(0.937) = -0.63 \\ y_{13} &= (0.92)(0.35) + (0.17)(0.937) = 0.48 \\ \text{usw.} \end{aligned}$$

Damit ist die Hauptkomponentenanalyse abgeschlossen. Für die Interpretation fassen wir die Resultate zusammen. Wir erhalten folgende Eigenwerte:

$$\begin{aligned}\lambda_1 &= 11.95 = 82 \% \\ \lambda_2 &= 2.56 = 18 \% \\ \lambda_1 + \lambda_2 &= 14.51 = 100\%\end{aligned}$$

Die aus den neuen Koordinaten zu konstruierende Ordination (Abb. 7.3) weist entlang der 1. Achse 82 % der Unterschiedlichkeiten (Varianz) der Aufnahme aus, die 2. Achse nur 18 %.

Die Faktorenladungen sehen wie folgt aus:

Art	1	2
Faktor		
1	0.35	0.937
2	-0.937	0.35

Der neue Faktor 1 steht mit einem Korrelationskoeffizienten von 0.937 der alten Art 2 nahe. Entsprechend liegt Faktor 2 in der Nähe der alten Art 1, ist aber gegenläufig zu dieser (-0.937).

Schliesslich erhielten wir die neuen Faktorenwerte:

Faktor	Aufnahme					
	1	2	3	4	5	6
1	-1.92	-0.635	0.476	2.35	0.889	-1.16
2	-0.09	-0.68	-0.80	-0.10	0.95	0.72

Ohne die Ordination zu konsultieren, können aus der Tabelle Aufnahmen herausgesucht werden, die besonders ähnlich sind. Dies ist dann der Fall, wenn alle Faktorenwerte (vor allem aber die ersten) etwa gleiche Grösse besitzen. In unserem Beispiel liegen jedoch alle Punkte deutlich auseinander, wie wir aus der bereits gezeichneten Ordination wissen (Abb. 7.3).

Eine Besonderheit pflanzensoziologischer Anwendungen der Hauptkomponentenanalyse ist, dass mit sehr vielen Variablen (Arten) gerechnet wird. Selbst nach einer Datenreduktion bleiben noch 50, 100 oder mehr Eigenwerte zu bestimmen. Es sind daher lange Rechenzeiten zu erwarten. Für die Interpretation stellt sich die Frage, wie viele Dimensionen überhaupt betrachtet werden sollen. Dazu gibt es Regeln, die für pflanzensoziologische Tabellen erfahrungsgemäss unzweckmässig sind (UEBERLA 1971). Im allgemeinen zeigt es sich, dass bei Tabellen der Grössenordnung 100 mal 100 mindestens 4, höchstens 6 bis 8 Dimensionen interpretiert werden sollten. Die damit erklärte Varianz kann um 15 bis 30% liegen. Mit diesem doch bescheidenen Anteil wird es aber meist möglich sein, einen guten Einblick in die Ähnlichkeitsstruktur der Aufnahmen zu erhalten. Der Rest geht auf Konto "Rauschen".

7.4 Q- und D-Technik

Die bereits gezeigte Hauptkomponentenanalyse wird auch als R-Analyse (R-Technik) bezeichnet. Der Begriff Technik besagt, dass es sich nicht eigentlich um verschiedene Methoden handelt, sondern lediglich um einen andere Wege zum selben Resultat.

Im vorigen Kapitel berechneten wir mit Hilfe der R- Technik Koordinaten für 6 Aufnahmen. Dazu benützten wir eine zweidimensionale Ähnlichkeitsmatrix der Arten. Stellen wir uns nun die Aufgabe, eine Ordination der Arten zu berechnen, so muss bei der R-Technik die sechsdimensionale Ähnlichkeitsmatrix der Aufnahmen analysiert werden. Das ist recht aufwendig, da insgesamt 6 Eigenwerte und 6 Eigenvektoren zu berechnen sind. Tun wir das anhand unseres Beispiels mit Hilfe eines Computerprogrammes, so sehen wir, dass fünf Eigenwerte gleich null sind: Die Ähnlichkeitsverhältnisse zweier Arten lassen sich nämlich zwangslos auf einer einzigen Achse darstellen. Der Rechenaufwand zur Bestimmung des einzigen brauchbaren Eigenwertes wäre also übermässig hoch.

Die Q-Technik bietet in diesem Fall die Möglichkeit, die Koordinaten der Arten direkt aus einer zweidimensionalen Artmatrix zu bestimmen. Identische Ergebnisse sind aber immer auch über eine R-Analyse zu erzielen.

Die Q-Technik wird so begonnen, als ob eine R-Ähnlichkeitsmatrix der Aufnahmen erstellt werden müsste. Dazu sind letztere zu zentrieren. Erinnern wir uns der Rohdaten A:

Aufn.	1	2	3	4	5	6
Art 1	1	2	2.5	2.5	1	0.5
Art 2	0	1	2	4	3	1

Die Mittelwerte der Aufnahmen \bar{x}_i , betragen

(0.5 1.5 2.25 3.25 2 0.75)

Nun wird von jeder Artmächtigkeit der Mittelwert der jeweiligen Aufnahme subtrahiert und es resultiert die zentrierte Tabelle B:

Aufn.	1	2	3	4	5	6
Art 1	0.5	0.5	-0.25	-0.75	-1	-0.25
Art 2	-0.5	-0.5	0.25	0.75	1	0.25

Obwohl nun also die Aufnahmen zentriert sind, wird das Skalarprodukt der Arten berechnet, die sogenannte Q- Matrix. Wir erhalten für die ersten beiden Elemente:

$$q_{11} = 0.5^2 + 0.5^2 - 0.25^2 - 0.75^2 - 1^2 - 0.25^2 = 2.1875$$
$$q_{12} = (0.5)(-0.5) + (0.5)(-0.5) + (-0.25)(0.25) + (0.75)(0.75) + (-1)(1) + (-0.25)(0.25) = -2.1875$$

Die gesamte Matrix ergibt

$$Q = \begin{pmatrix} 2.1875 & -2.1875 \\ -2.1875 & 2.1875 \end{pmatrix}$$

Nun bestimmt man die Eigenwerte wie bei der R-Technik. Für die ausgeschriebene Determinante erhält man:

$$(2.1875-\lambda)(2.1875-\lambda)-(-2.1875)(-2.1875) = 0$$

Nach den Multiplikationen ergibt sich

$$\lambda^2 - 4.375\lambda = 0$$

Wir dividieren durch λ und erhalten die einzige von null verschiedene Lösung

$$\lambda_1 = 4.375$$

Die Eigenvektoren sind unter diesen Umständen rasch zu bestimmen. Zu lösen ist die Gleichung

$$\begin{bmatrix} 2.1875 & -2.1875 \\ -2.1875 & 2.1875 \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = 4.375 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix}$$

Daraus ergeben sich die zwei Gleichungen

$$\begin{aligned} 2.1875\alpha_{11} - 2.1875\alpha_{21} &= 4.375\alpha_{11} \\ -2.1875\alpha_{11} + 2.1875\alpha_{21} &= 4.375\alpha_{21} \end{aligned}$$

Setzt man $\alpha_{11} = 1$ und löst die obere Gleichung nach α_{21} auf, so ergibt sich

$$\begin{aligned} -2.1875\alpha_{21} &= 2.1875 \\ \alpha_{21} &= -1 \end{aligned}$$

Auch mit $\lambda_2 = 0$ kann ein Eigenvektor bestimmt werden. Es ist wiederum zweckmässig, $\alpha_{12} = 1$ zu setzen. Für α_{22} erhält man dann -1. Zusammengefasst ergibt sich für die Faktorenladungen

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

Normalisiert betragen sie

$$\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & -0.707 \end{bmatrix}$$

Im Gegensatz zur R-Technik ist nun noch jeder Vektor der Grösse des Eigenwertes anzupassen:

$$\alpha_{11}^2 + \alpha_{21}^2 = \lambda_1 = 4.375$$

Wir multiplizieren alle aktuellen Werte mit $\lambda^{1/2}$ und erhalten die gesuchten Koordinaten

$$\alpha'_{11} = y_{11} = 1.479$$

$$\alpha'_{21} = y_{21} = 1.479$$

Für die zweite Achse betragen die Koordinaten

$$y_{12} = y_{22} = 0.$$

Das Beispiel hat deutlich gemacht, dass sich mit der Q-Technik unter Umständen der Rechenaufwand beträchtlich reduzieren lässt. Gelegentlich kommt es auch vor, dass statt einer Q-Matrix bereits eine Distanzmatrix verfügbar ist. OR-LOCI (1973, 1978) hat gezeigt, dass sich aus dieser, nach entsprechender Transformation, dieselben Eigenwerte und Eigenvektoren extrahieren lassen. Auch unser Beispiel eignet sich für die D-Technik. Ob die Ausgangsdaten zentriert sind oder nicht, spielt dabei keine Rolle. Wir erhalten:

$$D^2 = \begin{bmatrix} 0 & 8.75 \\ -8.75 & 0 \end{bmatrix}$$

Die d^2 -Distanzen sind nun in $-0.5d^2$ - Ähnlichkeitskoeffizienten umzuwandeln:

in spezifischer Weise transformiert. Schliesslich wird eine kombinierte R- und Q-Hauptkomponentenanalyse durchgeführt. Aus grundsätzlichen Ueberlegungen heraus dürften streng genommen nur Zähl- und Frequenzdaten in dieser Weise analysiert werden. Dass trotzdem Messdaten verwendet werden, ist in der Praxis die Regel.

Der Rechengang könnte nun ohne weiteres am Beispiel des vorigen Kapitels gezeigt werden. Dies wäre mit Nachteilen verbunden, wie in Kapitel 7.6 noch zu sehen sein wird. Deshalb bedienen wir uns eines Datensatzes mit 4 Aufnahmen und 3 Arten, der trotz seiner Kleinheit für Vegetationsaufnahmen typisch ist.

Diese Vegetationstabelle (Tabelle 7.1) wird als Kontingenztafel aufgefasst. Wir verwenden Notierungen, wie sie in Tab. 7.1, A, dargestellt sind. $f_{.2}$ steht z.B. für die Summe aller Artmächtigkeiten der Aufnahme 2, $f_{.3}$ für die Summe innerhalb der Art 3, usw. Die Rohdaten, in Tabelle 7.1, B, dargestellt, werden nun schrittweise transformiert, und zwar zunächst zu relativen Häufigkeiten bezüglich der Gesamtsumme $f_{..}$:

$$p_{ij} = f_{ij} / f_{..}$$

Im Zahlenbeispiel erhält man dadurch eine neue Matrix P, wie sie aus Tab. 7.2, A, ersichtlich ist.

Nun wird die Annahme getroffen, dass die Randsummen $f_{.j}$ und $f_{i.}$ bzw. $p_{.j}$ und $p_{i.}$ für den gesamten Datensatz charakteristisch sind. Jede Artmächtigkeit ist deshalb nur bedeutsam, wenn sie von der charakteristischen (d.h. zu erwartenden) Häufigkeit abweicht. Es sind also Erwartungswerte zu berechnen. Diese ergeben sich aus der relativen Häufigkeit einer Art (bezüglich des Gesamtdatensatzes) und der relativen Häufigkeit der Individuen innerhalb einer Aufnahme:

$$e_{ij} = (p_{.j})(p_{i.})$$

Für die erste Artmächtigkeit erhalten wir

$$-0.5D^2 = \begin{bmatrix} 0 & -4.375 \\ -4.375 & 0 \end{bmatrix}$$

Die Eigenwerte berechnen sich nach bewährtem Schema:

$$\begin{aligned} (0-\lambda)(0-\lambda) - 4.375^2 &= 0 \\ \lambda^2 &= 4.375^2 \\ \lambda &= 4.375 \end{aligned}$$

Das Ergebnis ist identisch mit demjenigen der Q-Analyse. Führt man die Analyse weiter, ergeben sich auch gleiche Eigenvektoren und schliesslich die gewünschten Koordinaten.

In der Praxis spielt die Q-Technik eine untergeordnete Rolle, denn es gibt Algorithmen, die nur die grössten Eigenwerte extrahieren und daher auch in ungünstigen Fällen einigermaßen effizient arbeiten. Die D-Technik erweist sich als interessante Alternative, wenn die übrigen Untersuchungen ohnehin auf einer Distanzmatrix basieren.

7.5 Korrespondenzanalyse

Die Korrespondenzanalyse ist ein Hauptkomponentenverfahren, das sich bei Pflanzenökologen grosser Beliebtheit erfreut. Insbesondere im französischen Sprachbereich hat sie sich als Standardmethode eingebürgert (BENZECRI 1969, PIELOU 1977). Die angelsächsische Welt ist durch die Arbeiten von HILL (1974, 1979a) auf ihre Nützlichkeit aufmerksam geworden. Während bei der gewöhnlichen Hauptkomponentenanalyse zwei Rechengänge erforderlich sind um Aufnahmen und Arten zu ordinieren, tut die Korrespondenzanalyse beides gleichzeitig. Mehr noch: Die beiden Koordinatensätze sind gleich skaliert, sodass sie übereinander projiziert gemeinsam interpretiert werden dürfen. Der Zusammenhang zwischen Arten und Aufnahmen ist so besonders leicht erkennbar.

Formal unterscheidet sich die Korrespondenzanalyse nur wenig von den Hauptkomponentenanalysen: Zunächst werden die Daten

$$e_{11} = (0.143)(0.286) = 0.041$$

Zusammengefasst ergeben sich die Erwartungswerte in Tab. 7.2, B. Die höchsten Erwartungswerte erzielt Aufnahme 2, die ja auch die reichhaltigste ist. Schliesslich ist die Abweichung der effektiven Häufigkeiten p_{ij} von den erwarteten e_{ij} zu berechnen:

$$d_{ij} = p_{ij} - e_{ij}$$

Für das erste Element erhalten wir

$$d_{11} = 0.143 - 0.041 = 0.102$$

Die vollständige D-Matrix ist in Tab. 7.2, C dargestellt. Auf Grund dieser Abweichungen von den erwarteten, relativen Häufigkeiten wird nun weitergerechnet. Es fällt sofort auf, dass sowohl die Artvektoren wie auch die Aufnahmen den Mittelwert null besitzen, d.h. dass sie bereits zentriert sind. Erinnern wir uns der Hauptkomponentenanalyse, so könnte sogleich die R- oder die Q-Technik angewandt werden. In der Tat werden beide gleichzeitig durchgeführt. Wir beginnen mit der Matrix der Skalarprodukte der Arten (mir der Matrix der Aufnahmen ergäbe sich genau dasselbe Resultat!). Wir erhalten für den Vergleich der Arten 1 und 2

$$s_{12} = (0.102)(-0.061) + (0.020)(-0.041) + (-0.081)(0.020) + (-0.041)(0.082) = 0.012$$

Die gesamte Matrix lautet:

$$S = \begin{array}{ccc} 0.0192 & -0.0121 & -0.0071 \\ -0.0121 & 0.0125 & -0.0004 \\ -0.0071 & -0.0004 & 0.0075 \end{array}$$

Wie bei der Hauptkomponentenanalyse werden die Eigenwerte aus der Determinantengleichung bestimmt:

Tabelle 7.1

Vegetationstabelle, die mit Hilfe der Korrespondenzanalyse untersucht werden soll. Formale Darstellung als Kontingenztafel (A) und Zahlenbeispiel (B).

A

Art/Aufn.	j=1	j=2	j=3	j=4	Σ
i=1	f_{11}	f_{12}	f_{13}	f_{14}	$f_{1.}$
i=2	f_{21}	f_{22}	f_{23}	f_{24}	$f_{2.}$
i=3	f_{31}	f_{32}	f_{33}	f_{34}	$f_{3.}$
Σ	$f_{.1}$	$f_{.2}$	$f_{.3}$	$f_{.4}$	$f_{..}$

F =

B

Art/Aufn.	1	2	3	4	Σ Arten
1	1	1			2
2		1	1	1	3
3		1	1		2
Σ Aufn.	1	3	2	1	Total = 7

Tabelle 7.2

Abgeleitete Tabellen des Zahlenbeispiels in Tab. 7.1, B. Relative Häufigkeiten (A), Erwartungswerte (B), Abweichungen von den Erwartungswerten (C).

A

	Art	Aufn. 1	2	3	4	Σ
P =	1	0.143	0.143	0.0	0.0	0.286
	2	0.0	0.143	0.143	0.143	0.429
	3	0.0	0.143	0.143	0.0	0.286
	Σ	0.143	0.429	0.286	0.143	1.0

B

	Art	Aufn. 1	2	3	4	Σ
E =	1	0.041	0.123	0.081	0.041	0.286
	2	0.061	0.184	0.123	0.061	0.429
	3	0.041	0.123	0.081	0.041	0.286
	Σ	0.143	0.429	0.286	0.143	1.0

C

	Art	Aufn. 1	2	3	4	Σ
D =	1	0.102	0.020	-0.081	-0.041	0
	2	-0.061	-0.041	0.020	0.082	0
	3	-0.041	0.020	0.061	-0.041	0
	Σ	0	0	0	0	0

$$\begin{vmatrix} 0.0192-\lambda & -0.0121 & -0.0071 \\ -0.0121 & 0.0125-\lambda & -0.0004 \\ -0.0071 & -0.0004 & 0.0075-\lambda \end{vmatrix} = 0$$

Multipliziert man diese Determinante aus, so erhält man eine kubische Gleichung dritten Grades. Schneller geht es mit einem Computerprogramm und wir finden

$$\begin{aligned} \lambda_1 &= 0.0297 \\ \lambda_2 &= 0.0094 \\ \lambda_3 &= 0 \end{aligned}$$

Nun ist für jeden Eigenwert ein Eigenvektor α (Faktorenladungen) zu finden. Das Gleichungssystem

$$\begin{bmatrix} 0.0192 & -0.0121 & -0.0071 \\ -0.0121 & 0.0125 & -0.0004 \\ -0.0071 & -0.0004 & 0.0075 \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \end{bmatrix} = 0.0297 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \end{bmatrix}$$

ist zu lösen. Normalisiert lauten die Ergebnisse

$$\begin{aligned} \alpha_{11} &= -0.79670 \\ \alpha_{21} &= 0.55311 \\ \alpha_{31} &= 0.24359 \end{aligned}$$

Analoge Berechnungen für $\lambda_2 = 0.0094$ ergeben

$$\begin{aligned} \alpha_{12} &= 0.17871 \\ \alpha_{22} &= 0.60061 \\ \alpha_{32} &= -0.77931 \end{aligned}$$

Die Ermittlung des dritten Eigenvektors ist auch möglich:

$$\begin{aligned} \alpha_{13} &= 0.57735 \\ \alpha_{23} &= 0.57735 \\ \alpha_{33} &= 0.57735 \end{aligned}$$

Nun geht es um die Bestimmung der neuen Aufnahmekoordinaten. Es gilt hier zu beachten, dass die alten Koordinaten nicht

unsere Rohwerte sind, sondern die transformierten Daten, die wir in der D-Matrix finden (Tab. 7.2, C):

Art	Aufnahme								
	1	2	3	4					
1	0.102	0.020	-0.081	-0.041	=	d_{11}	d_{12}	d_{13}	d_{14}
2	-0.061	-0.041	0.020	0.082		d_{21}	d_{22}	d_{23}	d_{24}
3	-0.041	0.020	0.061	-0.041		d_{31}	d_{32}	d_{33}	d_{34}

Das bewährte Rezept für die neuen Koordinaten Y lautet:

$$\begin{aligned}
 y_{1.\text{Achse},1.\text{Aufn.}} &= d_{11} \alpha_{11} + d_{21} \alpha_{21} + d_{31} \alpha_{31} + d_{41} \alpha_{41} \\
 &= (0.102)(-0.79670) + (-0.061)(0.55311) + (-0.041)(0.24359) \\
 &= -0.12510
 \end{aligned}$$

$$\begin{aligned}
 y_{1.\text{Achse},2.\text{Aufn.}} &= d_{12} \alpha_{11} + d_{22} \alpha_{21} + d_{32} \alpha_{31} + d_{42} \alpha_{42} \\
 &= (0.020)(-0.79670) + (-0.041)(0.55311) + (0.020)(0.24359) \\
 &= -0.03386 \\
 &\text{usw.}
 \end{aligned}$$

Für alle Aufnahmen und Achsen ergibt sich

Achse	Aufnahme			
	1	2	3	4
1	-0.12510	-0.03386	0.09124	0.06773
2	0.01327	-0.03677	-0.05004	0.07354
3	0.0	0.0	0.0	0.0

Der letzte Schritt der Korrespondenzanalyse ist die Ableitung der Koordinaten der Arten, die wir Z nennen wollen. Auch dieses Vorgehen ist nicht neu. Wir brauchen nämlich die Faktorenladungen α wie in der Q-Technik der Hauptkomponentenanalyse bloss mit den Eigenwerten λ zu skalieren:

$$z = \begin{matrix} -0.7967 \lambda_1^{1/2} & 0.17871 \lambda_2^{1/2} & -0.577 \lambda_3^{1/2} \\ 0.5531 \lambda_1^{1/2} & 0.60061 \lambda_2^{1/2} & -0.577 \lambda_3^{1/2} \\ 0.2436 \lambda_1^{1/2} & -0.77931 \lambda_2^{1/2} & -0.577 \lambda_3^{1/2} \end{matrix}$$

Ausmultipliziert finden wir

$$z = \begin{matrix} -0.13732 & 0.01736 & 0.0 \\ 0.09534 & 0.05836 & 0.0 \\ 0.04198 & -0.07572 & 0.0 \end{matrix}$$

Somit ist die Korrespondenzanalyse abgeschlossen. Die Y- und Z-Werte dienen der Ordination der Aufnahmen und Arten (Abb. 7.4). In unserem sehr einfachen Beispiel stehen sich verschiedene Punkte sehr nahe, was leicht zu interpretieren ist. Wir finden:

- Für Aufnahme 1 ist Art 1 typisch.
- Für Aufnahme 3 ist Art 3 typisch.
- Für Aufnahme 4 ist Art 2 typisch.

Das Resultat leuchtet sofort ein, wenn man es anhand der Rohdaten überprüft. Man findet, dass die Aufnahmen 1 bis 3 einen floristischen Gradienten repräsentieren, in den sich Aufnahme 4 nicht zwanglos einpassen lässt (Pfeil in Abb. 7.4). Es bestätigt sich, dass Aufnahme 2 nicht durch eine einzelne Art differenziert werden kann. Der Erklärungsgrad der beiden Ordinationsachsen ist ebenfalls klar:

$$\begin{aligned} \Sigma \lambda_i &= 0.0297 + 0.0094 = 0.0391 = 100\% \\ \lambda_1 &= 75.8\% \\ \lambda_2 &= 24.2\% \end{aligned}$$

Die x-Achse erklärt also 75.8% der Varianz zwischen den Ordinationspunkten, die y-Achse 24.2%. Die gesamte Punktekonfiguration ist in zwei Dimensionen verlustfrei darstellbar.

Die doppelte Zentrierung der Daten bringt es mit sich, dass

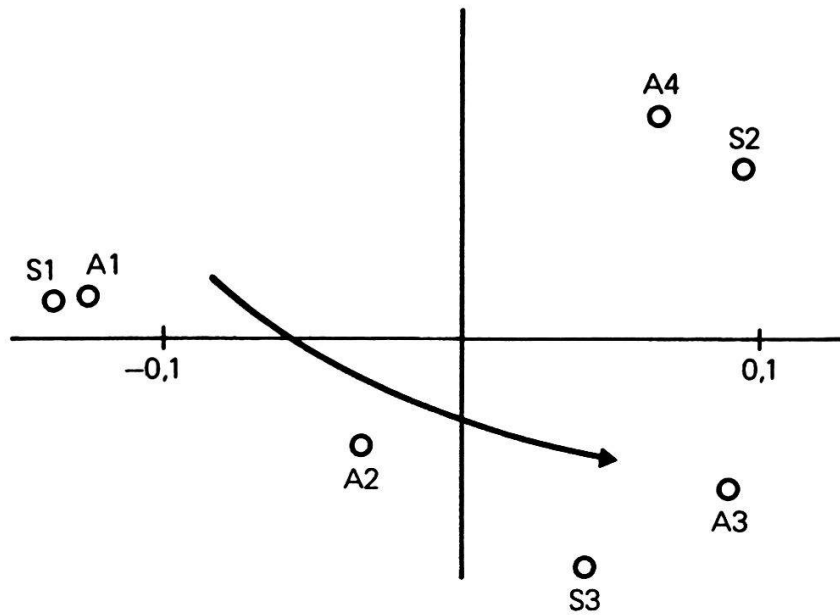


Abb. 7.4 Ordination als Resultat einer Korrespondenzanalyse. Mit A bezeichnete Punkte stehen für Aufnahmen, mit S bezeichnete für Arten.

die Aehnlichkeitsverhältnisse, mit denen die Korrespondenzanalyse operiert, deutlich von denjenigen der Rohdaten abweichen. Ein durchschnittlicher floristischer Gradient etwa präsentiert sich als Ordination nach einer Hauptkomponentenanalyse meist schlaufen- oder (dreidimensional) schraubenförmig (vgl. Abb. 7.3). Aus der Korrespondenzanalyse resultieren dagegen sanft bogenförmige Punktwolken, weshalb auch von einem Hufeiseneffekt gesprochen wird. HILL (1979a) entwickelte mit seiner Detrended Correspondence Analysis (DCA) ein Verfahren, das leicht gekrümmte Strukturen auf eine Achse projiziert. In vielen Fällen erzielt er damit eine lineare Darstellung eines Vegetationsgradienten.

7.6 Varianten der Hauptkomponentenanalyse

Mit dem bisher Behandelten sind die Hauptkomponentenverfahren bei weitem nicht erschöpft. Die Vorschläge für Aenderungen lassen sich unterteilen nach dem Analyseschritt, in welchem sie von der Hauptkomponentenanalyse abweichen:

Eine erste Gruppe von Verfahren bedient sich ausschliesslich einer besonderen Transformation der Rohdaten. Dem praktisch arbeitenden Pflanzenökologen bietet sich damit ein Spektrum leicht durchschaubarer, echter Alternativen. Wir werden im Folgenden zwei typische Beispiele genauer untersuchen.

Eine zweite Gruppe setzt bei den Aehnlichkeitsmassen an. Statt Skalarprodukt, Varianz oder Korrelationskoeffizient wird z.B. ein halb- oder nichtmetrisches Mass verwendet. Wir betrachten im folgenden den Effekt von VAN DER MAARELS (1979) similarity ratio. GOWER (1966,1967) zeigt dagegen einen Weg, wie ein beliebiges, nicht metrisches Mass für eine metrisch einwandfreie Ordination verwendet wird. Seine "Principle Coordinate Analysis" ist z.B. bei PIELOU (1977) im Detail beschrieben. Kennt man dagegen nur die Vegetationstypen der Aufnahmen und will ohne Berücksichtigung des Artenspektrums eine Ordination herleiten, so bietet sich die "Principle Axis Analysis" an (GOWER 1966, ORLOCI 1978).

Eine dritte Gruppe schliesslich versucht, die Nachteile der Linearität gewöhnlicher Hauptkomponentenanalysen zu umgehen. KRUSKAL (1964) erlaubt bei seiner "nichtmetrischen multidimensionalen Skalierung" die Verwendung beliebiger Aehnlichkeitsmasse und die Ordination versucht bloss, die relativen Positionen der Aufnahmen zu reproduzieren. Die Methode soll hier nicht weiter diskutiert werden.

In Anbetracht der Besonderheiten vegetationskundlicher Daten lohnt es sich, den Einflussmöglichkeiten mittels verschiedener Transformationen besondere Beachtung zu schenken. Bei der Hauptkomponentenanalyse sehen wir, dass Ordinationen typischer Vegetationsgradienten zu schlaufen-, kreis- oder spiralförmigen Strukturen führen können (Abb. 7.5 A). Dies ändert die früher besprochene Einheitslängentransformation der Aufnahmen. Ausgehend von den Rohdaten bestimmen wir die euklidische Länge der Aufnahmevektoren:

Art i	Aufnahme j					
	1	2	3	4	5	6
1	1	2	2.5	2.5	1	0.5
2	0	1	2	4	3	1
$l_j = [\sum_i a_{ij}^2]^{1/2}$	1	2.24	3.20	4.72	3.16	1.12

Durch Division jeder Artmächtigkeit mit der Länge des Aufnahmevektors erhalten wir die transformierten Werte a':

$$a'_{ij} = \frac{a_{ij}}{l_j}$$

Dies ergibt die neue Tabelle:

Art	Aufnahme					
	1	2	3	4	5	6
1	1	0.89	0.78	0.53	0.32	0.45
2	0	0.45	0.625	0.85	0.95	0.89

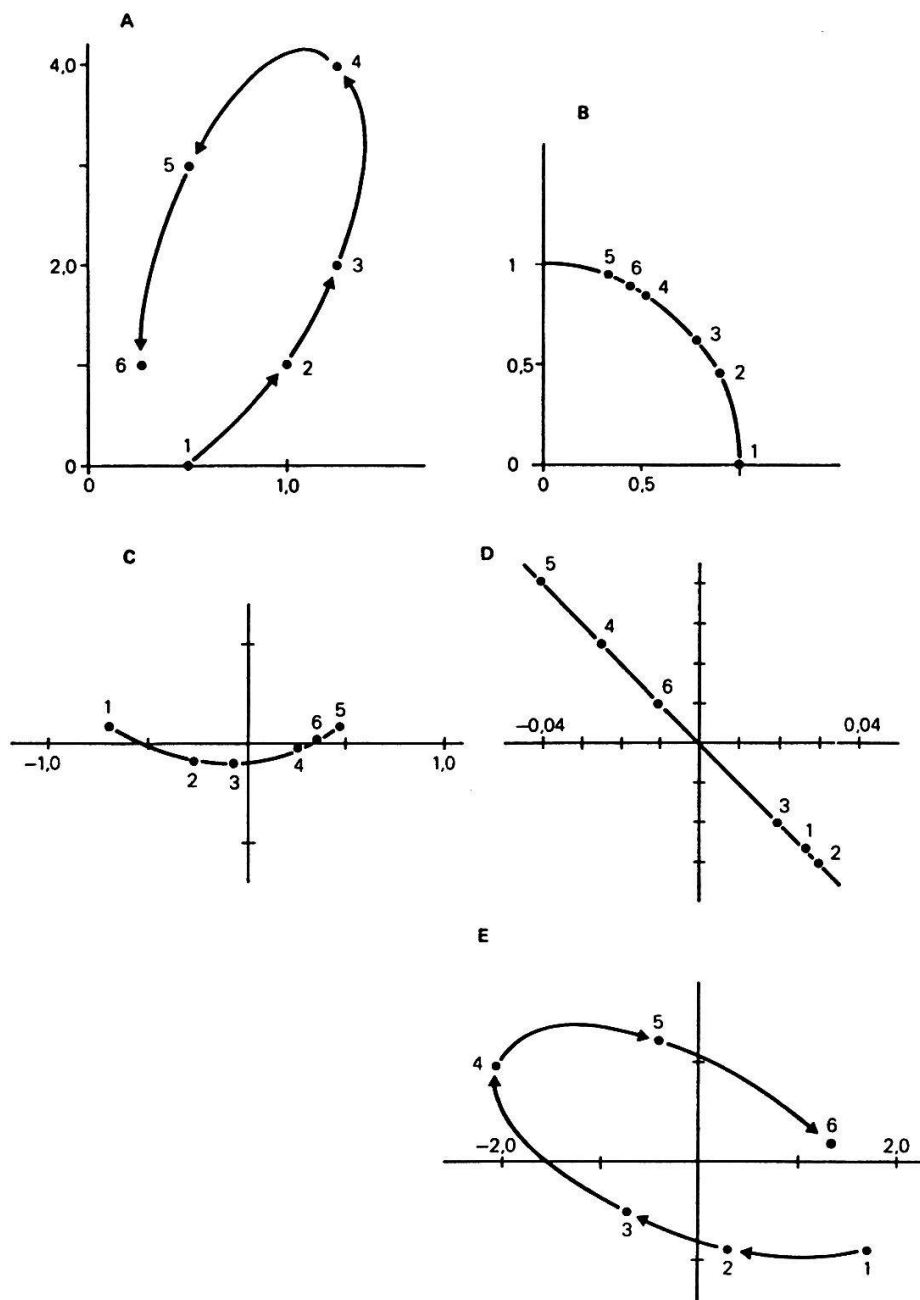


Abb. 7.5 Ordination als Resultat eines verschieden transformierten Datensatzes. A: Rohdaten. B: Aufnahmen auf Einheitslänge transformiert. C: Hauptkomponentenanalyse der auf Einheitslänge transformierten Aufnahmen. D: Transformation für eine Korrespondenzanalyse. E: Hauptkomponentenanalyse mit van der Maarels similarity ratio.

Aus diesen Werten ist die in Abb. 7.5, B gezeigte Ordination konstruiert. Die Einheitslängentransformation hat alle Aufnahmen auf ein Kreissegment projiziert! Die Ähnlichkeitsverhältnisse haben sich gegenüber den Rohdaten deutlich verändert. In der neuen Ordination sind nur noch die relativen Anteile der Arten einer Aufnahme für Vergleiche massgebend. Die Aufnahmen 4, 5 und 6 sind sich deshalb sehr nahe gekommen - sie besitzen alle hohe Werte bezüglich Art 2 und niedrige Werte bezüglich Art 1.

Aus Abbildung 7.5, A wird erneut klar, dass sich die Rohdaten nicht sinnvoll auf nur einer linearen Ordinationsachse darstellen lassen. Mit Abb. 7.5, B kommen wir einer solchen Möglichkeit schon näher. Unterziehen wir deshalb die transformierten Daten einer Hauptkomponentenanalyse. Wir erhalten

Faktor	Aufnahme					
	1	2	3	4	5	6
1	-0.706	-0.283	-0.073	0.256	0.463	0.342
2	0.098	-0.081	-0.095	-0.025	0.088	0.015

Das Resultat ist in Abb. 7.5, C als Ordination dargestellt. Der Kreisbogen ist nun so gedreht, dass er in Richtung der ersten Achse verläuft. Die zweite Achse ist für die Differenzierung der Aufnahmen kaum von Bedeutung. Diese Tatsache lässt sich auch an den Eigenwerten ablesen. Man findet nämlich

$$\lambda_1 = 0.981 = 96.7\%$$
$$\lambda_2 = 0.034 = 3.3\%$$

Eine je nach Datenstruktur noch tiefergreifendere Transformation haben wir bereits kennengelernt. Es sind die Abweichungen von den Erwartungswerten, welche in die Korrespondenzanalyse eingehen. Wie dort gezeigt, beginnen wir mit der Bildung der Randsummen der Aufnahmen und der Arten:

Art	Aufnahme						Σ Arten
	1	2	3	4	5	6	
1	1	2	2.5	2.5	1	0.5	9.5
2	0	1	2	4	3	1	11
Σ Aufn.	1	3	4.5	6.5	4	1.5	$\Sigma = 20.5$

Wir vollziehen nun alle bei der Korrespondenzanalyse gezeigten Operationen und erhalten

Art	Aufnahme					
	1	2	3	4	5	6
1	0.026	0.030	0.020	-0.025	-0.041	-0.010
2	-0.026	-0.030	-0.020	0.025	0.041	0.010

Abb. 7.5, D zeigt die entsprechende Ordination. Die Transformation hat die Aufnahmen auf eine Gerade projiziert! Generell reduziert nämlich die Korrespondenzanalyse die Dimensionalität um 1. Im Kapitel 7.5 musste aus diesem Grunde ein dreidimensionales Beispiel gewählt werden, sodass als Resultat zumindest noch eine zweidimensionale Ordination möglich war. Zu beachten ist die Reihenfolge der Aufnahmen auf der Ordinationsgerade. Sie dokumentiert, dass sich die Ähnlichkeitsverhältnisse grundlegend von denjenigen in den Rohdaten unterscheiden.

Im Kapitel Ähnlichkeitsmasse wurde gezeigt, dass jene öfters versteckt Transformationen beinhalten. Verwenden wir z.B. van der Maarels Koeffizient anstelle des Skalarproduktes, so werden die Daten bezüglich ihrer Ähnlichkeit anders beurteilt. Der Versuch ist in Abb. 7.5, E dargestellt. Wie leicht zu sehen ist, bleiben hier die Ähnlichkeitsverhältnisse der Rohdaten erhalten, doch liegen die Achsen anders als bei der normalen Hauptkomponentenanalyse. Van der Maarels similarity ratio ist ja kein metrisches Ähnlichkeitsmass. Die Hauptkomponentenanalyse interpretiert es jedoch metrisch. Man führt also formal

gesehen einen nicht kontrollierbaren, die Analyse meist nicht weiter störenden Fehler ein, der zu einem anderen Resultat führt. Ob ein solches Vorgehen akzeptabel ist, bleibe dem Urteil des Anwenders überlassen.