

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

Herausgeber: Geobotanisches Institut, Stiftung Rübel (Zürich)

Band: 90 (1986)

Artikel: Analyse vegetationskundlicher Daten : Theorie und Einsatz statistischer Methoden = Analysis of vegetation data : theory and application of numerical methods

Autor: Wildi, Otto

Kapitel: 9: Zusammenhang zwischen Vegetation und Standort

DOI: <https://doi.org/10.5169/seals-308832>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 31.01.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

9. Zusammenhang zwischen Vegetation und Standort

9.1 Zielsetzung und Schwierigkeiten

Mit der Analyse des Zusammenhanges zwischen Vegetation und Standort sind wir bei der schwierigsten und vielleicht wichtigsten Aufgabe der Pflanzenökologie angelangt (vgl. z.B. POORE 1955, S. 246). Dabei werden Lösungen gesucht auf Fragen, welche gleichermaßen von wissenschaftlichem und umweltpolitischem Interesse sind. Unsere Umwelt ist heute tiefgreifenden Veränderungen und Störungen ausgesetzt. Eingriffe können direkt auf die Vegetation erfolgen. Entwaldung, Schnitt, Weidegang oder die vollständige Entfernung der Pflanzendecke sind Beispiele. Ebenso häufig aber wird primär der Standort verändert. Grossflächig erfolgt dies bei landwirtschaftlicher Düngung, bei der Eutrophierung der Flüsse und Seen durch Abwässer oder durch saure Niederschläge. In jedem Falle geht die Aufgabe der Pflanzenökologie dahin, Veränderungen und mithin zukünftige Zustände vorauszusagen. Je nach der Art des Eingriffes sollen Aussagen nach folgenden Schemata gemacht werden:

1. Falls Standort A umgewandelt wird in Standort B, so folgt, dass Vegetation C sich umwandelt in Vegetation D.
2. Falls Vegetation C umgewandelt wird in Vegetation D, so folgt, dass Standort A sich umwandelt in Standort B.

Als Spezialfall eines Standortfaktors soll die Zeit mit in die Betrachtungen eingeschlossen sein. Voraussetzung für die postulierten Voraussagen ist die formale, im statistischen Sinne deskriptive Bestimmung des Zusammenhanges zwischen Vegetation und Standort. Natürlich kann man methodisch erheblich weiter gehen und die dynamischen Prozesse numerisch nachzubilden versuchen. Davon soll hier abgesehen werden.

Formal ergibt sich die Aufgabe, die durch mehrere Variablen charakterisierte Vegetation mit mehreren StandortsvARIABLEN

in Beziehung zu setzen. Gegeben sind also zwei multivariate Datensätze. Der Zusammenhang zwischen einem Vektor des einen Satzes (z.B. einer Art) und einem Vektor des andern Satzes (einem Standortsfaktor) kann durch eine normale Korrelationsanalyse gefunden werden. Für die multivariate Fragestellung muss zur Kanonischen Korrelationsanalyse gegriffen werden (PIELOU 1977). Deren Ergebnisse sind nicht sehr einfach zu interpretieren. GITTINS (1985) hat ihr eine umfassende Darstellung mit Anwendungsbeispielen gewidmet. Es bleibt abzuwarten, ob sie sich damit in der Ökologie doch noch durchsetzen kann. Zwei Haupthindernisse erschweren ihren Einsatz in der Pflanzenökologie: Zum einen ist es die grosse Zahl der Variablen (nämlich Arten), die zur Beschreibung der Vegetation nötig sind. Zum andern ist es die meist ausgeprägte Nichtlinearität in der Beziehung zwischen den beiden Datensätzen.

Zur Illustration des Linearitätsproblem es verwenden wir die Vegetationsdaten des Kapitels 7.3 und ergänzen diese mit einem Standortsfaktor, pH. Wir erhalten folgende Tabelle:

Aufn.	A	B	C	D	E	F
Art 1	1	2	2.5	2.5	1	0.5
Art 2	0	1	2	4	3	1
pH	4.0	4.2	4.3	4.8	5.2	5.7

Es wäre hier naheliegend, mittels der einfachen Korrelationsanalyse den Zusammenhang der Arten 1 und 2 mit dem pH zu untersuchen. In Abb. 9.1, A, sind die pH-Werte in Funktion der Abundanzwerte von Art 1 aufgetragen. Es ist offensichtlich, dass eine lineare Regression versagen muss: Sowohl bei pH 4.0 als auch beim Höchstwert von 5.7 erreicht Art 1 ein Minimum. Damit lässt sich aus der Häufigkeit ihres Auftretens innerhalb einer Probestfläche nicht eindeutig ein pH-Wert voraussagen. Ausserdem ist es wünschenswert, nicht eine einzige, sondern mehrere Arten gleichzeitig zu berücksichtigen. Man könnte also statt einer einzelnen Art die Faktorenwerte einer Komponentenanalyse mit dem pH

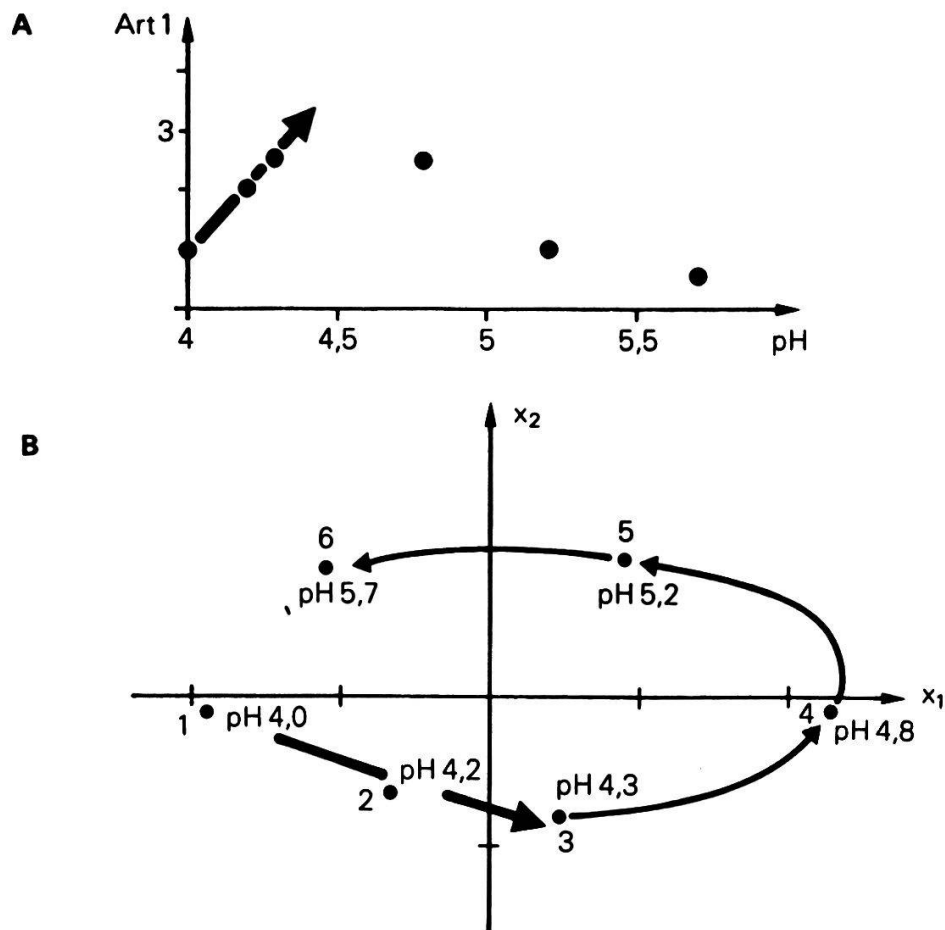


Abb. 9.1 Bestimmung der Beziehung zwischen dem Standortsfaktor pH und der Abundanz einer Pflanzenart (A), bzw. den Achsen einer Ordination (B). Fette Pfeile zeigen Bereiche, innerhalb welcher die Zusammenhänge annähernd linear sind.

korrelieren. Ein solcher Versuch ist in Abb. 9.1, B, dargestellt. Doch die nichtlineare Ähnlichkeitsstruktur der Aufnahmen lässt die Regressionsanalyse auch hier scheitern. Bei den ersten Punkten des Gradienten steigt der pH von links nach rechts, in der zweiten Hälfte von rechts nach links an. Die fett ausgezogenen Bereiche in Abb. 9.1, A und B weisen auf eine mögliche Lösung hin: Wird ausschliesslich in einem engen ökologischen Bereich gearbeitet, so stellt die lineare Regression tatsächlich eine brauchbare Näherungslösung dar. Leider tritt dieser Fall in der Praxis eher selten auf.

Günstiger stehen die Erfolgsaussichten einer Analyse, wenn die Daten bereits gruppiert sind. Es braucht bloss geprüft zu werden, ob die in der Vegetation gefundenen Gruppen sich auch standörtlich unterscheiden. Doch auch hier sind die Resultate kritisch zu interpretieren. Analysiert wird ja nicht die Struktur der Vegetation selbst, sondern das Resultat einer Gruppierungsanalyse. Dieses kann wiederum unterschiedlich treffend sein, je nach gewählter Methode.

Ziel einer Gruppierungsanalyse ist das Erkennen von Vegetationseinheiten, welche Umweltfaktoren möglichst genau vorausszusagen erlauben. Man wird also stets versuchen, eine geeignete Gruppierungsmethode zu finden (Kap. 5). Es kann aber sein, dass eine gute Klassifikationslösung unerwartet schlecht mit dem Standort übereinstimmt. In Abb. 9.2 ist ein solcher Fall dargestellt. Die beiden strukturell sauber getrennten Gruppen überlappen bezüglich pH deutlich. Eine solche Situation wurde von FEOLI und FEOLI-CHIAPELLA (1980) an reellen Daten nachgewiesen. Wir ersehen daraus, dass immer auch die Kombination der angewandten Methoden, nicht nur die einzelne Methode für sich kritisch auszuwählen ist.

9.2 Grafische Lösungen

Nicht immer muss auf eine rechnerische Lösung zurückgegriffen werden. Man kann ein Strukturmodell (Ordination, Dendrogramm) anhand standörtlicher Messungen direkt interpretieren. Im Prinzip wird in Abb. 9.1 so vorgegangen. Die

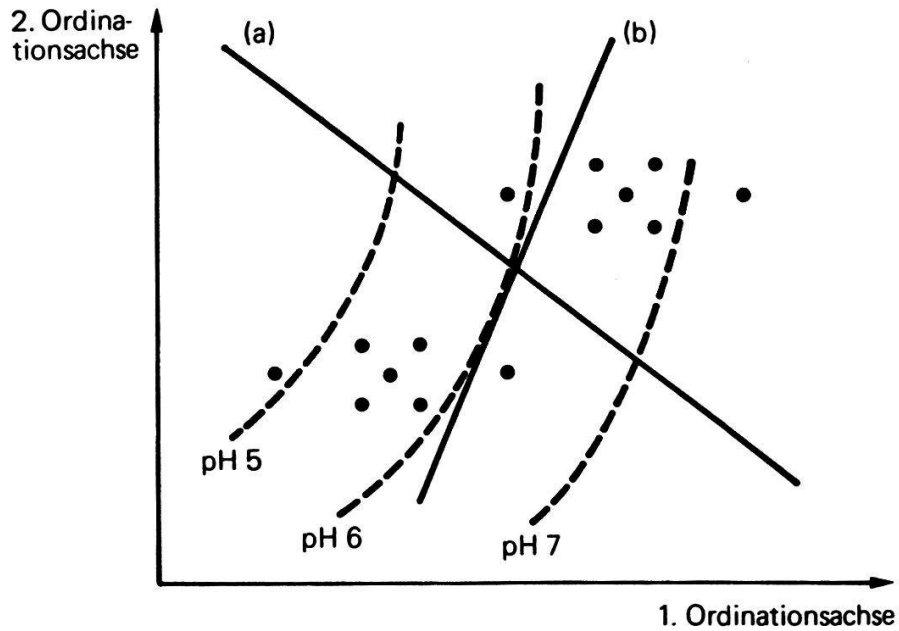


Abb. 9.2 Optimale Auftrennung einer Ordination in zwei Aufnahmegruppen auf Grund der Aehnlichkeiten der Aufnahmen (a) und des pH-Wertes (b).

Tabelle 9.1

Versuchsdaten zur Untersuchung einer Zeitreihe.

	1. Jahr			2. Jahr			3. Jahr		
Aufnahme- fläche Nr.	1	2	3	1	2	3	1	2	3
Art 1	1	1		1	1		2		
Art 2	1	2	1	2	1	1	2	1	1
Art 3		1	2		2	2	1	3	3
Art 4			1			2		1	3

dort gezeigte Ordination beruht ausschliesslich auf Vegetationsdaten (vgl. Kap. 8). Nachträglich werden die Aufnahmen mit gemessenen pH-Werten beschriftet. Damit lässt sich die gefundene Datenstruktur auf einfachste Weise ökologisch interpretieren. Im allgemeinen ist die Zahl der Aufnahmen bei pflanzenökologischen Untersuchungen so hoch, dass nicht mehr jeder Punkt der Ordination einzeln beurteilt werden kann. Dann ist es zweckmässig, Punkteschwärme innerhalb eines Standortbereiches zu umfahren und zu beschriften (Abb. 9.3). Dass sich die Bereiche meist überschneiden, beeinträchtigt die Uebersichtlichkeit kaum. Die Methode ist denn auch recht beliebt und Beispiele finden sich vielerorts (vgl. GREEN 1979, S. 215, CLYMO 1980, KOMARKOVA 1980).

Die Zeit kann als Spezialfall eines Standortfaktors betrachtet werden. Alle Aufnahmen einer Zeitreihe werden dazu in einem Datensatz vereinigt. In einer Ordination lassen sich in der Folge nicht nur die standörtlich bedingten Differenzierungen, sondern auch der zeitliche Trend darstellen. Das folgende Beispiel lehnt sich der Arbeit VAN DER MAARELS (1969) an. Wir gehen von Tabelle 9.1 aus. Die Population besteht aus 9 gleichberechtigten Individuen (Aufnahmen). Um die Analyse überblickbar zu halten, sollen die Daten nicht transformiert werden. Es wird eine Hauptkomponentenanalyse gerechnet. Im vorliegenden Beispiel basiert sie auf van der Maarels Koeffizient als Ähnlichkeitsmass. Man erhält als Eigenwerte $\lambda_1 = 2.10$, $\lambda_2 = 1.25$, $\lambda_3 = 0.377$, und $\lambda_4 = 0.277$. Bereits zwei Dimensionen reproduzieren 83.7% der gesamten Varianz. Die Aufnahmen, welche derselben Probestfläche entstammen, werden in der Reihenfolge der Erhebungszeitpunkte durch Pfeile verbunden. Bei der sehr einfachen Datenstruktur in Abb. 9.4 erkennt man leicht Entwicklungen in zwei verschiedene Richtungen. Sie werden durch die Aufnahmen 1 einerseits sowie 2 und 3 andererseits repräsentiert.

Als Ordination dargestellt (Abb. 9.4) sind die Veränderungen rascher und unter komplexeren Verhältnissen stets leichter erkennbar als in einer Vegetationstabelle. Man erkennt, dass sich die Aufnahme fläche 1 dem Anfangszustand der Fläche 2 annähert. Fläche 2 entwickelt sich in Richtung Fläche 3 und

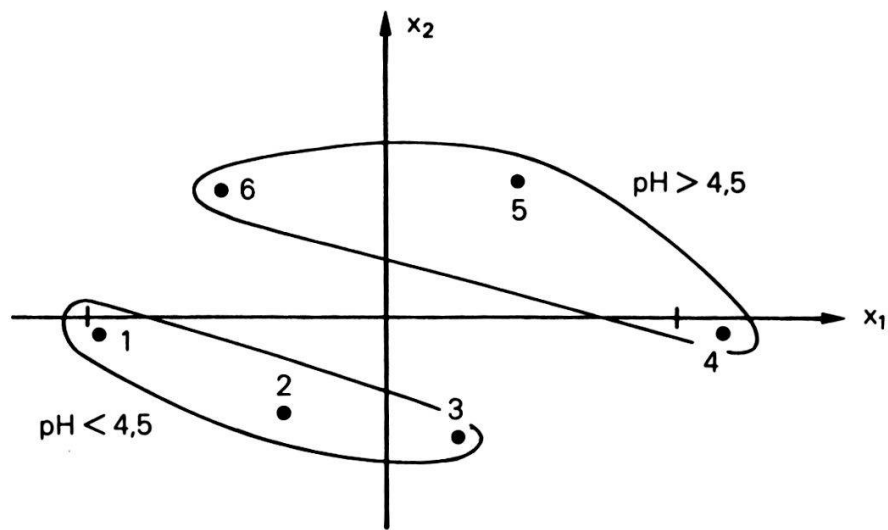


Abb. 9.3 Standörtliche Interpretation der Ordination aus Abb. 9.1. Aufnahmen innerhalb festgelegter Standortsbereiche werden umrandet und beschriftet.

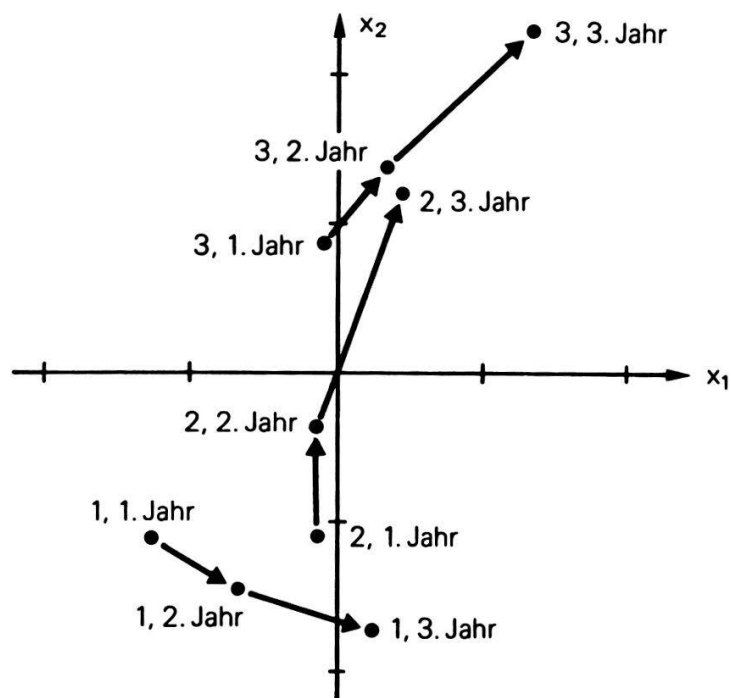


Abb. 9.4 Ordination der Daten einer Sukzessionsuntersuchung über drei Jahre (Tab. 9.1). Die Pfeile zeigen die Entwicklungsrichtung der Aufnahmeflächen.

erreicht als Endzustand etwa deren Zusammensetzung während des zweiten Jahres.

Aehnliche Möglichkeiten ergeben sich, wenn aus verschiedenen Jahren stammende Aufnahmen gemischt und gruppiert werden. Liegt eine deutliche Gruppenstruktur vor, so kann die "Wanderung" jeder sich verändernden Aufnahme durch ein Dendrogramm durch grafisches Hervorheben desselben verdeutlicht werden.

9.3 Korrelation von Ordination und Standort

Bereits in Abb. 9.1 wurde gezeigt, dass der Zusammenhang von Ordinationsachsen und Standortsmessungen mit der linearen Korrelationsanalyse nur in sehr einfachen Fällen erfassbar ist. Ordinationsverfahren wie die Hauptkomponenten- oder die Korrespondenzanalyse führen nur zum Ziel, wenn die Ähnlichkeitsstrukturen einigermaßen kontinuierlich und linear sind. Wesentlich einfacher ist die Aufgabe, wenn die Individuen und die Merkmale zu Gruppen zusammengefasst werden können. Dies ist in Abb. 8.2, Abschnitt 8.2 der Fall. Die dort gezeigte Konzentrationsanalyse liefert auch Koordinaten für die Gruppen. Wir bilden also die Gruppenmittelwerte der verschiedenen Standortsfaktoren (Tab. 9.2). Nun werden die Standortsfaktoren mit der ersten Ordinationsachse korreliert und man erhält

$$\begin{aligned} r(x, \text{pH}) &= -0.977 \\ r(x, \text{Höhe}) &= -0.991 \\ r(x, \text{Neigung}) &= 0.959 \end{aligned}$$

Die hohen Korrelationswerte weisen zunächst auf eine Uebereinstimmung zwischen der Gruppenstruktur und den Standortsfaktoren hin. Rein deskriptiv betrachtet bestätigen die Korrelationskoeffizienten die Zusammenhänge, die sich leicht aus Abb. 8.2 herauslesen lassen. Sehr vertrauenswürdig ist die auf nur drei Referenzgruppen basierende Aussage nicht. Einem Statistikbuch (BARTEL 1974) entnehmen wir, dass der Zufalls-Höchstwert im obigen Beispiel bei einer Irrtumswahrscheinlichkeit von 5% ganze 0.997 beträgt! Die soeben

berechneten Korrelationen sind offenbar mit Vorsicht zu interpretieren. Die Methode wird erst dann sinnvoll, wenn die Zahl der Gruppen grösser ist als drei. Natürlich müssten nun auch die Korrelationen der Standortsmessungen mit der zweiten und dritten Ordinationsachse überprüft werden. Ob bestehende Zusammenhänge auf diese Weise gefunden werden, bleibt ungewiss. Die Ähnlichkeitsstruktur der Aufnahmegruppen ist wie diejenige der ungruppierten Aufnahmen mehr oder weniger nichtlinear!

Die Erkenntnis, dass Aufnahmen durch viele Ordinationsmethoden als Kreise oder Spiralen dargestellt werden, machten sich FEOLI und FEOLI-CHIAPELLA (1980) zu Nutze. Weiss man mit einiger Sicherheit, dass die Vegetationsaufnahmen einem einfachen Gradienten entstammen, so lässt sich deren Position auf dem Ordinationskreis mit einem Winkelmass gut beschreiben. Solche Voraussetzungen sind im hier verwendeten Beispiel in extremer Weise gegeben (Abb. 9.3, 9.5). Zunächst muss jedoch eine Ursprungsrichtung gefunden werden, von welcher aus die Winkel zu messen sind. In unserem Beispiel (Abb. 9.5) liegt diese zwischen den Aufnahmen 6 und 1, also dem Anfangs- und dem Endpunkt des Gradienten. Wir verwenden der Einfachheit halber die x-Achse als Ursprungsrichtung und messen nun die Winkel im mathematisch positiven Sinne. Man kann sie berechnen oder direkt Abb. 9.5 entnehmen und findet dann folgenden Zusammenhang:

Aufnahmen	1	2	3	4	5	6
Positionswinkel α	3	44	120	179	225	322
pH	4.0	4.2	4.3	4.8	5.2	5.7

Die Korrelation ergibt:

$$r(\alpha, \text{pH}) = 0.979$$

Man kann sich auch hier fragen, ob dieser Wert über dem zu erwartenden Zufallshöchstwert liegt. Die Zahl der Frei-

Tabelle 9.2

Korrelation von Ordinationskoordinaten mit Standortsfaktoren von Aufnahmegruppen.

Aufnahmegruppe	1	2	3
x-Koordinate aus Konzentrationsanalyse	1.249	-0.481	-1.001
pH	4.75	5.60	6.23
Höhe über Meer	475	493.3	503.3
Neigung in Grad	11.9	7.0	2.33

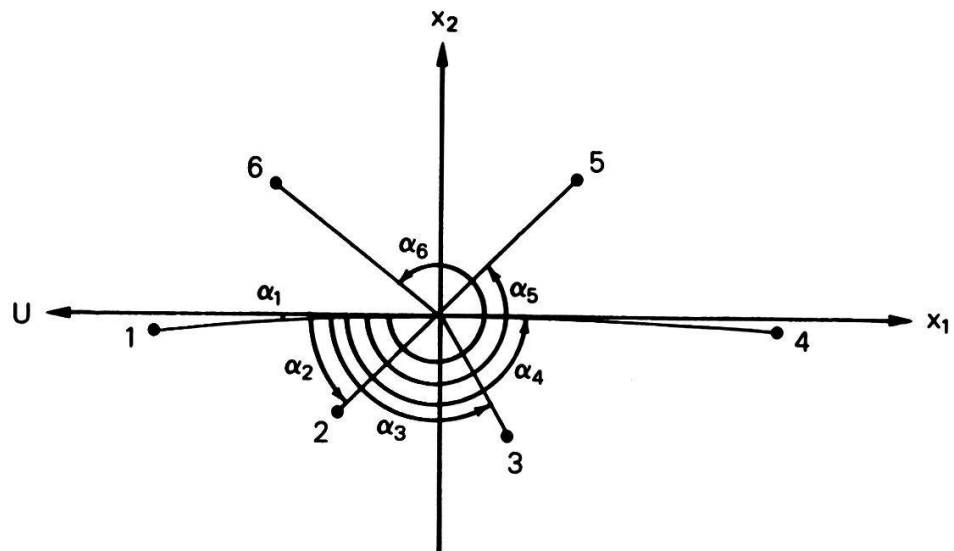


Abb. 9.5 Bestimmung des Winkels zwischen der Ursprungsrichtung U und den Aufnahmen in einer zweidimensionalen Ordination.

heitsgrade beträgt im obigen Beispiel $df = 5$. Einem Statistikbuch entnehmen wir den Zufallshöchstwert bei einer Irrtumswahrscheinlichkeit von 5% (BARTEL 1972) und finden ein r von 0.754. Unser $r(\alpha, pH)$ ist sogar noch signifikant von Null verschieden bei einer Irrtumswahrscheinlichkeit von 1%, denn der Zufallshöchstwert beträgt dann 0.874.

9.4 Die Diskriminanzanalyse

Zur Bestimmung des Zusammenhanges zwischen Vegetationseinheiten und Standortsfaktoren eignet sich die Diskriminanzanalyse. Sie erfordert Ausgangsdaten in einer Form, wie sie bei pflanzenökologischen Untersuchungen häufig anfallen. Von jeder Aufnahme ist die Gruppenzugehörigkeit vorzugeben. Man bedient sich zu deren Bestimmung beispielsweise einer Gruppierungsanalyse. Die Standortdaten müssen hingegen ordinal oder metrisch vorliegen. Geeignet sind also pH-Werte, Konzentrationen, Temperaturen, die Höhenlage usw. Das Resultat der Diskriminanzanalyse ist als Ordination darstellbar. Deshalb kommen dem Anwender Erfahrungen aus der Interpretation von Hauptkomponentenanalysen sehr zustatten. Die Nützlichkeit der Diskriminanzanalyse zeigt sich auch an der wachsenden Zahl praktischer Anwendungen (DEL MORAL 1975, WILDI 1977, FEOLI 1979). Darstellungen der Methode finden sich bei COOLEY und LOHNES 1971, MORRISON 1976, PIELOU (1977), SUCHARD-FICHER et al. (1982) und anderen. Ähnlich wie bei den Hauptkomponentenanalysen weichen die Auffassungen bezüglich der durchzuführenden Transformationen der Ausgangsdaten voneinander ab. Im Falle pflanzenökologischer Daten muss davon ausgegangen werden, dass Messwerte mit sehr unterschiedlichen Skalierungen vorliegen. Eine Normalisierung oder Standardisierung der Standortsvektoren ist deshalb stets angezeigt.

Wir fassen im folgenden die Diskriminanzanalyse als geometrisches Verfahren auf. Wie bei der Hauptkomponentenanalyse wird für einen mehrdimensionalen Punkteschwarm eine günstige Projektion gesucht. Die Ausgangskoordinaten sind die Standortsmessungen. Die neuen Diskriminanzachsen

(meist Diskriminanzfunktionen genannt) sollen nun so zu liegen kommen, dass die Aufnahmegruppen möglichst klar getrennt erscheinen. In Abb. 9.6 ist das in Kap. 1.3, Abb. 1.11 gezeigte Beispiel nochmals wiedergegeben. Es ist leicht erkennbar, dass weder die x-Achse noch die y-Achse allein die beiden Gruppen zu trennen vermag. Die erste (und einzige) Diskriminanzfunktion hingegen erlaubt eine klare Zuordnung aller Individuen aufgrund ihrer neuen Koordinaten. Wo immer eine solcherart verdeckte Gruppenstruktur vorliegt, hilft die Diskriminanzanalyse weiter. Abb. 9.6 zeigt ausserdem, dass die erste Achse der Hauptkomponentenanalyse hier das erwünschte Ergebnis nicht erbringen kann.

Anhand eines übersichtlichen Beispiels sollen Rechengang und Wirkungsweise dargestellt werden. Wie schon bei der Hauptkomponentenanalyse verzichten wir auf Beweisführungen und verweisen auf statistische Fachliteratur (z.B. MORRISON 1976). In Abb. 9.7 ist ein Datenbeispiel dargestellt, bei welchem die Diskriminanzanalyse ein aufschlussreiches Resultat verspricht. Gegeben sind drei Gruppen von Aufnahmen, welche im dreidimensionalen Raum liegen. Die Achsen sind mit A1, A2 und A3 bezeichnet. Wir nehmen an, dass es sich dabei um Standortsfaktoren handelt wie pH, Einstrahlungssumme oder Jahresniederschlag. Die Individuen der Gruppen liegen je auf einem Achsenkreuz, welches der Uebersichtlichkeit wegen innerhalb eines Würfels liegend dargestellt ist. Kein einziger Standortsfaktor x, y oder z ist geeignet, die Gruppen eindeutig zu trennen. Auch in zwei Dimensionen (Abb. 9.8, A) ist dies nicht möglich. Hingegen erahnt man, dass bei einer speziellen Blickrichtung schiefwinklig zu den drei Achsen alle drei Gruppen getrennt erscheinen könnten. In der Tat liesse sich die Gesamtpopulation so projizieren wie in Abb. 9.8, B. Dies ist denn auch das im folgenden gesuchte Endresultat.

Wie bei den Ordinationsmethoden dargestellt, bilden Artmächtigkeiten oder Standortsfaktoren den Ausgangspunkt der Analysen. Sie sind mit Hilfe von Faktorenladungen so zu transformieren, dass die gewünschte Projektionsrichtung erzielt wird. Bei der Hauptkomponentenanalyse ging es darum,

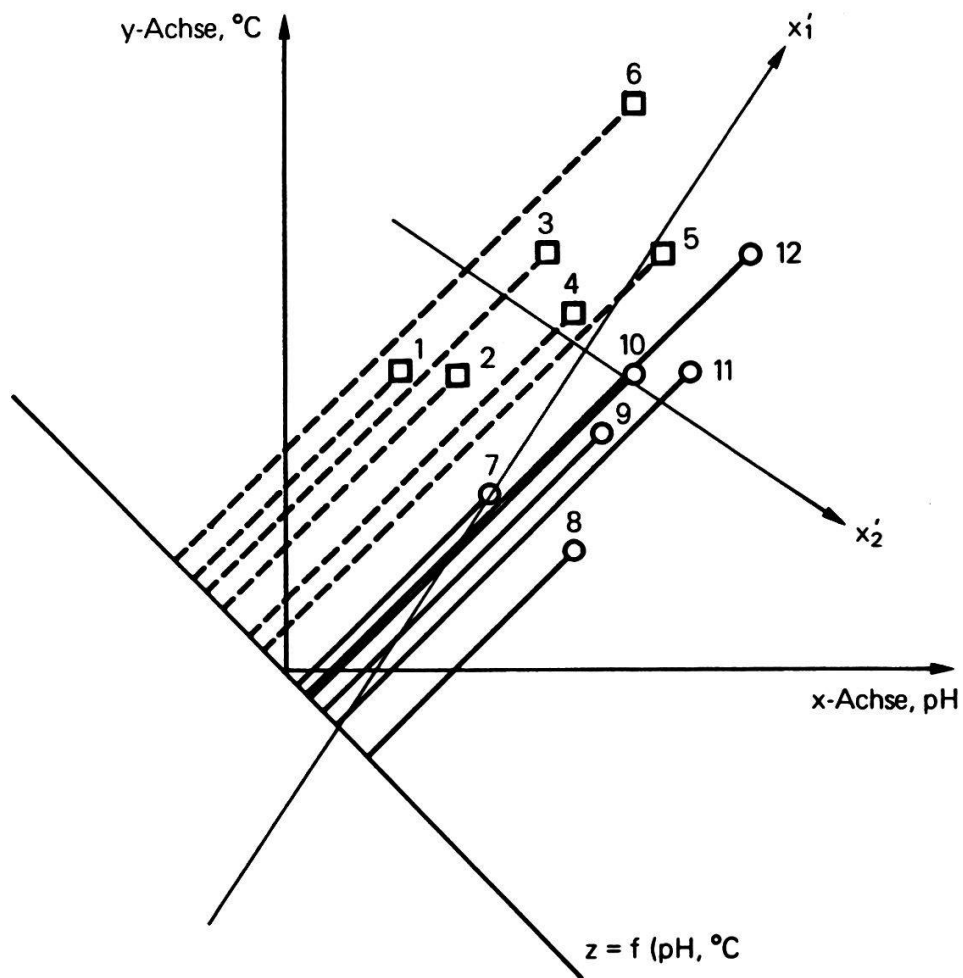


Abb. 9.6 Diskriminanzfunktion $z = f(\text{pH}, ^\circ\text{C})$ für zwei Gruppen einer Stichprobe. Zum Vergleich sind die beiden Achsen (x_1', x_2') einer Hauptkomponentenanalyse eingetragen.

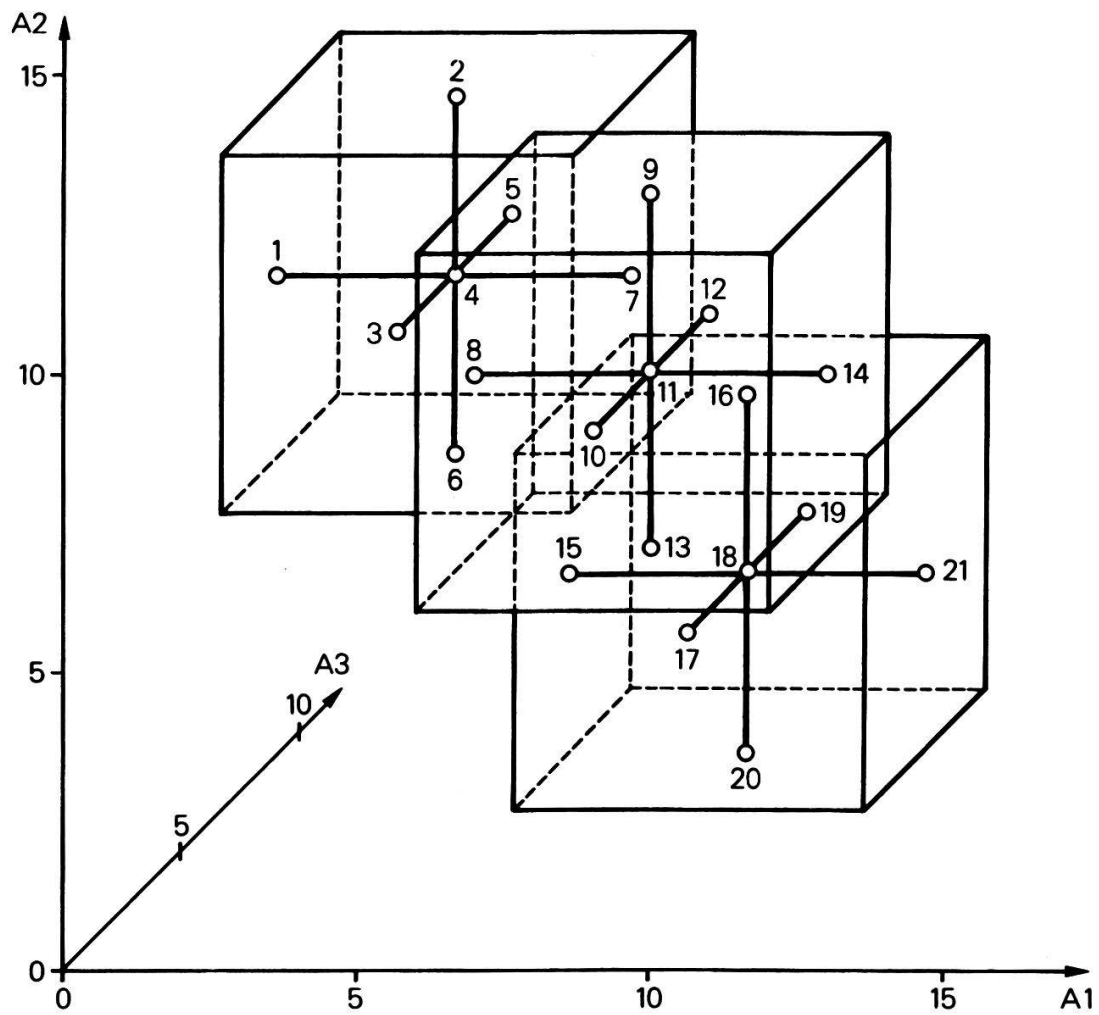


Abb. 9.7 Dreidimensionaler Punkteschwarm, bestehend aus drei durch Achsenkreuze verbundene Punktgruppen. Aus dieser Perspektive überschneiden sich die Gruppen. Die Würfel sollen die Lage der Punkte im Raum verdeutlichen.

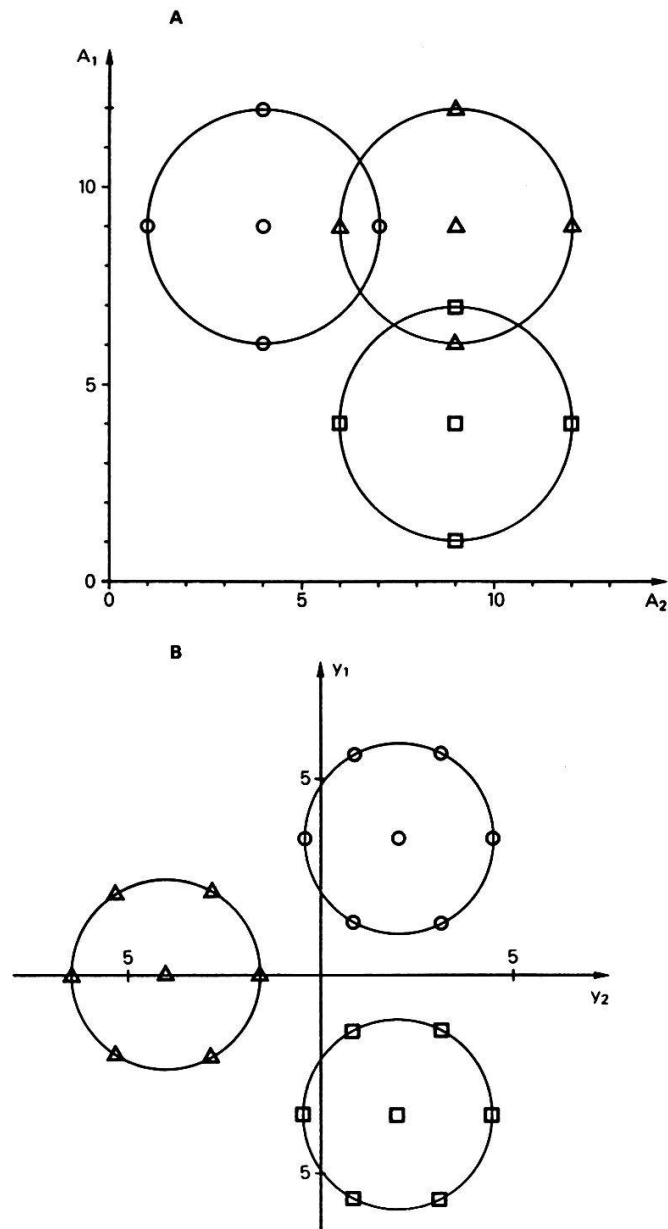


Abb. 9.8 Projektion des Punkteschwarmes in Abb. 9.7 in den Dimensionen a_1 und a_2 (A). Die Diskriminanzanalyse ist in der Lage, zwei neue Achsen y_1 und y_2 so zu finden, dass sich die drei Gruppen nicht überschneiden (B).

die Varianz der neuen Koordinaten auf der ersten Achse zu maximieren. Die Diskriminanzanalyse hingegen versucht, die Varianz zwischen den Gruppen möglichst gross erscheinen zu lassen. Das Problem der Trennung von Gruppen kennen wir bereits von der Varianzanalyse her (Kap. 6.3). Dabei gilt ja

$$\begin{array}{lll} \text{Totalvarianz} & \text{Varianz} & \text{Varianz} \\ \text{der Stichprobe} = & \text{innerhalb der} + & \text{zwischen den} \\ & \text{Gruppen} & \text{Gruppen.} \end{array}$$

Neu ist nur, dass bei der Diskriminanzanalyse mehrere Merkmale gleichzeitig berücksichtigt werden. Es sind somit stets Varianz- Kovarianzmatrizen zu analysieren. Liest man aus Abb. 9.7 alle Koordinaten heraus, so erhält man die Datenmatrix A (Tabelle 9.3). Wie schon bei der Hauptkomponentenanalyse, so sollen auch hier die Kolonnen zentriert werden. Dazu ist von jeder Koordinate der Mittelwert zu subtrahieren:

$$x_{ij} = a_{ij} - \bar{a}_j$$

Die neuen, transformierten x_{ij} - Werte finden sich ebenfalls in Tabelle 9.3. Nun ist die Aehnlichkeitsstruktur der zentrierten Standortsfaktoren x_1, x_2 und x_3 zu bestimmen. Als Aehnlichkeitsmass wählen wir der Einfachheit halber das Skalarprodukt. Wir bezeichnen die so entstehende Matrix mit T (für total). Bei zwei beliebigen Standortsfaktoren n und m gilt:

$$t_{nm} = \sum_{i=1}^N x_{im} x_{in}$$

N ist die Gesamtzahl der Aufnahmen. Für das Element t_{12} erhält man somit als Beispiel

$$\begin{aligned} t_{12} &= (-6.33*1.66)+(-3.33*4.66)+ \dots + (4.66*-1.333) \\ &= -58.33 \end{aligned}$$

Die gesamte Matrix ergibt

Tabelle 9.3

Datenmatrix für das Rechenbeispiel in Abb. 9.6.

Aufnahme	Gruppe	Standortsfaktor			Zentrierte Daten		
		A1	A2	A3	X1	X2	X3
1	1	1	9	9	-6.33	1.66	1.66
2	1	4	12	9	-3.33	4.66	1.66
3	1	4	9	6	-3.33	1.66	-1.33
4	1	4	9	9	-3.33	1.66	1.66
5	1	4	9	12	-3.33	1.66	4.66
6	1	4	6	9	-3.33	-3.33	1.66
7	1	7	9	9	-0.33	1.66	1.66
8	2	6	9	4	-1.33	1.66	-3.33
9	2	9	12	4	1.66	4.66	-3.33
10	2	9	9	1	1.66	1.66	-6.33
11	2	9	9	4	1.66	1.66	-3.33
12	2	9	9	7	1.66	1.66	-0.33
13	2	9	6	4	1.66	-1.33	-3.33
14	2	12	9	4	4.66	1.66	-3.33
15	3	6	4	9	-1.33	-3.33	1.66
16	3	9	7	9	1.66	-3.33	1.66
17	3	9	4	6	1.66	-3.33	-1.33
18	3	9	4	9	1.66	-3.33	1.66
19	3	9	4	12	1.66	-3.33	4.66
20	3	9	1	9	1.66	-6.33	1.66
21	3	12	4	9	4.66	-3.33	1.66
$\bar{x} =$		7.33	7.33	7.33	0	0	0

$$T = \begin{pmatrix} 170.66 & -58.33 & -58.33 \\ -58.33 & 170.66 & -58.33 \\ -58.33 & -58.33 & 170.66 \end{pmatrix}$$

Die Gleichförmigkeit verschiedener Elemente ergibt sich aus dem gewählten Beispiel: Alle Gruppen in Abb. 9.7 sind ja gleich gross und haben die gleiche Form. Im folgenden Schritt sind die Aehnlichkeitsverhältnisse innerhalb der einzelnen Gruppen zu bestimmen. Das Vorgehen ist genau gleich wie für die Gesamtstichprobe: Die Daten werden (innerhalb der Gruppen!) zentriert und die Matrix der Skalarprodukte W_1 (für "within") berechnet. Man erhält für die erste Gruppe:

$$W_1 = \begin{pmatrix} 18.0 & 0 & 0 \\ 0 & 18.0 & 0 \\ 0 & 0 & 18.0 \end{pmatrix}$$

Es ist leicht einzusehen, dass die Resultate für die Gruppen 2 und 3, also W_2 und W_3 , identisch ausfallen müssen. Aus der Summe der drei Matrizen ergeben sich schliesslich die totalen Abweichungsquadrate innerhalb der Gruppen:

$$W = W_1 + W_2 + W_3 = \begin{pmatrix} 54.0 & 0 & 0 \\ 0 & 54.0 & 0 \\ 0 & 0 & 54.0 \end{pmatrix}$$

Nun sollen also die neuen Koordinaten gesucht werden, welche die Punktwolken optimal auflösen. Der Weg führt wie bei der Hauptkomponentenanalyse über ein Eigenwertproblem. Zunächst sind geeignete Eigenwerte Lambda (λ) zu suchen. Sie ergeben sich durch Auflösung der Determinantengleichung

$$[W^{-1} (T-W) - \lambda I] = 0.$$

Das Problem unterscheidet sich von demjenigen der Hauptkomponentenanalyse nur insofern, als zunächst der Ausdruck $W^{-1} (T-W)$ zu berechnen ist. $T-W$ soll mit A bezeichnet werden und wir erhalten

$$A = T-W = \begin{bmatrix} 170.66 & -58.33 & -58.33 \\ -58.33 & 170.66 & -58.33 \\ -58.33 & -58.33 & 170.66 \end{bmatrix} - \begin{bmatrix} 54.0 & 0 & 0 \\ 0 & 54.0 & 0 \\ 0 & 0 & 54.0 \end{bmatrix}$$

$$= \begin{bmatrix} 116.66 & -58.33 & -58.33 \\ -58.33 & 116.66 & -58.33 \\ -58.33 & -58.33 & 116.66 \end{bmatrix}$$

W^{-1} ist die Inverse von W, d.h. die Lösung der Gleichung $WW^{-1}=I$. Die Bedeutung der Matrixinversion ist z.B. bei BATSCHULET (1980) erklärt. Dort findet sich auch ein Lösungsweg (S. 518-520). Folgt man den beschriebenen Rechenregeln, so scheitert man mit unserem Beispiel: Die Determinante der Matrix W ist Null, die Inverse W^{-1} ist undefiniert. Glücklicherweise wurde von Mathematikern ein sicherer Weg zur Bestimmung der Eigenstruktur von $W^{-1} A$ gefunden. Die einfachen, aber rechenintensiven Operationen findet man bei COOLEY und LOHNES (1971), S. 192 bis 194, zusammen mit einem Computerprogramm. Wir verzichten deshalb auf die detaillierte Darstellung des Rechenganges. Für unser Beispiel lauten die Eigenwerte

$$\lambda_1 = 3.24 ; \lambda_2 = 3.24$$

Die Eigenvektoren alpha (α) betragen

Alte Koordinate	Diskriminanzachse	
	1	2
1	-0.408	-0.707
2	-0.408	0.707
3	0.817	0.0

Die Interpretation der Eigenvektoren erfolgt genau gleich wie bei der Hauptkomponentenanalyse. Demzufolge ist $\alpha_{11} = -0.408$ die Korrelation (der Cosinus) der alten Achse 1 mit der Diskriminanzachse 1. Diese ist mit $\alpha_{13} = 0.817$ ähnlich orientiert wie die alte Achse 3. Die Diskriminanzachse 2 steht in 45° zu den alten Achsen 1 und 2 (Cosinus $45^\circ =$

Tabelle 9.4

Koordinaten auf den Diskriminanzachsen

Aufnahme	Gruppe	Koordinaten	
		Y1	Y2
1	1	3.266	5.657
2	1	0.817	5.657
3	1	-0.408	3.536
4	1	2.041	3.536
5	1	4.491	3.536
6	1	3.266	1.414
7	1	0.817	1.414
8	2	-2.858	2.121
9	2	-5.307	2.121
10	2	-6.532	0.000
11	2	-4.082	0.000
12	2	-1.633	0.000
13	2	-2.858	-2.121
14	2	-5.307	-2.121
15	3	3.266	-1.141
16	3	0.817	-1.141
17	3	-0.408	-3.536
18	3	2.041	-3.536
19	3	4.491	-3.536
20	3	3.266	-5.657
21	3	0.817	-5.657

0.707). Zur Achse 3 steht sie rechtwinklig. Dass dem so ist, zeigt auch Abb. 9.8, B. Nun sind noch die dort verwendeten Koordinaten Y zu bestimmen. Dies geschieht wiederum wie bei der Hauptkomponentenanalyse. Für jede Dimension p gilt

$$y_{1p} = x_{1p} \alpha_{11} + x_{2p} \alpha_{21} + x_{3p} \alpha_{31}$$

Die aktuellen Werte (Tabelle 9.4) sind

$$\begin{aligned} y_{11} &= (-0.408)(-6.33) + (-0.408)(1.667) + (0.817)(1.667) \\ &= 3.26 \end{aligned}$$

$$\begin{aligned} y_{12} &= (-0.707)(-6.33) + (0.707)(1.667) + (0.0)(1.667) \\ &= 5.65 \end{aligned}$$

Alle weiteren Koordinatenwerte sind in Tabelle 9.4 wiedergegeben. Wie in Abb. 9.8, B, gezeigt, präsentieren sich die drei Punkteschwärme in zwei Dimensionen klar getrennt. Sind die ursprünglichen Koordinaten Standortsfaktoren, so können wir jetzt festhalten, dass diese die - aufgrund floristischer Angaben - vorgegebenen Gruppen einzeln nicht zu unterscheiden erlauben. In Kombination charakterisieren sie jedoch jede Gruppe eindeutig.