

VORLESUNGEN ÜBER DIE VIELDIMENSIONELLE ANALYSE IN DEN NATURWISSENSCHAFTEN

Prof. Dr. Richard Reyment, Universität Uppsala, Schweden

INHALT

1. Vorrede
2. Einleitung
3. Kurze Einführung in die Matrizenrechnung und die Differentialgeometrie
4. Einführung in die Multivariatanalyse
5. Anwendungsbereich: die quantitative Genetik
6. Anwendungsbereich: Die Analyse von Grösse, Zuwachs und Form
7. Anwendungsbereich: Chemometrie
8. Anwendungsbereich: Ökologie
9. Anwendungsbereich: Ozeanographie
10. Anwendungsbereich: Taxonomie
11. Anwendungsbereich: Anthropometrie
12. Anwendungsbereich: Petrologie

1. VORREDE

Der Zweck dieser Arbeit läßt sich ganz einfach beschreiben, nämlich: die Methoden und Begriffe der modernen multivariatstatistischen Datenanalyse in deutscher Sprache darzustellen. Warum dieses Interesse, kann man sich wohl fragen? Es gibt ja z.Z. in Europa ungefähr 95 Millionen Menschen im deutschen Sprachgebiet, was für einen Sondereinsatz zu sprechen scheint - auch wenn dieser Einsatz von einem Engländer kommt. Ungeachtet, daß Englisch heutzutage die lingua franca der wissenschaftlichen Kongreße geworden ist, und es nunmehr sehr ungewöhnlich ist, daß in internationalen Zusammenhängen ein Vortrag auf Deutsch gehalten wird (das gleiche Schicksal trifft jetzt auch das Französische), ist, meines Erachtens, eine Übersicht angebracht in der die oft schwierigen mathematischen Begriffe verständlich gemacht werden.

Man kann nicht die Tatsache wegerklären oder verschweigen, daß das Zentrum der quantitativen Forschung in der Biomathematik und Geomathematik in den Vereinigten Staaten liegt und, daß man in Europa technisch und wissenschaftlich mehr und mehr ins Hintertreffen gerät, - beinahe sämtliche einschlägige Lehrbücher sind in englischer Sprache verfasst.

Diese kurze Einleitung in die multidimensionale Analyse ist folgendermaßen eingeteilt: Die Grundbegriffe der für die Multivariatanalyse so notwendige Matrizenrechnung werden zuerst durchgenommen. Danach folgt eine Einführung in die Methoden der multivariatstatistischen Analyse. Dieser Abschnitt ist in herkömmlicher Weise dargestellt worden. Man findet dort die in der

biometrischen Literatur häufig zitierten Verfahren, mit erläuternden Beispielen versehen. Gewi ße Aspekte der Differentialgeometrie werden auch angeführt, da dieses Thema eine sehr wichtige Rolle bei der neulich vorgelegten geometrischen Morphometrie spielt.

Der Hauptteil des Textes besteht aus praktischen Beispielen, welche ich unter geeigneten Titeln gruppiert habe. Ich habe soweit möglich verschiedene Teile der Naturwissenschaften durch repräsentative Erläuterungen berücksichtigt. Folglich finden sich Beispiele für quantitative Genetik, analytische Chemie, die Analyse von Zuwachs und Form, Ökologie, Ozeanographie, Petrographische Analyse, Sedimentologie, Serologie und Anthropomorphometrie.

Ich gebe eine kurzgefaßte Einführung in die neuere Methodik der mathematischen Statistik, die von unmittelbarer Bedeutung in den Erdwissenschaften und Lebenswissenschaften ist. Das vorgelegte Material zerfällt in zwei Teile, einerseits: die klassische multivariatstatistische Analyse und, zweitens, auf Differentialgeometrie basierte Verfahren. Die geometrische Morphometrie muss als eine echte Bereicherung unseres Denkens gesehen werden. Einige Bemerkungen über die Verwendbarkeit der Fraktalgeometrie in der Paläobiologie werden beigefügt.

Folgende Fragen werden behandelt:

1. **Chemische Analysen** : Mit der allgemeinen Zugänglichkeit von kleinen kraftvollen Rechengeräten, sowie der weite Verbreitung von Rechenprogrammen, hat jeder Forscher die Möglichkeit, ziemlich komplizierte Analysen ausführen zu können. Aber mit dieser augenscheinlich glücklichen Lage folgt etwas Unangenehmes. Die fertigen Programme sind fast immer nur für theoretisch einwandfreies Material gedacht. Damit sei gesagt, da ß Abweichungen von der Theorie in manchen Fällen die Ergebnisse einer Untersuchung mehr oder weniger ungültig machen können. Die Geochemie stellt zum Beispiel einen derartigen Fall dar, weil die Daten aus Bestandteilvariablen bestehen.

Geometrie: (a) Die Multivariatstatistik ist auf die analytische Geometrie von Descartes (1596-1650) gegründet, was bedeutet, daß die Eigenschaften des cartesianischen Raumes gelten. Zu erwähnen ist, daß dieser enge Raum nunmehr überwunden worden ist, und zwar durch die Einführung einer ganz neuen Geometrie, d.h. der Geometrie der Fraktalen, einer von Benoit Mandelbrot vorgelegten Theorie. Diese Fragen sind im vierten Kapitel meines neulich bei Pergamon-Press erschienenen Buchs **Multidimensional Palaeobiology** behandelt worden (1991).

(b) Erörtert wird auch die Frage der Analyse von Zuwachs, Größe und Gestalt (= Form), eine Frage welche die Menschheit seit der Zeit der alten Griechen interessiert hat. Die algebraisch mehr oder weniger automatische Methode der multivariaten Analyse von Zuwachs und Form ist durch neue Idéen, im Sinne von D. W. Thompson (1917), ersetzt worden. In diesem Zusammenhang werde ich die Deformation, das heißt die **Verbiegung** von Gittern und

die biologische Anwendung der Theorie des Krümmens von dünnen Blechschablonen (*théorie des plaques minces*) einführen.

Sämtliche dieser Methoden sind auf neuen Ideen begründet, auf Auffassungen und Begriffen, die vor zehn Jahren kaum existierten. Die dynamische Biometrie befindet sich ohne Zweifel in einer Phase beschleunigten Wachstums.

Zweiter Teil EINLEITUNG

Die Anwendung quantitativer Methoden in den Naturwissenschaften hat eine verhältnismäßig kurze Geschichte, besonders was statistisch orientierte Verfahren betrifft. Ich gebe hier eine kurze Zusammenfassung der wichtigsten Fortschritte:

Die ersten Einsätze findet man in biologischen Zusammenhängen und zwar bei Untersuchungen mit taxonomischer Einrichtung. Abgesehen von einem etwas behelfsmässigen Versuch von Pearson (1901), wobei er einen Vorläufer der Hauptkomponentenanalyse andeutete, sieht man R. A. Fisher (England) und P. C. Mahalanobis und C. R. Rao (Indien) als die eigentlichen Erfinder der multivariaten Biometrie an (Das Wort Biometrie stammt von Karl Pearson, W. F. R. Weldon und F. Galton.)

P. C. Mahalanobis, der erste Vorsteher des Indischen Statistischen Institutes in Calcutta, wurde anfangs der Zwanzigerjahre mit der anthropomorphometrischen Untersuchung Indiens beauftragt. Arbeitend mit einer verhältnismässig großen Anzahl von Schädel- und Körpervariablen, wollte man eine morphologische Klassifikation erreichen, welche die verschiedenen Kasten des Landes richtig bestimmte. Heute ist man darüber erstaunt: eine notwendige Voraussetzung des Grundmodelles ist, daß jede Kaste ihre eigenen morphologischen Eigenheiten besitzt - derartige Auffassungen gehören zu den meist kritisierten Seiten der Rassenbiologie. Ich kenne jedoch keine ablehnenden Bemerkungen zu den Ergebnissen der anthropomorphometrischen Untersuchungen der Calcutta-Schule und kann nur den Schluß ziehen, daß Indier um Indier, so was gehe!

Mahalanobis hat einen statistischen Abstand (siehe S.) benutzt, wobei die Differenzen zwischen den Mittelwerten der Stichproben (der Differenzmittelvektor) als statistischer Abstand verwendet wurde, aber dadurch justiert, daß die Korrelationen zwischen Variablen berücksichtigt wurden. Dieser Abstand ist dimensionsfrei und kann deswegen nicht direkt gemessen werden. Pearson hatte schon früher eine ähnliche Idee gehabt, hatte aber die Frage der Rolle der Korrelationen nicht verstanden (Pearsons Koeffizient der Rassenähnlichkeit). Leider führte dieses Mißverständnis zu einem unbehaglichen Streit, bevor sich Pearson ergab. Die Berechnungen des Abstandes von Mahalanobis sind sehr umständlich und zeitraubend und er benutzte ein kleines Heer von Mathematikern dabei. Heute ist diese Arbeit nahezu banal und jeder Tischrechner kann die notwendigen Berechnungen leisten.

Dem englischen Mathematiker, Biometriker und Genetiker, R.A. Fisher wurde anfangs der Dreißigerjahre das folgende Problem gestellt: ein amerikanischer Botaniker wollte drei Arten des Geschlechtes *Iris* auf Grund von vier Meßungen bestimmen. Die Lösung Fishers war die nunmehr sehr gut bekannte Methode der **Diskriminanzfunktion**. Auf Grund der benutzten Meßungen konnte Fisher die drei Arten richtig identifizieren (er sagte leider klassifizieren) und dadurch, war es möglich ein neues Exemplar richtig bestimmen zu können.

Eine natürliche Fortsetzung kam gleich danach. Diesmal handelte es sich um Mumien aus Ägypten, welche Fisher mit der gleichen Methode analysierte. Er konnte jedoch die Signifikanz seiner Ergebnisse nur ganz ungefähr schätzen - ein Problem auch für die indischen Statistiker.

Schon 1932 hatte der Nordamerikaner H. H. Hotelling ein statistisches Testverfahren entwickelt, das eine Art von Verallgemeinerung des nunmehr klassischen *t*-tests von Student war. Nach einiger Zeit verstand man die Zusammenhänge zwischen den Verfahren von Fisher, Mahalanobis und Hotelling und sah ein, daß sie eng mit einander verbunden sind. Sämtliche Methoden bilden jetzt eine einheitliche Theorie. Die Signifikanz des generalisierten Abstandes wird durch das T von Hotelling geschätzt. Die Scheidelinie zwischen zwei Gruppen stellt die erwähnte Diskriminanzfunktion dar.

Gleichzeitig mit den Biometrikern wurde auf psychometrischer Seite gearbeitet. Hierbei wurde die weit bekannte **Faktoranalyse** entwickelt (Thurstone, Cattell usw.) und die schon erwähnte Methode der Hauptkomponentenanalyse von Hotelling als eine mit der Faktoranalyse gleichberechtigte, aber dennoch verschiedene Methode vorgelegt. Die Hauptkomponentenanalyse ist vermutlich das wichtigste einzelne Verfahren der Multivariatstatistik.

Seit der Jugendzeit der Statistik hat man eine ganze Reihe von neuen Methoden eingeführt. Alle diese sind jedoch Verbesserungen und Erweiterungen der Grundbegriffe; zwar wichtig, aber im Großen und Ganzen haben die ursprünglichen Forscher die Reichweite der verschiedenen Probleme gut verstanden. Die Statistik ist nur eine Hilfskraft der Wissenschaft, nicht mehr, nicht weniger. Fast jeder Zweig hat jedoch seine eigenen Sonderverfahren entwickelt. Die Chemometrie und, analytisch betrachtet, verwandte Wissenschaften arbeiten mit ganz speziellen Daten, nämlich Meßvektoren mit konstanter Summe. Beispiele sind: Gesteinsanalysen, serologische Analysen (Blutgruppen), chemische Bestandteile einer Stichprobe, usw. Solche Beobachtungen verwenden technische Spezialisierungen, welche die räumlichen Verhältnisse berücksichtigen. Diese sind erst ganz neulich entwickelt worden, obwohl die Notwendigkeit dafür bekannt war, seit den früheren Arbeiten von Pearson und Galton (über das Vorkommen von so genannten Scheinkorrelationen).

Die statistische Auswertung **atypischer Beobachtungen** stellt ein bisher nur unbefriedigend behandeltes Problem dar. Solche Beobachtungen sind in der Regel biologisch oder geologisch

einwandfrei, weichen aber von den statistischen Normen mehr oder weniger stark ab. Ein gutes Beispiel machen polymorphe Erscheinungen aus. Es gibt nunmehr verwendbare Methoden (welche leider kaum in den zugänglichen Handbüchern vertreten sind), die eine wissenschaftlich vertretbare Analyse erlauben. Hier spricht man von atypischen Werten und einflußreichen Werten (W. Krzanowski (1988), W. Hampel *et al.* (1985), usw).

Die **quantitative Genetik** gründet sich auf die Fortschritte Fishers bei der Ausarbeitung eines wesentlichen Teils der heutigen genetischen Theorie, mit wichtigen Beiträgen von S. Wright. Diese Ergebnisse wurden dann erweitert und ausgebaut von D. S. Falconer, der als der eigentliche Gründer der quantitativen Genetik anzusehen ist. Unter mehreren Jahren wurde die quantitative Genetik ausschließlich auf den Gebieten der Haustierkunde und Pflanzenveredelung verwendet, bis R. Lande eine wichtige Erweiterung der Begriffe und eine neue Synthese hervorbrachte. Diese Synthese vereinigt im gleichen Zusammenhang die verschiedenen Abschnitte der quantitativen Genetik in ein Modell für das Studium der Evolution des Phänotyps. Auswahlprinzipien werden mit der Hypothese der genetischen Drift S. Wrights vereinigt.

Die multivariatstatistischen Aspekte der Landeschen Synthese kommen durch die Deutung seiner Matrizengleichungen zum Ausdruck. Die theoretischen Darlegungen von R. Lande machen von gewissen in der theoretischen Physik und Mechanik vorkommenden Darstellungsweisen Gebrauch, die sonst ziemlich unbekannt in der biometrischen Literatur sind. Ein ähnliches Sachverhältnis gilt für die Arbeiten von F. L. Bookstein, der die Analyse von Größe, Zuwachs und Form völlig revolutioniert hat. Die hergebrachte analytische Ordnung ist nach G. Teissier, P. Jolicoeur, J. E. Mosimann und anderen auf die Aufspaltung mittels Eigenanalyse der Kovarianzmatrize von (Distanz)-Variablen gegründet. Diese Verfahrensweise basiert nicht auf einem biomathematischen Modell, sondern beruht mehr oder weniger auf einem Zufall. Es finden sich eine Reihe von verwandten Methoden, welche auf die Eigenspaltung einer positiv-definierten Matrize zurückzuführen sind. Neuerdings, haben einige Forscher begonnen, die Analyse der Form eines Organismus auf die entsprechende Silhouette zu basieren.

Das neue Verfahren Booksteins zeichnet sich dadurch aus, daß es die Formvariation bei einem Organismus als eine Deformation erfasst und dadurch der ursprünglichen Gitterdeformationen von D. W. Thompson (1917) eine mathematisch durchführbare Behandlung geben kann. Die Frage gehört nicht zu den einfachsten der Biomathematik. Neue (aber ziemlich unpraktisch dargestellten) Ideen sind jüngst von J.-F. Durand (Montpellier) entwickelt worden, die für die Wiederherstellung und Deutung problematischer Fossilien (z. B. den Organismen des Kambriums) interessant sein könnten, aber unter logischen Auseinandersetzungen leiden. Diese Methode ist "kontra-intuitiv" dadurch, daß die Ergebnisse gegen den "gesunden Menschverstand" streiten.

Worin bestehen die Verbesserungen der Darstellungen Booksteins

(1978, 1991) im Vergleich mit der Zerlegung nach gewöhnlichen Hauptkomponenten und verschiedenen Prokrustes-Ausgleichsverfahren? Diese Frage läßt sich ganz einfach beantworten. Die Methode des Interpolierens mittels der s.g. *Dünnblechsspline* kann ohne weiteres aufgebaut und an interessante theoretische Überlegungen angeschlossen werden. Dazu kommt, daß die ganze Arbeitsweise der klassischen Multivariatstatistik zugänglich wird.

ÖKOLOGIE läßt sich in vielen Arten und Weisen quantifizieren. Die statistische Ökologie bezieht sich hauptsächlich auf die Analyse von Artenfrequenzentabellen, welche man, soweit ausführbar, mit physikalischen und chemischen Faktoren der Umwelt korrelieren möchte. Derartige Fragestellungen sind heutzutage viel wichtiger zuvor, dadurch daß Umweltfragen, darunter die Ökotoxikologie, eine hervorragende Rolle spielen. Hier sind die Möglichkeiten zu praktischen Anwendungen offenbar sehr zahlreich.

Die Gestaltung der multivariatstatistischen Analyse ist bei völlig spezifizierten Situationen ziemlich kompliziert - man braucht nur die Interaktion zwischen mehreren Mengen ungleicher Variablen zu bedenken. Eine verwendbare Methodik ist die von C.J.C. ter Braak angeführte kanonische Korrespondenzanalyse. Dieses Verfahren vereinigt die Vielseitigkeit der normalen Korrespondenzanalyse im Sinne von J.-P. Bénécri mit dem Modelle der kanonischen Korrelationsanalyse nach H. Hotelling (1936), dadurch gekennzeichnet, daß man Teilmengen von Variablen verschiedener Herkunft zu berücksichtigen hat (beispielsweise, morphologische Charaktere, eine Teilmenge von klimatologischen Variablen, und eine Teilmenge von chemischen Bestimmungen). Weitere Fragen beziehen sich auf die sedimentologischen und naturgeographischen Seiten der Ökologie, worunter die Geochemie häufig eine bedeutende Rolle spielt. Die Analyse chemischer Daten stellt jedoch ein schwieriges Problem dar.

Die analytische Chemie, sei sie geochemisch ausgerichtet oder eine normale Äußerung der chemischen Forschung, muß mit besonderer Vorsichtigkeit behandelt werden. Die normalen Methoden der multivariaten Statistik lassen sich nicht einsetzen, wenn die analytischen Tabellen so beschaffen sind, daß die Summe jeder Reihe konstant bleibt: in anderen Worten, wenn jede Reihe zu Eins oder zu Hundert summiert. Wie sprechen also hier von Proportionen und Frequenzen. Um die Art des Problems verständlich zu machen, können wir Tabellen der Allelen des ABO-Komplexes von Blutgruppen betrachten. Jede Population [abgesehen von Sondergruppen wie den Indianern Südamerikas (nur O), den australischen Eingeborenen (nur A), den Basken, vielen Gaskogner (B fehlt)] enthält hauptsächlich die Allele O . An zweiter Stelle kommt, normalerweise, die Allele A_1 , und am dritten Platz die Allele B (bei den meisten europäischen Populationen). Bei fast allen europäischen Populationen findet man A_2 am vierten Platz (die Lappen Skandinaviens bilden eine Ausnahme). Beschäftigt man sich nur mit einer Zusammenstellung von Blutgruppenfrequenzen beschäftigt, braucht man nur drei Allele zu kennen. Die vierte findet man ganz einfach durch Subtraktion der Summe von drei Allelen von 100. Folglich sagt man, daß so eine geschlossene Datenreihe $p-1$ Freiheitsgrade besitzt. Dazu kommt die Tatsache,

daß man nicht mehr mit dem Cartesianischen Raum zu tun hat, sondern mit einem Teilraum, Simplexraum genannt. Die statistische Analyse der Simplexdaten benötigt besonders angepaßte Methoden. Solche Daten findet man überall in der Chemie, der Sedimentologie, der Petrologie und der Genetik, und manchmal auch in der Ozeanographie.

Dritter Teil: Einführung in die Matrizenrechnung

Einleitendes

Grundbegriffe. Vektoren und Matrizen. Geometrische Darstellung der Eigenschaften derselben.

Eigenvektoren und ihre Anwendung in der Naturwissenschaft.

Zerlegung einer rechteckigen Matrix (z. B. eine Datenmatrix) und das s.g. Zweierverhältnis Q-R. Die Grundstruktur einer Matrix (Ansatz von Eckart und Young) und ihre statistische Rolle. Die quadratische Form $\mathbf{x}^T \mathbf{A} \mathbf{x}$.

Differentialgeometrie. Affine (einheitliche) und nicht-affine (uneinheitliche) Transformationen eines starren Körpers. Statistische Verhältnisse im Simplexraum.

Vektorrechnung

Die Vektorrechnung hat sich aus den Bedürfnissen der Mechanik und Physik des 19. Jahrhunderts entwickelt. Beispiele für vektorielle Größen sind Translation, Geschwindigkeit, Beschleunigung, Kraft, magnetische und elektrische Feldstärke, usw. Die Methoden der Vektoranalyse kommen seit lange in der multivariaten Statistik vor.

Als *Vektor* bezeichnet man ein geordnetes Punktpaar, d.h. ein Punktpaar, dessen Punkte in einer bestimmten Reihenfolge liegen. Den ersten Punkt bezeichnet man als den **Anfang**, den zweiten als das **Ende** oder die **Spitze** des Vektors. Fallen beide Punkte zusammen, wird der vector **Nullvektor** genannt und durch die gewöhnliche Null bezeichnet.

Zwei vom Nullvektor verschiedene Vektoren **a** und **b** heißen *kollinear*, wenn sie parallel sind oder zusammenfallen. Der Nullvektor ist zu jedem Vektor kollinear.

Als den *absoluten Betrag* eines von Null verschiedenen Vektors **a** bezeichnet man die Länge

$$|\mathbf{a}| = (a_1^2 + a_2^2 + \dots + a_n^2)^{\frac{1}{2}}. \quad (1)$$

Vektoren der Länge 1 heißen *Einheitsvektoren*. Ein Vektor **a** beliebiger Länge läßt sich durch Division durch seine Norm auf

die Länge 1 normieren.

Unter dem *skalaren Produkt* zweier Vektoren \mathbf{a} , \mathbf{b} von je n Komponenten a_i bzw b_i versteht man die Zahl:

$$\mathbf{a}'\mathbf{b} = a_1b_1 + a_2b_2 + \dots + a_nb_n. \quad (2)$$

Offenbar ist also

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}$$

Dyadisches Produkt

Außer dem skalaren Produkt zweier Vektoren \mathbf{a} , \mathbf{b} in der oben angeführten Form, Zeile mal Spalte, gibt es als zweite Möglichkeit multiplikativer Verknüpfung die der Form Spalte mal Zeile, das dyadische Produkt \mathbf{ab}' .

$$\mathbf{ab}' = \begin{pmatrix} a_1 \\ \cdot \\ a_m \end{pmatrix} (b_1, \dots, b_n) = \mathbf{C} \quad (3)$$

wo \mathbf{C} ist eine $m \cdot n$ - Matrix mit den Elementen

$$c_{ik} = a_i b_k.$$

also $m \cdot n$ Zahlenprodukten, eine Matrix freilich von einfacher Bauart. Das dyadische Produkt trifft man bei der Berechnung von Kovarianzmatrizen. Die Matrix $\mathbf{C} = \mathbf{ab}'$ ist somit vom kleinsten möglichen Range: d.h. $r_c = 1$.

Ist $m = n$, so ist \mathbf{C} quadratisch **singulär**.

Winkel zwischen zwei Vektoren

$$\cos\varphi = \frac{\mathbf{a}'\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}$$

Der Winkel zwischen den Vektoren \mathbf{a} und \mathbf{b} läßt sich definieren nach (4)

Wenn $\mathbf{a}'\mathbf{b}$ Null ist, sind die Vektoren \mathbf{a} and \mathbf{b} zueinander *orthogonal*; ist $\mathbf{a} = k\mathbf{b}$ sind sie einander *parallel*.

Matrizen und Matrizenoperationen

Definitionen

Ein rechteckiges Schema von Zahlen nennt man eine *Matrix*:

$$\begin{array}{ccccccc}
 a_{11} & a_{12} & \cdot & \cdot & \cdot & a_{1n} & \\
 a_{21} & a_{22} & \cdot & \cdot & \cdot & a_{2n} & \\
 & & \cdot & \cdot & \cdot & & \\
 & & \cdot & \cdot & \cdot & & \\
 a_{m1} & a_{m2} & \cdot & \cdot & \cdot & a_{mn} &
 \end{array} \quad (5)$$

Ist $m = n$, so heißt diese Matrix **quadratisch** und die Zahl m beziehungsweise n ihre *Ordnung*. Im allgemeinen Fall ist die Matrix *rechteckig*, d.h. vom Typ (m, n) . Die Zahlen, aus denen sich die Matrix zusammensetzt, werden ihre *Elemente* genannt. Die Matrix wird auch durch einen Buchstaben bezeichnet, z B. **A**.

Die Elemente sind mit zwei Indizes versehen, deren erster die Zeile (Zeilenindex) und deren zweiter die Spalte (Spaltenindex) bezeichnet, in der das entsprechende Element steht. Die Determinante von **A** bezeichnet man mit $|\mathbf{A}|$ oder $\det \mathbf{A}$.

Eine Matrix, die nur aus einer Spalte besteht, wird *Spaltenmatrix* und (siehe auch die Definition eines Vektors) eine Matrix, die nur aus einer Zeile besteht, wird *Zeilenmatrix* genannt.

Eine quadratische Matrix

$$\begin{array}{cccc}
 d_1 & 0 & \dots & 0 \\
 0 & d_2 & \dots & 0 \\
 & & \cdot & \cdot & \cdot \\
 0 & 0 & \dots & d_n
 \end{array}$$

bei der alle Elemente, die außerhalb der Hauptdiagonalen stehen, gleich Null sind, heißt *Diagonalmatrix*.

Addition und Multiplikation von Matrizen

Gegeben sind zwei Matrizen gleichen Typs **A** und **B**. Die Matrix **C** vom gleichen Typ heißt die *Summe der Matrizen A und B*, wenn ihre Elemente die Summen der entsprechenden Elemente der gegebenen Matrizen sind:

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \quad (7)$$

wenn

$$c_{ik} = a_{ik} + b_{ik} \quad (i = 1, 2, \dots, m; k = 1, 2, \dots, n)$$

Nach dieser Definition kann man nur Matrizen gleichen Typs addieren. Die Matrizenaddition (und Subtraktion) kommutativ und assoziativ ist:

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$

Dabei sind **A**, **B**, und **C** beliebige Matrizen gleichen Typs.

Multiplikation einer Matrix mit einer Zahl

$$\mathbf{C} = t\mathbf{A}$$

gibt die Matrix \mathbf{C} deren Elemente die Produkte der entsprechende Elemente von \mathbf{A} mit der Zahl t sind.

Die Matrix \mathbf{C} ist das Produkt der beiden Matrizen \mathbf{A} and \mathbf{B} wenn das Element c_{ik} , das in der i -ten Zeile und der k -ten Spalte der Produktmatrix steht, gleich dem *inneren Produkt* der i -ten Zeile von \mathbf{A} mit der k -ten Spalte von \mathbf{B} ist:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

(8)

($i = 1, 2, \dots, m; j = 1, 2, \dots, q$).

Die Multiplikation zweier Matrizen ist nur dann ausführbar, wenn die Anzahl der Spalten des ersten Faktors gleich der Anzahl der Zeilen des zweiten ist. Die Matrizenmultiplikation ist nicht immer kommutativ; d.h. \mathbf{AB} gibt nicht unbedingt die gleiche Matrix wie \mathbf{BA} . So ist z. B.

$$\begin{array}{cccccc} 1 & 2 & 2 & 0 & 8 & -2 \\ & & & & = & \\ 3 & 4 & 3 & -1 & 18 & -4 \end{array}$$

aber

$$\begin{array}{cccccc} 2 & 0 & 1 & 2 & 2 & 4 \\ & & & & = & \\ 3 & -1 & 3 & 4 & 0 & 2 \end{array}$$

Ist $\mathbf{AB} = \mathbf{BA}$ so sagt man, die Matrizen \mathbf{A} und \mathbf{B} seien *vertauschbar*.

Eine derartige Relation kann man leicht für die beiden Matrizen \mathbf{A} und \mathbf{B} unten bekräftigen:

$$\mathbf{A} = \begin{array}{cc} 1 & 2 \\ -2 & 0 \end{array} \quad \mathbf{B} = \begin{array}{cc} -3 & 2 \\ -2 & -4 \end{array}$$

Die Matrizenmultiplikation ist *associativ* und *distributiv* (bezüglich der Matrizenaddition):

1. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
2. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$
3. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

Eine häufig angewendete Matrizenoperation ist der Uebergang zur *transponierten* oder *gespiegelten Matrix* \mathbf{A}' durch Vertauschen von Zeilen und Spalten. z B.

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}; \quad \mathbf{A}' = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \\ c_1 & c_2 \end{pmatrix}$$

Beispiel:

$$\mathbf{A} = \begin{pmatrix} 5 & -3 & 1 \\ 2 & 0 & -1 \\ -2 & 3 & -4 \end{pmatrix}$$

und

$$\mathbf{A}' = \begin{pmatrix} 5 & 2 & -2 \\ -3 & 0 & 3 \\ 1 & -1 & -4 \end{pmatrix}$$

Eine quadratische Matrix heißt *symmetrisch*, wenn sie ihrer Transponierten gleich ist.

Eine quadratische Matrix heißt *schiefssymmetrisch* wenn sie ihrer Transponierten entgegengesetzt gleich ist.

$$\mathbf{A} = -\mathbf{A}' \quad (10)$$

Beispiel:

$$\mathbf{A} = \begin{pmatrix} 0 & 2 & 4 \\ -2 & 0 & -1 \\ -4 & 1 & 0 \end{pmatrix}$$

und

$$\mathbf{A}' = \begin{pmatrix} 0 & -2 & -4 \\ 2 & 0 & 1 \\ 4 & -1 & 0 \end{pmatrix}$$

Jede quadratische Matrix \mathbf{A} ist zerlegbar in die Summe eines symmetrischen und eines schiefssymmetrischen Anteiles:

$$\mathbf{A} = \mathbf{A}_S + \mathbf{A}_A \quad (11)$$

mit

$$\mathbf{A}_S = \frac{1}{2}[\mathbf{A} + \mathbf{A}']$$

und

$$\mathbf{A}_A = \frac{1}{2}[\mathbf{A} - \mathbf{A}'] \quad (12)$$

Potenzen

Durch p -malige Multiplikation einer quadratischen Matrix \mathbf{A} mit sich selbst entsteht die p -te Potenz \mathbf{A}^p mit positiv ganzem p . Besonders einfach gestaltet sich das Potenzieren von *Diagonalmatrizen* $\mathbf{D} = \text{Diag}(d_i)$. Das gilt hier auch noch für nicht ganze Exponenten, beispielsweise:

$$\mathbf{D}^{\frac{1}{2}} = \text{Diag}(\sqrt{d_i}), \quad (13)$$

die positiv genommenen Quadratwurzeln aus den d_i .

Im allgemeinen

$$\mathbf{A}^p \mathbf{A}^q = \mathbf{A}^{p+q}. \quad (14)$$

Matrizen mit der bemerkenswerten Eigenschaft $\mathbf{G}^2 = \mathbf{G}$ werden *idempotent* genannt.

Kehrmatrix

Inversion einer quadratischen symmetrischen Matrix

Für jede reguläre Matrix \mathbf{A} existiert eine eindeutig bestimmte inverse Matrix \mathbf{A}^{-1} mit der Eigenschaft

$$\mathbf{A}^{-1} \mathbf{A} = \mathbf{A} \mathbf{A}^{-1} = \mathbf{I}. \quad (15)$$

Für zwei reguläre Matrizen \mathbf{A} und \mathbf{B} hat man, daß

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}. \quad (16)$$

Für drei Dimensionen:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} b_2 c_3 - b_3 c_2 & a_3 c_2 - a_2 c_3 & a_2 b_3 - a_3 b_2 \\ b_3 c_1 - b_1 c_3 & a_1 c_3 - a_3 c_1 & a_3 b_1 - a_1 b_3 \\ b_1 c_2 - b_2 c_1 & a_2 c_1 - a_1 c_2 & a_1 b_2 - a_2 b_1 \end{pmatrix} \quad (17)$$

Die quadratische Matrix \mathbf{A} heißt *obere (untere) Dreiecksmatrix*, wenn alle Elemente unterhalb (oberhalb) der Hauptdiagonalen gleich Null sind. Die Determinante einer Dreiecksmatrix ist das Produkt ihrer Diagonalelemente.

Orthogonale Matrizen

Eine bedeutsame Klasse quadratischer Matrizen bilden die *orthogonalen*. Diese sind gekennzeichnet dadurch, daß β ihre Spaltenvektoren ein System *orthogonaler Einheitsvektoren* bilden: d.h. $\mathbf{a}'_i \mathbf{a}_k = \delta_{ik}$. Hier ist δ_{ik} das Kronecker-Delta, das die Zahl 0 für $i \neq k$ und 1 für $i = k$ besitzt.

Diagonalmatrix

Eine quadratische Matrix, deren sämtliche Elemente außerhalb der Hauptdiagonalen Null sind, bei beliebigen Diagonalelementen d_i , wird *Diagonalmatrix* genannt.

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \end{pmatrix}$$

$$0 \quad \dots \quad 0 \quad \dots \quad d_n$$

Sind alle Diagonalelemente gleich 1, so hat man die *Einheitsmatrix*, \mathbf{I} .

Lineare Abhängigkeit

Gegeben sei ein System von p Vektoren \mathbf{a}_k zu je n Komponenten. Diese Vektoren werden linear abhängig genannt, wenn es p Konstanten c_k gibt, die nicht sämtlich verschwinden sollen, derart daß eine lineare Beziehung der Form

$$c_1 \mathbf{a}_1 + \dots + c_p \mathbf{a}_k = \mathbf{0} \quad (18)$$

besteht. Hier ist $\mathbf{0}$ der Nullvektor.

Der *Rang* einer mn -Matrix \mathbf{A} läßt sich nun auffassen als das System ihrer n Spaltenvektoren oder auch als das ihrer m Zeilenvektoren.

Eine quadratische Matrix wird *singulär* genannt, wenn ihre Spalten (und Zeilen) linear abhängig sind; andernfalls heißt sie *nichtsingulär*. Der Unterschied $d = n - r$ wird *Defekt* oder *Rangabfall* oder auch *Nullität* der n -reihigen Matrix genannt.

Determinanten

Einer quadratischen Matrix \mathbf{A} ist ihre Determinante als eine nach bestimmter Vorschrift aus den Elementen a_{ik} berechenbare Zahl zugeordnet.

$$\begin{aligned} \det \mathbf{A} = |\mathbf{A}| &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \\ &= a_{11}a_{22} - a_{12}a_{21} \end{aligned} \quad (19)$$

für zwei Dimensionen. Eine Determinante wird genau Null, wenn die Zeilen oder Spalten der Matrix linear abhängig sind, d.h., wenn die Matrix *singulär* ist. Eine *singuläre* Matrix ist somit gekennzeichnet durch $|\mathbf{A}| = 0$. Für *nichtsinguläres* \mathbf{A} gilt stets $|\mathbf{A}| \neq 0$.

Die Determinante $\det(\mathbf{AB})$ eines Matrizenprodukts \mathbf{AB} zweier quadratischen Matrizen ist gleich dem Produkt der Determinante der beiden Faktoren:

$$\det(\mathbf{AB}) = \det \mathbf{A} \cdot \det \mathbf{B}$$

Es gilt daher auch

$$\det(\mathbf{AB}) = \det(\mathbf{BA}).$$

Die *Spur* einer quadratischen Matrix ist die Summe der

Hauptdiagonalelemente a_{ii} .

Für die Spur des Produktes zweier nicht notwendig quadratischer Matrizen, einer $m \times n$ Matrix \mathbf{A} und einer $n \times m$ Matrix \mathbf{B} :

$$\text{Spur } (\mathbf{AB}) = \text{Spur } (\mathbf{BA})$$

WICHTIGE BEMERKUNG: Ein Matrizenprodukt kann Null sein, ohne daß einer der beiden Faktoren selbst Null ist.

Quadratische Formen

Eine quadratische Form hat das folgende allgemeine Format:

$$Q = \mathbf{x}'\mathbf{A}\mathbf{x}.$$

Sie ist ein homogener Ausdruck zweiten Grades mit reellen Koeffizienten a_{ik} . Die symmetrische Matrize \mathbf{A} wird Matrix der Form genannt; sie legt die Form d.h. die quadratische Funktion der n Variablen x_i eindeutig fest. Quadratische Formen treten in den Anwendungen vielfach als *Energieausdrücke* auf, beispielsweise als kinetische und potentielle Energie mechanischer elastischer Systeme, als Formveränderungsarbeit in der Baustatik, und in der multivariaten Statistik: z. B. der generalisierte statistische Abstand von Mahalanobis.

Positiv definite quadratische Formen

$$Q = \mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \text{ für jedes } \mathbf{x}.$$

Derartige Formen heißen *positiv definit*. Falls $Q = 0$ für nur $\mathbf{x} = \mathbf{0}$ sagt man *eigentlich definit*, hingegen *positiv semidefinit* wenn es Werte $\mathbf{x} \neq \mathbf{0}$ gibt, für die $Q = 0$. Diese Definitionen sind für die allgemeine Multivariatanalyse besonders wichtig.

Die Eigenwertaufgabe

Eigenwerte und Eigenvektoren

Es handelt sich darum, zu einer quadratischen, sonst aber beliebigen (reellen oder komplexen) Matrix \mathbf{A} Vektoren \mathbf{x} derart zu suchen, daß der mit \mathbf{A} transformierte Vektor $\mathbf{y} = \mathbf{A}\mathbf{x}$ dem Ausgangsvektor \mathbf{x} proportional, ihm parallel ist; also:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad (20)$$

mit einem zunächst noch unbestimmten Parameter λ .

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}. \quad (21)$$

Die Matrix dieses Gleichungssystems wird oft die *charakteristische Matrix* der Matrix \mathbf{A} genannt. Hier, \mathbf{I} ist die Einheitsmatrix. Die Lösung des Gleichungssystems erfolgt durch die charakteristische Determinante:

$$D(\lambda) = \det (\mathbf{A} - \lambda \mathbf{I}) = 0 \quad (22)$$

die genau n Wurzeln besitzt; diese sind die Eigenwerte. Die dazu hörenden Eigenvektoren \mathbf{x}_i findet man durch die Lösung der Gleichung:

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i \quad (i = 1, 2, \dots, n).$$

Definition

1. Eine Matrix hat wenigstens einen Eigenwert $\lambda = 0$, wenn sie singularär ist, $\det \mathbf{A} = 0$.
2. Die Eigenvektoren \mathbf{x}_i sind nicht eindeutig. Man kann sich von der Unbestimmtheit des Eigenvektors befreien, indem man ihn in passender Weise *normiert*. Z. B. man macht seine Euklidische Norm zu 1:

$$\mathbf{x}'_i \mathbf{x}_i = 1 \quad (23)$$

3. Eigenvektoren zu verschiedenen Eigenwerten sind stets linear unabhängig.
4. Zu einem einfachen Eigenwert gibt es genau einen linearen unabhängigen Eigenvektor, also eine *Eigenrichtung*.

Das statistische Verfahren der Hauptkomponentenanalyse basiert auf der oben angeführten Eigenwertaufgabe. Die Eigenwertaufgabe kann jedoch verallgemeinert werden.

Die allgemeine Eigenwertaufgabe

In den Anwendungen, insbesondere in der multivariaten Varianzanalyse, der Methode der kanonischen Variablen, und in der Schwingungstechnik, kommt neben der bisher behandelten *speziellen Eigenwertaufgabe* eine Verallgemeinerung vor, die als *allgemeine Eigenwertaufgabe* bezeichnet wird, nämlich

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{B}\mathbf{x} \quad (24)$$

beziehungsweise

$$(\mathbf{A} - \lambda \mathbf{B})\mathbf{x} = \mathbf{0} \quad (25)$$

mit zwei n -reihig quadratischen Matrizen \mathbf{A} und \mathbf{B} , und die das spezielle Problem als Sonderfall mit $\mathbf{B} = \mathbf{I}$ enthält. Als Bedingung für die Existenz nichttrivialer Lösungen \mathbf{x} (den generalisierten Eigenvektoren) haben wir hier die verallgemeinerte charakteristische Gleichung

$$\det(\mathbf{A} - \lambda\mathbf{B}) = 0. \quad (26)$$

Spektralzerlegung diagonalähnlicher Matrizen

Hat man für eine diagonalähnliche Matrix \mathbf{A} einen der Eigenwerte, etwa λ_1 nebst zugehörigem Rechts- und Linksvektor \mathbf{x}_1 beziehungsweise \mathbf{y}_1 bestimmt, so läßt sich aus \mathbf{A} eine neue Matrix \mathbf{B} gleicher Reihenzahl n gewinnen, die bei gleichen Eigenvektoren die selben Eigenwerte λ_i wie \mathbf{A} besitzt bis auf den Wert λ_1 , der in Null übergeht.

$$\mathbf{B} = \mathbf{A} - \lambda_1 \mathbf{x}_1 \mathbf{y}'_1. \quad (27)$$

Man spricht von einer *Deflation* oder *Schrumpfung* der Matrix. Bei einer Matrix von Range r erhält man nach genau r oben angegebenen Schritten eine Matrix von Range Null, das ist aber die Nullmatrix. Wir gewinnen so die Beziehung:

$$\mathbf{A} = \lambda_1 \mathbf{x}_1 \mathbf{y}'_1 + \lambda_2 \mathbf{x}_2 \mathbf{y}'_2 + \dots + \lambda_n \mathbf{x}_n \mathbf{y}'_n$$

(28)

Für eine n -reihige diagonalähnliche Matrix \mathbf{A} vom Range r mit den Eigenwerten λ_i und den biorthonormierten Rechts- und Linkseigenvektoren \mathbf{x}_i bzw \mathbf{y}_i gilt die oben angeführte Spektralzerlegung. Die Matrix erscheint aufgebaut aus r dyadischen Produkten, d.h. Matrizen vom Range 1.

In der anglophonen statistischen Literatur spricht man vom Eckart-Young-Ansatz, der aber eigentlich eine Verallgemeinerung der Gleichung (28) ist, da dieser Lehrsatz die Dekomposition einer rechteckigen Matrix behandelt. Die allgemeine Lösung des Problems der Zerlegung einer rechteckigen Matrix wurde von Eckart und Young (1936) gemacht.

Vierter Teil. Einführung in die Multivariatanalyse

Einleitendes

Zuerst soll man die allgemeinen Fragestellungen und Bedingungen betrachten. Unter diesen trifft man auf die Wishart'sche Verteilung und die multivariate Gauß'sche Verteilung. Mit diesen sind einschlägige geometrische Begriffe verknüpft, u.a., die Tatsache, daß die meisten multivariatstatistischen Verfahren auf der multivariaten Normalverteilung basieren. Folgende Fragen werden besprochen: die Korrelation zwischen Variablen und die Geometrie der Streuellipsoide.

Die Methode der **Hauptkomponentenanalyse (HKA)** (als Hauptbeispiel R-Mode-Analyse): Deutung der Ergebnisse einer HKA.

Die mit der HKA mathematisch verwandte Faktoranalyse und ihre Varianten werden danach behandelt: (die als *echt* betrachtete Faktoranalyse - d.h. die klassische Darstellung der Psychometriker); die Ersatzvariante (eine Art von (Komponentenfaktoranalyse mit oder ohne Rotation der Eigenvektoren)).

Die **Pfadkoeffizienten** von S. Wright und ihre Anwendung bei der Analyse von Zuwachs und Form werden eingehend berücksichtigt. Danach richte ich die Aufmerksamkeit auf die s.g. Q-Modus-Klasse von analytischen Methoden: Das Verfahren von Hauptkoordinaten und das Einordnen von N Individuen im Objektraum.

Eine bedeutungsvolle Gruppe machen die Q-R-Modus-Methoden aus: Die Biplot-Methode K. R. Gabriels, die Korrespondenzanalyse nach H. O. Hirschfeld (Hartley) (späterhin neuentdeckt von R. A. Fisher und noch später von J.-P. Bénécri; und sie wird sicher dereinst nochmals wiederentdeckt). Der s.g. Biplot benutzt kontinuierlich verteilte Variablen während die Korrespondenzanalyse für Häufigkeitstabellen beabsichtigt ist).

Die Stabilität der Eigenvektoren. Die Analyse geschlossener Systeme in der Geochemie usw (die Logquotient-Modelle J. Aitchisons) wird erörtert.

Vergleich von zwei multivariaten Grundgesamtheiten. Die Diskriminanzanalyse, der verallgemeinerte Abstand zwischen zwei Grundgesamtheiten (Abstand zwischen multivariaten Mittelwerten) und das Testverfahren Hotellings. Taxonomische Bedeutung des statistischen Abstandes.

Die Analyse mehrerer Grundgesamtheiten folgt: Kanonische Variablen und die allgemeine Eigenwertaufgabe bei mehreren Stichproben (d.h. eine Art von Diskriminanzanalyse bei mehreren Populationen). Die guten Eigenschaften des Verfahrens; die schlechten Eigenschaften des Verfahrens (u.a. Unstabilität bei den Vektoren nach N. A. Campbell).

Die nächste Frage bezieht sich auf die Methode der kanonischen Korrelation als multivariates Modell für die Korrelationen zwischen Teilmengen. Anwendung auf Häufigkeitsbeobachtungen. Die Verbindung zwischen Artenanzahl und Ökologie.

Einleitendes

Unter **Individuen** versteht man voneinander abgrenzbare Einheiten, z. B. Menschen, Tiere, Pflanzen, Gesteinsarten, usw. Die an den Individuen gewonnenen Meßwerte sind stochastischer Natur. Sie sind Realisierungen von Zufallsgrößen. Die Meßvariablen werden als *Merkmale* bezeichnet. Körpergewicht und Körperlänge sind zwei am Mensch meßbare Merkmale. Die entsprechenden Meßresultate, z. B. 90 kg und 175 cm, heißen *Merkmalswerte*, *Beobachtungswerte*, oder *Meßwerte*. Eine mehrdimensionelle Analyse umfasst eine Reihe von Merkmalen, einen *Merkmalsvektor*. Transformierte Meßwerte nennen wir **Rechenmaßwerte**

(abk. **RMW**).

Mehrdimensionelle Normalverteilung

Die Theorie der multivariatstatistischen Analyse basiert auf bestimmten Gesetzmäßigkeiten der statistischen multidimensionalen Normalverteilung.

Normalverteilung

Eine eindimensionale (skalare) Zufallsgröße y ist normalverteilt, wenn für ihre Wahrscheinlichkeitsdichte folgendes Verhältnis (29) gilt:

$$f(y) = \frac{1}{\sqrt{(2\pi)\sigma}} \frac{e^{-(y-\mu)^2/2\sigma^2}}{2\sigma^2} \quad (29)$$

wobei μ der Mittelwert und σ die Standardabweichung sind. Die p -dimensionale (vektorielle) Zufallsgröße $\mathbf{y}' = (y_1, y_2, \dots, y_p)$ heißt normalverteilt, wenn sie die Wahrscheinlichkeitsdichte

$$f(y_1, \dots, y_p) = \frac{1}{\sqrt{(2\pi)^p \sqrt{|\Sigma|}}} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y}-\boldsymbol{\mu})} \quad (30)$$

besitzt. Der $(p,1)$ -Vektor $\boldsymbol{\mu}$ und die (p,p) -Matrix Σ sind die Parameter dieser Verteilung, d.h. der Mittelwertsvektor bzw Kovariansmatrix, die man mit $N(\boldsymbol{\mu}, \Sigma)$ bezeichnet. Hierbei wird immer Σ als eine positiv definite symmetrische Matrix angenommen.

Aus der Dichtefunktion (30) erkennt man, daß die Punkte \mathbf{y} gleicher Dichte auf einem Ellipsoid im p -dimensionellen Raum, welches einer Gleichung

$$(\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) = c^2 \quad (31)$$

genügt, liegen. Dieses Ellipsoid wird *Konzentrationellipsoid* genannt.

Die Wishart-Verteilung

Die Wishart-Verteilung $W(\Sigma, m)$ wird definiert als die mehrdimensionale Verteilung der $p \times p$ -Matrix

$$\mathbf{X} = \mathbf{y}_1 \mathbf{y}_1' + \dots + \mathbf{y}_m \mathbf{y}_m' \quad (32)$$

wobei die \mathbf{y}_i von einander unabhängige p -dimensionale Zufallsgrößen sind, die jeweils die Verteilung $N(\mathbf{0}, \Sigma)$ besitzen. Die Wishart-Verteilung ist eine Verallgemeinerung der χ^2 -Verteilung der univariaten Statistik.

Methoden für die Analyse einer einzigen Stichprobe

Die Datenmatrix: Die Datenmatrix \mathbf{X} ist eine rechteckige Matrix bei der jede Reihe p Elemente (den Variablen entsprechend) und jede Zeile n Elemente (den Individuen entsprechend) enthält. Die Datenmatrix ist der Baustein der multivariaten Morphometrie.

Hauptkomponentenanalys

Die Theorie der Hauptkomponentenanalyse (HKA) stützt sich auf die einfache Eigenwertaufgabe für quadratische symmetrische Matrizen. Obwohl Pearson schon 1900 gewisse Gedanken über die Zerlegung einer Kovarianzmatrix ausgedrückt hatte, wurde die heutige Methodik von dem amerikanischen Statistiker Harold Hotelling (1932) eingeführt, und zwar als eine Art von Faktoranalyse (siehe unten!) bei der Untersuchung pädagogischer Analysen. (Damals war die Faktoranalyse noch nicht entwickelt.) Die Hauptzüge der HKA sind wie folgt:

Gegeben sind ein Beobachtungsvektor $\mathbf{x} = (x_1, \dots, x_p)$.

$$\mathcal{E}(\bar{\mathbf{x}}) = \boldsymbol{\mu}$$

$$\bar{\mathbf{x}}$$

mit Stichprobenschätzung

und Kovarianzmatrix

$$\text{Cov}(\mathbf{x}, \mathbf{x}') = \Sigma$$

mit Stichprobenschätzung \mathbf{S}

Diese beiden Schätzungen der Populationsparameter sind erwartungstreu.

Einer HKA liegt die Lösung der charakteristischen Gleichung

$$|\mathbf{S} - \lambda \mathbf{I}| = 0 \quad (33)$$

zugrunde, wobei \mathbf{S} ist die geschätzte Kovariansmatrix einer gegebenen Stichprobe bzw die Korrelationsmatrix derselben ist. Es gibt dann

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$$

$$(i, j = 1, \dots, p) \quad (34)$$

Zum Beleg des Verfahrens betrachten wir die (2 x 2)-Matrix \mathbf{R} . Die erwünschte Lösung wird geschrieben:

$$\begin{aligned} (r_{11} - \lambda)u_1 + r_{12}u_2 &= 0 \\ r_{21}u_1 + (r_{22} - \lambda)u_2 &= 0 \end{aligned} \quad (35)$$

Die allgemeine Form dieser Verhältnisse ist

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}' \quad (36)$$

Diese Gleichung teilt uns mit, daß β die symmetrische quadratische Matrix \mathbf{R} als das Produkt der drei Matrizen **Eigenvektoren, Eigenwerte und transponierte Eigenvektoren** ausgedrückt werden kann, eine sehr bequeme Zerlegungsmöglichkeit.

BEISPIELE DER HAUPTKOMPONENTANALYSE

1. Männchen und Weibchen von Schildkröten

Dimensionen: Länge, Höhe und Breite des Panzers
Beobachtungen: 24 Männchen und 24 Weibchen

Anzahl der

1	93.0	74.0	37.0
2	94.0	78.0	35.0
3	96.0	80.0	35.0
4	101.0	84.0	39.0
5	102.0	85.0	38.0
6	103.0	81.0	37.0
7	104.0	83.0	39.0
8	106.0	83.0	39.0
9	107.0	82.0	38.0
10	112.0	89.0	40.0
11	113.0	88.0	40.0
12	114.0	86.0	40.0
13	116.0	90.0	43.0
14	117.0	90.0	41.0
15	117.0	91.0	41.0
16	119.0	93.0	41.0
17	120.0	89.0	40.0
18	120.0	93.0	44.0
19	121.0	95.0	42.0
20	125.0	93.0	45.0
21	127.0	96.0	45.0
22	128.0	95.0	45.0
23	131.0	95.0	46.0
24	135.0	106.0	47.0
25	98.0	81.0	38.0
26	103.0	84.0	38.0
27	103.0	86.0	42.0
28	105.0	86.0	40.0
29	109.0	88.0	44.0
30	123.0	92.0	50.0

31	133.0	99.0	51.0
32	123.0	95.0	46.0
33	133.0	102.0	51.0
34	133.0	102.0	51.0
35	134.0	100.0	48.0
36	136.0	102.0	49.0
37	137.0	98.0	51.0
38	138.0	99.0	51.0
39	141.0	105.0	53.0
40	147.0	108.0	57.0
41	149.0	107.0	55.0
42	153.0	107.0	56.0
43	155.0	115.0	63.0
44	155.0	117.0	60.0
45	158.0	115.0	62.0
46	159.0	118.0	63.0
47	162.0	124.0	61.0
48	177.0	132.0	67.0

Mittelwerte: 124.68750 95.43750 46.33333
Standardabw. 20.48160 12.67584 8.39284

Korrelationsmatrix

	1	2	3
1	1.0000	.9783	.9647
2	.9783	1.0000	.9606
3	.9647	.9606	1.0000

HKA der Korrelationsmatrix

Eigenwerte
.02142 .04284 2.93574
Prozent der Spur
.71395 1.42813 97.85792
Eigenvektoren

	1	2	3
1	.7479	-.3250	.5788
2	-.6574	-.4835	.5780
3	-.0920	.8128	.5753

Der erste Vektor stellt gleichwertige Kovariation dar.

Die transformierten Variablen

1	.008	.086	-4.305
2	-.013	.011	-4.179
3	-.018	-.011	-3.926
4	-.028	.021	-2.902
5	-.029	-.010	-2.893
6	.009	-.002	-3.274
7	-.004	.019	-2.835
8	.006	.013	-2.738
9	.021	-.003	-2.885
10	-.009	-.034	-1.861
11	.004	-.030	-1.891
12	.025	-.017	-1.999
13	.000	.005	-1.237
14	.009	-.039	-1.423
15	.001	-.047	-1.345
16	-.003	-.069	-1.092
17	.034	-.061	-1.474
18	-.003	-.012	-.691
19	-.009	-.071	-.722
20	.022	-.009	-.332
21	.010	-.039	-.001
22	.023	-.034	-.030
23	.038	-.024	.232
24	-.026	-.104	1.403
25	-.020	.035	-3.399
26	-.016	-.006	-2.922

27	-.037	.059	-2.296
28	-.023	.012	-2.434
29	-.024	.063	-1.615
30	.011	.106	.080
31	.010	.038	1.229
32	-.005	.002	-.155
33	-.013	.014	1.463
34	-.013	.014	1.463
35	.012	-.033	1.003
36	.006	-.036	1.374
37	.039	.033	1.344
38	.037	.022	1.471
39	.004	.005	2.320
40	.007	.041	3.315
41	.028	.003	3.098
42	.048	.009	3.409
43	-.013	.080	4.953
44	-.024	.004	4.757
45	.004	.050	4.981
46	-.015	.043	5.381
47	-.041	-.054	5.761
48	-.031	-.046	7.817

2. *Zuckerkrankheit bei englischen Frauen*: Die Beobachtungen sind Todesfälle/10⁶ Personen während der Jahre 1937-1957. Die Einflüsse der während des zweiten Weltkrieges eingeführten Nahrungsveränderungen werden dargelegt.

Die Variablen:

1. Die Sterblichkeit/Million
2. Fibergehalt
3. Fett
4. Zucker
5. Eiweiß
6. Kohlenhydrate
7. Kkal/Tag

Anzahl der Dimensionen= 7 Kostarten

Beobachtungsvektoren: = 19 Jahre

Mittelwertsvektor

112.7895 0.3216 43.9474 92.4737 84.4737 422.0300 51.5790

Standardabweichungen

20.4900 0.1914 4.4904 16.1051 3.0617 12.7541 99.8478

Kovarianzmatrix

	1	2	3	4	5	6	7
1	419.842	-0.035	-33.512	-206.117	-9.506	-181.500	-1580.205
2		0.037	-0.4860	-1.715	0.396	0.3828	-6.5304
3			20.163	57.082	-8.807	7.3889	316.7544
4				259.374	-24.681	73.2778	1359.211
5					9.3743	14.1111	-44.6784
6						162.666	818.889
7							9969.591

Die Korrelationsmatrix

	1	2	3	4	5	6	7
1	1.0000	-0.0088	-0.3642	-0.6246	-0.1515	-0.6945	-0.7224
2		1.0000	-0.5654	-0.5563	0.6764	0.1568	-0.3416
3			1.0000	0.7893	-0.6406	0.1290	0.7065
4				1.0000	-0.5005	0.3567	0.8452
5					1.0000	0.3614	-0.1461
6						1.0000	0.6430
7							1.0000

(Statistisch wichtige Werte mit fetter Schrift markiert.)

Hauptkomponentenanalyse der Korrelationsmatrix

Eigenwerte
 3.72469 2.22504 0.39953 0.27104 0.21448 0.11539 0.04982
Prozent der Spur

53.21 31.79 5.71 3.87 3.06 1.65 0.71

(Der Hauptteil der "Information" kommt bei den zwei ersten Vektoren vor = 85% der Spur.)

Eigenvektoren

	1	2	3	4	5	6	7
1	-0.3542	0.4137	-0.1521	0.6128	-0.2416	0.4963	-0.0063
2	-0.2985	-0.4234	0.7941	0.1699	0.1185	0.2321	0.0654
3	0.4446	0.2063	0.4159	0.3191	-0.4893	-0.3637	-0.3364
4	0.4947	0.0403	0.0625	-0.2109	0.1849	0.6832	-0.4520
5	-0.2437	-0.5388	-0.2480	-0.1981	-0.6494	0.1234	-0.3357
6	0.2389	-0.5223	-0.3269	0.6374	0.3386	-0.1186	-0.1683
7	0.4760	-0.2059	-0.0314	0.0622	-0.3429	0.2672	0.7329

(Der siebente Vektor ist ein Invarianzvektor: hier haben wir ein invariantes Verhältnis zwischen den Variablen (3), (4), (5) (positive) und (7)(negativ).)

Die Rechenmaßwerte

	1	2	3	4	5
1	1.2809	4.1018	-0.1429	0.5046	0.0820
2	-3.5515	5.4564	-0.0761	-0.1606	0.0014
3	-3.6488	2.8529	-0.4303	0.0898	0.0122
4	-5.6234	-0.0710	0.8389	-0.1548	-0.2398
5	-5.0765	0.2577	0.3604	-0.0321	0.1401
6	-2.0781	-1.2601	-0.1102	0.3326	-0.0243
7	-3.2235	-1.4261	-1.1112	-0.0764	-0.2985
8	-4.3612	-2.0419	0.1901	-0.2612	0.2062
9	-2.9892	-3.0624	-0.3572	-0.1428	0.2303
10	0.0082	-3.0376	-0.1009	0.1177	0.0909
11	2.1106	-2.5026	0.0634	0.5883	-0.0064
12	0.5911	-0.8606	0.4127	0.0123	-0.5626
13	2.4156	0.2021	0.3404	0.3607	-0.0649
14	1.1638	0.3631	0.1243	-0.0926	0.3257
15	2.7938	-0.0203	0.2556	-0.1505	0.2678
16	5.4217	-0.2469	0.0515	-0.0663	0.0880
17	4.8545	0.6229	-0.1027	-0.1584	0.0407
18	4.6316	0.1645	-0.1353	-0.3729	-0.1655
19	5.2804	0.5080	-0.0704	-0.3374	-0.1233

Prüfung der Gleichheit der letzten N-1 Eigenwerte gibt
 $\chi^2 = 107.9519$ bei 20 Freiheitsgraden.

Die Varimaxfaktormatrix

VAR	Komm.	1	2	3	4
1	0.9591	-0.9241	-0.0195	-0.0848	-0.3122
2	0.9906	-0.0814	-0.3782	0.9145	0.0673
3	0.9277	0.3584	0.8587	-0.2150	0.1253
4	0.9286	0.6672	0.5746	-0.3481	0.1796
5	0.9022	0.0913	-0.8028	0.4104	0.2846
6	0.9724	0.4206	-0.0969	0.0986	0.8811
7	0.9398	0.7137	0.3921	-0.1850	0.4923

VARIANZ 30.41 28.84 17.48 17.85

Rechenmaßwerte bezüglich der Varimaxmatrix

	1	2	3	4	5
1	-1.7583	1.6682	-1.0072	0.4707	-1.7878
2	-1.3821	0.3548	-1.0817	-2.0258	-1.4612
3	-1.5507	-0.3415	-1.0738	-0.4462	-0.4813
4	-0.3489	-0.0859	2.1179	-1.5494	0.8251
5	-0.7494	-0.3468	1.1168	-0.8774	0.1719
6	-0.8845	-0.2872	0.4330	1.0778	0.4985
7	-0.4811	-2.3271	-1.6313	0.8162	1.8050
8	0.2850	-1.2897	1.0809	-0.6749	1.1517
9	0.2849	-1.8175	0.1546	0.4549	1.4783
10	0.3141	-0.6495	0.5193	1.1515	0.8251
11	-0.4104	0.6727	0.7926	2.2524	0.4985
12	0.3856	0.4291	0.9535	-0.0438	0.8251
13	-0.3015	1.3609	0.6031	0.7866	-0.4813

14	0.3775	0.2924	0.0071	-0.3312	-0.8079
15	0.9113	0.5468	0.1644	-0.3518	-0.8079
16	1.1627	0.7783	-0.3909	0.3118	-0.8079
17	1.0418	0.5645	-0.8813	-0.0618	-0.8079
18	1.5578	0.0828	-0.9450	-0.4820	-0.1547
19	1.5460	0.3948	-0.9318	-0.4777	-0.4813

Quellen: Program PCA3.exe: Die Daten befinden sich in
Diabetes.dat

Bemerkungen zur Analyse: Die erste HKA stellt eine Relation zwischen (1), (3), (4) und (7) dar = Sterblichkeit als Funktion der anderen Variablen.

Sonderfragen der Hauptkomponentenanalyse

1. Anzahl der gültigen Hauptkomponenten

Diese Frage ist von praktischem Belang. Jolliffe (1986, S. 95) schlägt den folgenden Schneidepunkt vor:

(37)

$$\lambda_i^* = 0.7 \frac{(\sum \lambda_i)}{p}$$

wo die λ_i die Eigenwerte der Kovarianz- bzw Korrelationsmatrix der Stichprobe sind und p die Anzahl der Variablen ist. Dieses Verfahren ist sehr leicht zu berechnen. Es gibt auch andere Möglichkeiten für die Bestimmung eines Schneidepunktes.

2. Gewichtete Hauptkomponenten

In der Regel wird die HKA dadurch gewichtet, da β die Eigenvektoren mit den entsprechenden Quadratwurzeln der Eigenwerte multipliziert werden.

$$\mathbf{A} = \mathbf{U}_k \mathbf{L}_k^{\frac{1}{2}}. \quad (38)$$

Die Zeilen von \mathbf{A} sind dann die gesuchten Hauptkomponenten. Diese Gewichtung bedeutet, daß die Quadrate jeder Zeile von \mathbf{A} dem entsprechenden Eigenwert gleicht.

3. Robuste Hauptkomponenten

Daten, die von der multivariaten Normalverteilung stark abweichen, sollen eigentlich durch eine robuste Methode analysiert werden. Campbell (1979) hat die Frage eingehend berücksichtigt und eine verwendbare Methode entwickelt. Diese wird jetzt an einem Beispiel gezeigt.

Beispiel: Robuste Hauptkomponentenanalyse

Morphfrequenzen einer Ostrakodenart

Hier mache ich einen Vergleich zwischen der robusten Bestimmungsmethode und der üblichen Methode bei der Hauptkomponentenanalyse von Morphfrequenzen

```
(1) Übliche Berechnungsweise
Anzahl der Dimensionen = 4
Beobachtungsvektoren = 10
Datenmatrix
Morph 1 Morph 2 Morph 3 [Va]
1 0.73 0.21 0.06 0.25
2 0.33 0.66 0.01 0.30
3 0.16 0.79 0.05 0.35
4 0.26 0.58 0.16 0.35
5 0.13 0.74 0.13 0.50
6 0.39 0.57 0.04 0.40
7 0.22 0.71 0.07 0.32
8 0.08 0.88 0.04 0.31
9 0.23 0.72 0.05 0.34
10 0.29 0.65 0.05 0.39
Kovarianzmatrix
2 3 4
1 0.0334 -0.0320 -0.0014 -0.0060
2 -0.0320 0.0327 -0.0006 0.0046
3 -0.0014 -0.0006 0.0020 0.0013
4 -0.0060 0.0046 0.0013 0.0046
Übliche Eigenwerte
2 3 4
0.065993 0.004988 0.001747 0.000003
Prozentzahl der Spur
90.73514 6.85842 2.40235 0.00409
Eigenvektoren
1 2 3 4
1 0.7072 -0.1840 0.3572 -0.5818
2 -0.6965 -0.3294 0.2644 -0.5801
3 -0.0107 0.5087 -0.6452 -0.5700
4 -0.1212 0.7739 0.6215 -0.0106
(2) Robuste Bestimmung nach Campbell (1979)
Robuste Eigenwerte
1 2 3 4
0.018418 0.0060405 0.0020708 0.0000007
Robuste Eigenvektoren
1 2 3 4
1 0.7046 -0.2214 0.3376 -0.5834
2 -0.7091 -0.1869 0.3554 -0.5797
3 -0.0008 0.3965 -0.7207 -0.5687
4 0.0273 0.8711 0.4901 -0.0139
```

 Quellen: GENSTAT-Program Robpca.gpi (Daten einbegriffen)

Einflußreiche und atypische Beobachtungsvektoren und HKA

Eastment und Krzanowski (1982) und Krzanowski (1987) haben eine Synthese von analytischen Methoden zusammengestellt, um die Sicherheit einer HKA zu befestigen. Besonders die Frage von atypischen (aber wissenschaftlich richtigen) Beobachtungen usw sind berücksichtigt worden. Ich gebe nun ein Beispiel.

Beispiel: Die Anatomie einer Hauptkomponentenanalyse

Die *ad hoc* Methode von Wojtek Krzanowski (1987, 1988).
 Die Beobachtungen bestehen aus geochemischen Bestimmungen von vulkanischem Stoff in Tiefseematerial
 Die O.D.P Abschnitte:
 Elemente: SiO₂ TiO₂ Al₂O₃ FeO MnO MgO CaO K₂O Na₂O

HKA mittels s.g. JACK-KNIFING (=Schnappmesserverfahren).

VARIABLEN = 9
Beobachtungsanzahl = 26

Eigenwerte der Kovarianzmatrix

6.6688 1.3601 0.4101 0.3024 0.1657 0.0668 0.0207 0.0053

Hauptkomponenten der Kovarianzmatrix

1	0.37665	-0.31464	-0.20678	-0.37670	-0.32302	-0.37177	-0.38276
0.29812	0.31069						
2	-0.17074	0.41676	0.70876	-0.15383	-0.20135	0.00770	-0.04268
0.30915	0.36491						
3	0.12138	0.18067	-0.03303	-0.07571	-0.11950	-0.11714	0.06350
-0.80784	0.51103						
4	-0.03909	-0.05184	0.03486	0.07913	-0.89171	0.06863	0.20700
-0.11770	-0.36298						
5	0.14075	0.69509	-0.20918	0.13075	0.00624	-0.57209	-0.10742
0.05190	-0.30761						
6	-0.20243	0.33165	-0.58893	0.19829	-0.19741	0.43168	-0.06748
0.26554	0.40309						
7	0.27652	0.30573	-0.09368	-0.74153	0.06571	0.43291	0.06392
-0.05469	-0.26948						
8	-0.17678	-0.05226	-0.18814	-0.26173	0.04452	-0.32878	0.82132
0.21298	0.17919						
9	0.80263	0.05138	0.13348	0.38426	0.01968	0.18479	0.33171
0.16592	0.12478						

Das Program gibt auch die Hauptkoordinaten nach Gower. Diese sind wertvoll für die Herstellung von Diagrammen-

Hauptkoordinaten

-3.0446	-1.5059	-0.0476	0.3689	-0.4820	-0.1440	0.0619
-0.0686	0.0016					
-3.4959	-2.1887	-0.2997	-0.0944	-0.2576	0.2442	0.2013
0.0001	-0.0085					
-2.8902	-1.5961	0.0805	-0.0550	-0.2653	0.0927	-0.2424
-0.0745	0.0034					
2.6780	-1.1980	1.1211	0.2239	0.2471	-0.3784	0.1147
0.1357	0.0092					
-0.6038	-1.4990	0.3639	-0.2646	0.2333	-0.2091	-0.0711
-0.0671	0.0066					
3.6894	-1.9192	1.1564	1.0520	0.4492	0.2518	-0.0407
0.0580	-0.0073					
-2.6270	-1.1848	-0.1686	0.0012	0.6454	-0.1245	-0.1284
-0.0849	-0.0015					
-1.2384	-0.1830	-0.8233	-0.3542	0.4681	-0.0911	-0.2796
0.0830	-0.0050					
-2.3161	-0.1045	-1.1109	0.6967	-0.4775	-0.2303	-0.0792
0.1906	0.0040					
1.5307	0.8560	-1.5420	0.8396	0.8621	0.1993	0.0257
-0.0469	0.0064					
-1.7204	-0.0159	0.3256	-0.7867	0.5558	-0.0407	0.3630
-0.0396	0.0028					
2.3140	1.5527	0.7995	0.2698	-0.2436	-0.0891	-0.2297
-0.0613	-0.0051					
0.0735	1.2644	0.1174	0.0027	0.4705	0.2273	0.0047
-0.0096	0.0009					
-2.1354	1.4670	0.0688	0.2899	-0.1881	-0.3028	0.1242
-0.0139	0.0054					
5.6739	-0.6939	-1.0969	-0.3399	-0.4132	0.0573	0.1431
-0.0461	0.0003					
-1.8173	1.3559	0.4219	0.1795	-0.0701	0.2327	0.0526
0.0136	0.0010					
2.2547	0.9979	-0.1075	0.7026	-0.0365	-0.7414	0.0549
-0.1068	-0.0074					
-1.6057	1.1560	0.3347	-0.3878	-0.0658	0.0242	0.0876
-0.0060	-0.0056					
2.4263	0.0271	-0.1014	0.0327	-0.1809	0.3079	0.0181
-0.0104	-0.0014					
5.5232	-0.7858	-0.3617	-0.5514	-0.5315	0.1869	-0.0153

-0.0144	0.0038						
-0.5033	0.4134	0.1141	-0.6104	-0.2535	0.0150	-0.1633	
-0.0655	0.0020						
-2.5606	0.9694	0.1762	0.9833	-0.4258	0.3872	0.1138	
0.0188	-0.0023						
0.1481	0.8667	0.3872	-0.0979	-0.5273	0.0843	-0.0438	
-0.0006	0.0042						
-0.3680	1.0275	0.6352	-0.6089	0.1513	0.2402	-0.1276	
0.0494	0.0057						
0.3043	0.2644	-0.1712	-1.0325	-0.0814	-0.3110	0.0263	
0.1171	-0.0082						
0.3107	0.6564	-0.2718	-0.4590	0.4174	0.1114	0.0293	
0.0499	-0.0048						

Überflüssige Variablen: Prüfung

Variable abgezogen	1	2	3
1	1.2459	1.1360	1.1349
2	1.6323	2.3827	2.4814
3	1.0418	5.0962	5.0406
4	1.2217	1.1338	1.1390
5	1.4319	1.6933	1.8193
6	1.1790	1.1789	1.2285
7	1.0807	1.0731	1.0927
8	1.4543	2.3916	16,6937
9	1.5601	2.5606	6.7018

Die beiden markierten Werte deuten auf wichtige Variablen (fette Schrift) hin, deren Eliminierung zum Nachteil der Analyse führen würde.

Die Effekte der Ausschaltung von einzelnen Individuen.

Objekt abgezogen	Größter Winkel beim Ausschalten Größe der verglichenen Komponentenmannigfaltigkeiten					
	1	2	3	-3	-2	-1
1	2.2896	2.4097	2.4424	1.2821	1.2217	0.0754
2	3.6860	3.3304	2.8323	3.9072	0.3432	0.3432
3	2.2044	1.4693	1.4464	2.0632	4.0042	0.2197
4	1.9696	3.9809	12.446	5.0080	4.9898	1.1882
5	0.4183	2.0539	2.9331	1.2613	1.1504	0.2635
6	4.1138	8.7823	27.2455	3.0662	2.4871	0.5766
7	1.6097	1.9411	2.1653	2.3695	2.5163	0.0926
8	0.5029	0.6577	8.6192	3.6157	5.6021	0.3961
9	1.2791	1.3185	20.1468	4.3796	5.0359	0.9098
10	1.2578	3.7388	33.1951	2.6845	2.2718	0.4548
11	0.6984	0.6983	4.7889	5.1312	4.9501	0.2455
12	1.8274	3.8910	7.6914	2.5205	3.3428	0.2994
13	0.0435	1.4168	1.4806	0.1292	0.1295	0.0125
14	1.4799	1.6461	1.6907	2.5092	0.4257	0.1337
15	4.0073	4.1871	19.2649	2.6827	2.0322	0.0162
16	1.1725	1.8569	2.3003	0.8007	0.2025	0.0153
17	1.3539	2.3604	2.7329	8.1681	6.4676	0.9657
18	0.8981	1.5502	3.2532	0.4551	0.0959	0.0710
19	0.3434	0.3421	0.4405	0.3679	0.1469	0.0201
20	3.3341	2.9228	5.8723	0.4436	0.2771	0.0802
21	0.1597	0.6378	1.5039	1.2729	1.9698	0.0832
22	1.6166	2.7739	3.9912	4.3014	0.7835	0.0845
23	0.0678	1.2805	2.3856	0.5082	0.0472	0.0472
24	0.2107	2.2633	8.9225	2.3203	1.4319	0.2111
25	0.1292	0.6327	2.8449	2.2846	2.2771	0.6688
26	0.1206	0.9934	2.9113	0.5820	0.5432	0.1334

Die markierten Werte gehören zu Exemplaren, welche im Vergleich mit der Gesamtheit atypisch auftreten. Diese Exemplare könnten die Berechnungen in negativer Richtung beeinflussen. Die mit positivem Vorzeichen versehenen Spalten beziehen sich auf die drei größten Eigenwerte (Einflußbereich Variationen), während die mit einem negativen Vorzeichen versehenen Spalten beziehen sich auf die drei kleinsten Eigenwerte (Einflußbereich = Korrelationen).

Schnappmesser-Schätzungen der Eigenwerte
 6.5258 1.3941 0.3443 0.3695 0.2412 0.0901 0.0267
 0.0083 0.0000

Schnappmesser-Schätzungen der Standardabweichungen der Eigenwerte
 1.7220 0.2655 0.0982 0.0933 0.0619 0.0269 0.0080 0.0026

Schnappmesser-Schätzungen der Elemente der Eigenvektoren

1	0.38241	-0.33044	-0.22010	-0.38018	-0.31715	-0.37117	-0.38888
0.30634	0.30881						
2	-0.16938	0.43445	0.73386	-0.16244	-0.21361	-0.00621	-0.04461
0.30264	0.38554						
3	0.19637	0.25917	-0.09796	-0.07131	-0.35258	-0.21620	0.13416
-1.21992	0.60255						
4	-0.04319	0.15234	-0.03383	0.19928	-1.39542	-0.08930	0.27951
0.03387	-0.79834						
5	0.11129	1.00394	-0.41713	0.23098	0.13476	-0.63133	-0.20991
0.28589	-0.30133						
6	-0.25983	0.25295	-0.61818	0.20785	-0.10436	0.56201	-0.06205
0.27214	0.58806						
7	0.29927	0.34604	-0.10855	-0.74632	0.06131	0.47892	0.00590
-0.07694	-0.27319						
8	-0.15359	-0.03891	-0.17812	-0.31464	0.05927	-0.32417	0.85323
0.22867	0.15503						
9	0.80495	0.05060	0.13636	0.38884	0.01850	0.18741	0.32286
0.16179	0.12404						

Schnappmesser-Schätzungen der Standardabweichungen der Komponenten der Eigenvektoren

1	0.02253	0.06869	0.09663	0.02139	0.05581	0.03143	0.01387
0.06505	0.05485						
2	0.04477	0.08640	0.04498	0.05887	0.11378	0.05871	0.06587
0.16070	0.08272						
3	0.04973	0.16494	0.11488	0.10619	0.77808	0.12821	0.17920
0.14984	0.39536						
4	0.12888	0.49867	0.18618	0.15202	0.26205	0.33670	0.09240
0.66076	0.45595						
5	0.08685	0.28573	0.27620	0.10430	0.41244	0.13099	0.13282
0.36230	0.35240						
6	0.09039	0.20454	0.07699	0.12355	0.19855	0.20501	0.11917
0.16853	0.19720						
7	0.05241	0.08039	0.10423	0.06178	0.07489	0.12532	0.15732
0.06807	0.09200						
8	0.07024	0.09000	0.08517	0.14967	0.04567	0.08585	0.02587
0.05585	0.08303						
9	0.00696	0.00570	0.00981	0.01254	0.00473	0.01276	0.02749
0.00737	0.00882						

Anzahl bedeutungsvoller Hauptkomponenten

0	PRESS=	0.9615	Ergebnis	0.0000
1	PRESS=	0.3346	Ergebnis=	10.9036
2	PRESS=	0.1986	Ergebnis=	3.5553
3	PRESS=	0.1885	Ergebnis=	0.2427
4	PRESS=	0.1556	Ergebnis=	0.8233 Schwelle hier
5	PRESS=	0.1290	Ergebnis=	0.6607
6	PRESS=	0.1117	Ergebnis	0.3837
7	PRESS=	0.1074	Ergebnis=	0.0679
8	PRESS=	0.1050	Ergebnis	0.0206

Diese Ergebnisse deuten darauf hin, daß wir wenigstens zwei bedeutungsvolle Hauptkomponenten haben, möglicherweise 4: der Schneidepunkt soll nicht viel kleiner als 1 sein- hier > 0.8.

Die sogenannte Faktoranalyse von S. Wright

Die Faktoranalyse von S. Wright wurde schon 1932 eingeführt.

Sie ist von besonderem Interesse, weil sie die Methode von *Pfadkoeffizienten* umfaßt. Das Verfahren läßt sich am geeignetsten vermittels eines Beispielles beleuchten.

Beispiel: Die Morphometrie der Hühnchen Wrights.

Das klassische Beispiel der Pfadkoeffizienten behandelt eine Sammlung von Hühnchen. Es gibt zwei Schädelvariablen, zwei Tibiavariablen und zwei Femurmessungen. Hier folgt die Analyse: Die Variablen spiegeln den Wunsch Wrights, die Dimensionen des Schenkels und des Schienbeines mit denen des Schädels zu vergleichen.

KORRELATIONEN:

1.000	.877	.195	.572	.780	.814
.877	1.000	.265	.547	.760	.900
.195	.265	1.000	.119	.156	.215
.572	.547	.119	1.000	.595	.428
.780	.760	.156	.595	1.000	.744
.814	.900	.215	.428	.744	1.000

Einzelfaktorlösung

.925	.957	.229	.591	.845	.881
------	------	------	------	------	------

RESIDUEN:

.144	-.009	-.017	.025	-.002	-.002
-.009	.084	.046	-.019	-.048	.056
-.017	.046	.948	-.016	-.037	.014
.025	-.019	-.016	.650	.096	-.093
-.002	-.048	-.037	.096	.287	-.000
-.002	.056	.014	-.093	-.000	.223

Vierfaktorielle Lösung

.932	.964	.231	.591	.840	.877
------	------	------	------	------	------

RESIDUEN:

.132	-.021	-.020	.021	-.003	-.003
-.021	.071	.042	-.023	-.050	.055
-.020	.042	.947	-.018	-.038	.012
.021	-.023	-.018	.650	.098	-.091
-.003	-.050	-.038	.098	.294	.007
-.003	.055	.012	-.091	.007	.231

Die Pfadkoeffizienten

Für I=	1	und J =	4	Pfadkoeff. =	.1443
Für I=	2	und J =	3	Pfadkoeff. =	.2052
Für I=	2	und J =	6	Pfadkoeff. =	.2336
Für I=	3	und J =	2	Pfadkoeff. =	.2052
Für I=	3	und J =	6	Pfadkoeff. =	.1103
Für I=	4	und J =	1	Pfadkoeff. =	.1443
Für I=	4	und J =	5	Pfadkoeff. =	.3132
Für I=	5	und J =	4	Pfadkoeff. =	.3132
Für I=	5	und J =	6	Pfadkoeff. =	.0839
Für I=	6	und J =	2	Pfadkoeff. =	.2336
Für I=	6	und J =	3	Pfadkoeff. =	.1103
Für I=	6	und J =	5	Pfadkoeff. =	.0839

Quellen: Program Wright2.exe

Daten Wright.dta

Hauptkoordinatenanalyse

Die oben in der vorherstehenden Sektion behandelte Methode der HKA ist eine sogenannte R-Modus-Methode, d.h. sie gehört zum p -dimensionellen Raum der Variablen (die Benennung entstammt der gewöhnlichen Bezeichnungsweise der Korrelationsmatrix, \mathbf{R}). Anwendung der selben Mathematik im n -dimensionellen Raum (Raum der Individuen) wird eine Analyse nach Q-Modus genannt. Der englische Statistiker Gower (1966) schlug die Q-Modus-Methode genannt Hauptkoordinatenanalyse (HKO) vor. In der Regel benutzt man als Ausgangsmatrix die Assoziationen zwischen Individuen. Gower (1970) führte eine wertvolle Assoziationsmatrix ein, die eine Mischung von allerlei Merkmale erlaubt:- quantitative, qualitative und dichotome Variablen. Wenn man so möchte, kann man die Berechnungsweise für Korrelationen anwenden, um eine Art von "Korrelationen" zwischen Individuen (statt Variablen) herzustellen.

HKO erlaubt eine geometrische Darstellung der **Entfernungen** zwischen Individuen. Die Methode kann auch bei der Analyse von $k \geq 3$ Gruppen verwendet werden, wobei man die Mahalanobis-Abstände (siehe unten!) als eine $(k \times k)$ -Assoziationsmatrix einsetzt. Eine solche Analyse gleicht der Analyse mit Hilfe der Methode der **kanonischen Variablen** (siehe unten!).

BEISPIEL VON HAUPTKOORDINATEN

Das hier analysierte Material entstammt einer Untersuchung über die Ostrakodengattung *Soudanella* aus dem Paläozän Nigeriens (Reyment und Aranki, 1991). Diese sehr interessante Gattung kommt in Westafrika und Brasilien (Stinnesbeck und Reyment, 1987) vor und bildet ein wichtiges Verbindungsglied zwischen den beiden Kontinenten während der Frühzeit des Atlantischen Ozeans.

Anzahl der Variablen = 9
Anzahl der Individuen = 32

Die hier verwendete Assoziationsmatrix basiert auf den Euklidischen Abständen

Die Diagonalelemente der transformierten Matrix sind wie folgt:

206.44754	55.40063	74.70062	154.35696	110.91940	141.06757	235.42255
287.15503	214.94562	276.57251				
272.56378	275.01312	182.94751	179.34622	161.59436	95.29376	47.98065
281.34314	296.56067	359.49945				
270.98309	403.46942	424.13818	467.15063	624.16821	275.57739	213.13745
289.09875	300.39996	229.06195				
392.81696	503.73816					

Spur = 0.8302870E+04

Die drei ersten Eigenwerte der transformierten Assoziationsmatrix sind:

5595.24072 1237.72009 670.04083

Die Quadratwurzeln der Eigenwerte:

74.80134 35.18125 25.88515

Koordinaten der Individuen

	1	2	3
1	-0.84660	-8.37847	-9.78613
2	3.18184	-3.17419	-3.53040
3	5.25565	-3.95445	2.71528
4	-6.90254	-7.45869	3.47743
5	9.16919	-2.62374	-2.92249
6	3.06149	-9.16383	4.80127
7	5.35825	-11.88888	7.79865
8	13.52048	0.44320	-9.25646
9	13.32684	2.21746	-4.20138
10	12.19606	3.71732	-6.75817
11	14.83507	4.38283	4.29360
12	15.18169	3.60061	3.87428
13	-4.89493	-11.69647	0.18548
14	12.19958	1.72586	4.64318
15	-4.73511	-9.48191	5.58138
16	0.89392	5.37222	6.04660
17	4.14022	-1.20761	4.50318
18	14.54780	8.03904	1.35469
19	13.65833	9.65776	1.03811
20	17.60412	3.18885	-2.33559
21	14.14229	7.52529	2.56696
22	16.81489	-5.74329	-5.72655
23	-18.14061	6.05589	0.90384
24	-20.05052	5.17519	-5.03528
25	-21.49799	10.18742	2.68133
26	-14.05210	-3.31863	-3.39584
27	-13.82882	-1.88372	1.17784
28	-15.28344	-1.03266	-1.05925
29	-15.67653	-2.45420	-5.48638
30	-14.47016	-1.47578	0.57572
31	-18.04806	6.12607	4.60606
32	-20.66030	7.52152	-3.33092

Die Residuen

Eigenwert eliminiert	Prozentresiduum
2	17.70
3	9.63

Prozent der Spur für jeden Eigenwert

67.38924 14.90714 8.06999

N.B.- Die Residuen sind auffallend klein, was darauf hindeutet, daß die Koordinatenanalyse besonders wirkungsvoll sein muß.

Die Rangliste der Individuen auf Koordinatenachse I

25 32 24 23 31 29 28 30 26 27 4 13 15 1 16 6 2 17 3 7
5 10 14 9 8 19 21 18 11 12 22 20

Quellen: Program PCOORD.EXE
Daten: Soudanel.pcd

R-Q-MODUS-Methoden*Korrespondenzanalyse*

Oben habe ich die beiden Modalitäten der Multivariatanalyse eingeführt, R-Modus und Q-Modus. Die Korrespondenzanalyse ist ein Specialfall der HKA mit einem gemeinsamen Beziehungsmaß zwischen den Variablen und den Proben (mittels χ^2 , ein

gewichteter Euklidischer Abstand) wobei auch ein gemeinsames Koordinatensystem zur Darstellung der Variablen und Probenverteilungen verwendet wird. Die Methode der Korrespondenzanalyse vereinigt R- und Q-Modus Hauptkomponentenanalyse bzw Faktoranalyse (siehe unten!). Diese Bemerkungen gelten jedoch nur für die graphische Darstellung der Ergebnisse. Der Geltungsbereich der Korrespondenzanalyse sind Daten, die als Häufigkeitstabellen vorkommen.

Die Theorie der KA wurde zuerst von dem deutschen Mathematiker M. O. Hirschfeld (1936) (später Hartley) ausgearbeitet. Der englischer Statistiker Fisher (1940) führte explizite statistische Anwendungen des Begriffs (in bezug auf Kontingenztabelle) durch. Viel später wurde in Frankreich die Idee von J._P. Bénécri (1973) neuentdeckt, der die wichtigsten graphischen Erweiterungen des heutigen Verfahrens eingeführt hat. Der englische Biometriker M. Hill (1974) hat die historische Seite der KA geklärt und eine Verallgemeinerung der Methode entwickelt.

BEISPIEL VON KORRESPONDENZANALYSE

Die Untersuchung über die französischen Schlemmer: Aufgabe von Bénécri und Greenacre (1984). Die Variablen beziehen sich auf schlechte Gewohnheiten: Rauchen, Saufen, Freßen. Die Beobachtungen wurden in mehreren Pariser Bezirken gemacht.

Die Datenmatrix

	X			
	1	2	3	4
1	4.000	2.000	3.000	2.000
2	4.000	3.000	7.000	4.000
3	25.000	10.000	12.000	4.000
4	18.000	24.000	33.000	13.000
5	10.000	6.000	7.000	2.000

(Zeilen: (1) Nichtraucher, (2) raucht < 5/Tag, (3) raucht 5-10/Tag, (4) raucht > 10/Tag.

Reihen: (1) Vorsteher I, (2) Vorsteher II, (3) Angestellter (4), Angestellter II, (5) Sekretär.)

Ergebnisse der Zerlegung der Eingangsmatrix

Die Matrix **F**

		1	2	3	4
		R1	R2	R3	R4
1	VI	0.2387	-0.0574	0.4621	-0.8333
2	VII	0.3054	0.2892	0.7424	0.5061
3	AI	0.5141	-0.7155	0.0548	0.1303
4	AII	0.6752	0.5753	-0.3896	-0.1098
5	S	0.3599	-0.2647	-0.2838	0.1430

Die Matrix **G**

		1	2	3	4
1		0.5622	-0.8087	0.1713	0.0246
2		0.4829	0.1756	-0.6806	-0.5223
3		0.5668	0.4070	-0.0417	0.7151
4		0.3599	0.3867	0.7112	-0.4639

Die Projektionen der Variablen

RMW [1] rmw [2] rmw [3]
0.393 0.030 -0.002


```

0.099 -0.141  0.022
0.196 -0.007 -0.026
0.294  0.198  0.026
0.130  0.005  0.001

```

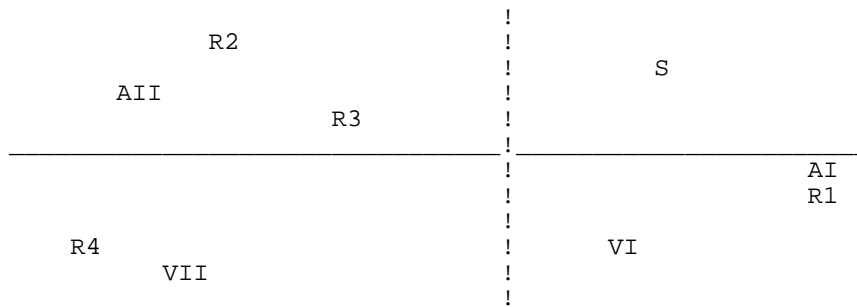
Die Projektionen der Individuen

```

          m[2]  v. Rechem[1]
Rechem[1] Rechem[2] Rechem[3]
-0.066  0.194  0.071
  0.259  0.243 -0.034
-0.381  0.012 -0.005
  0.233 -0.079 -0.008

```

Besprechung der Ergebnisse: Angestellten II und Sekretäre haben ähnliche Gewohnheiten. Die Kategorie Vorsteher I liegen zwischen diesen und einer Gruppe "Angestellten II - Vorsteher II".



Quellen : Genstatprogram Corresp.gpi

Die Biplotmethode

Eine der Korrespondenzanalyse ähnlich konstruierte Methode ist der Biplot von Gabriel (1968). Der Biplot ist für kontinuierliche Variablen geeignet, während die Korrespondenzanalyse für Häufigkeitsbeobachtungen gedacht ist (sie wird aber auch mit kontinuierlichen Variablen benutzt). Die Biplotmethode hat theoretische Vorteile, was mit sich führt, da β die "Biplotstruktur" nunmehr bei vielen neuen Q-R-Verfahren vorkommt.

Man beginnt mit der zentrierten Datenmatrix \mathbf{X} , was bedeutet, daß der Mittelwert jeder Variable gleich null ist. \mathbf{X} besitzt Rang r ; daher kann die Matrix als das Produkt zweier Matrizen ausgedrückt werden, also, $\mathbf{G}_{N,r}$ und $\mathbf{H}_{r,p}$. Wo \mathbf{X} nur den Rang 2 (siehe Beispiel unten!) hat, kann man ohne weiteres eine Biplotdarstellung durchführen. Ist doch der Rang > 2 , verwendet man die Spektralzerlegung einer rechteckigen Matrize, um die erwünschte Graphik herstellen zu können.

Die Spektralzerlegung schreiben wir:

$$\mathbf{X} = \mathbf{T}\mathbf{D}\mathbf{U} \quad (39)$$

wo \mathbf{T} und \mathbf{U} orthogonale Matrizen sind und \mathbf{D} ist eine Diagonalmatrix. Die kleinste Quadratannäherung von \mathbf{X} (um Rang 2 anzunähern) schreiben wir

$$\mathbf{X}_2 = \mathbf{T}_2 \mathbf{D}_2 \mathbf{U}_2. \tag{40}$$

Die gesuchten Matrizen sind dann

$$\mathbf{G} = \sqrt{(N-1)} (\mathbf{t}_1, \mathbf{t}_2) \tag{41}$$

und

$$\mathbf{H} = (d_{11} \mathbf{u}_1, d_{22} \mathbf{u}_2) / \sqrt{(N-1)} \tag{42}$$

Die Einzelheiten der Methode la ß en sich am besten durch ein Beispiel zeigen:

Beispiel: Die Matrix von Gordon (1981)

Die Daten von Gordon auf S. 116 seines Buches - ein konstruiertes Beispiel mit einer Matrix von genau Range 2.

$\mathbf{X}_{4,3}$ die Datenmatrix
 4.444000E-01 3.333000E-01 -1.333300
 4.111000E-01 -3.167000E-01 -3.333000E-01
 -1.889000E-01 4.833000E-01 -3.333000E-01
 -3.330000E-02 -6.500000E-01 1.000000
 -6.333000E-01 1.500000E-01 1.000000

Diese Matrix hat Rang = 2

Die kanonischen Gewichte durch die Q-R-Spektralzerlegung (siehe (40) von

$\mathbf{X} = \mathbf{T}_2 \mathbf{D}_2 \mathbf{U}'_2$.
 Spaltengewichte: -.307669 -.230761 .923087
 -0.6 0.8 0.0
 Zeilengewichte: -.667 -.1667 -.1667 .5 .5
 0.0 -0.5 0.5 -0.5 0.5

Die

Biplotkoordinaten

Erster Zeilenvektor
 -.7203 -.1801 -.1801 .5402 .5402
 zweiter Zeilenvektor
 .0000 -.5000 .5000 -.5000 .5000
 Erster Spaltenvektor
 -.6170 -.4628 1.8512
 Zweiter Spaltenvektor
 -.6000 .8000 .0000

Zusammenfassung: Punkte zwecks graphischer Darstellung

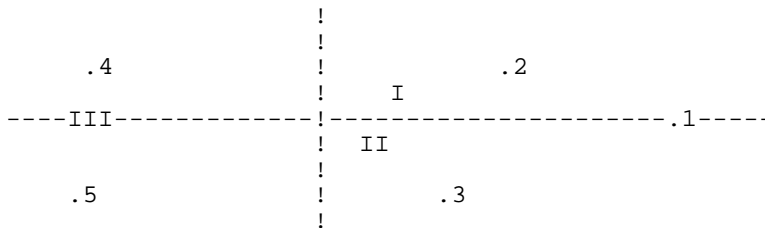
A. eine Matrix **G** (siehe Gleichung 41)

1 -.6170 -.6000
 2 -.4628 .8000
 3 1.8512 .0000

B. eine Matrix **H** (siehe Gleichung 42)

4 -.7203 .0000
 5 -.1801 -.5000
 6 -.1801 .5000
 7 .5402 -.5000
 8 .5402 .5000

Diese Koordinaten werden verwendet, um die Biplotdarstellung herzustellen.



Quellen: Program Gabriel.exe
 Daten Gordon.dat

Kanonische Korrelation

Wir kommen jetzt zu den Spezialmethoden bei welchen man Teilmengen verschiedener Herkunft vergleicht und kontrastiert. Die älteste davon ist die Methode der **kanonischen Korrelationen**, die schon vor 45 Jahren von Hotelling entwickelt worden ist. Ich fange mit einem Beispiel an, die Analyse von Gesteinen:

Variablen: linke Teilmenge = 3
 Variablen: rechte Teilmenge = 10

R_{11} Korrelationsmatrix der linken Seite

1	1.0000		
2	0.0667	1.0000	
3	-0.0389	0.0032	1.0000

R_{11} Korrelationen der rechten Seite

1	1.0000									
2	0.9962	1.0000								
3	0.9924	0.9975	1.0000							
4	0.9053	0.9338	0.9436	1.0000						
5	-0.5305	-0.4781	-0.4502	-0.1549	1.0000					
6	0.2192	0.2680	0.3230	0.5488	0.5782	1.0000				
7	0.6179	0.6600	0.7015	0.8560	0.2355	0.8971	1.0000			
8	0.3569	0.4050	0.4566	0.6657	0.4804	0.9887	0.9530	1.0000		
9	-0.0341	0.0249	0.0046	0.2585	0.5872	0.2315	0.2177	0.2410	1.0000	
10	0.2342	0.3003	0.3010	0.5767	0.5829	0.5739	0.6045	0.6048	0.9081	1.0000

R_{12} Korrelationen zwischen Mengen

1	0.9986	0.9971	0.9937	0.9064	-0.5278	0.2207	0.6195	0.3584		-0.0318
0.2360										
2	0.0636	0.1103	0.1682	0.3924	0.6355	0.9650	0.7958	0.9301		0.1663
0.4791										
3	-0.0409	0.0117	-0.0192	0.2011	0.4785	0.0710	0.0885	0.0877		0.9849
0.8405										

Signifikanzprüfung der Korrelationen

WILKS LAMBDA = < 0.0000
 CHI² = sehr hoch
 FG = 30.
 LAMBDA (2) = < 0.0000
 CHI² = 137.3281
 FG = 18.
 LAMBDA (3) = 0.0384
 CHI² = 42.3595
 FG = 8.

Kanonische Korrelationen

1 ung. 1.00

Eigenvektoren

Rechte Teilmenge
 (1) 0.3923 (2) -2.3055 (3) 1.8094 (4) 1.2550 (5) 0.3640 (6) 2.0617 (7)
 1.3604 (8) -3.6014 (9) -0.2345 (10) -1.0575

Linke Teilmenge

(1) 0.5740 (2) -0.0599 (3) -0.7972

Rechenmasswerte

Objekt	Rechte Werte	Linke Werte
1	1.3553	1.3229
2	1.4796	1.4912
3	0.0424	-0.0115
4	-0.2990	-0.3207
5	-0.2309	-0.1788
6	-1.6057	-1.6411
7	-0.9427	-0.9479
8	-0.8187	-0.7602
9	0.0840	0.0683
10	-0.7374	-0.6803
11	-0.4946	-0.5258
12	-0.8066	-0.7991
13	-0.5808	-0.5507
14	-0.4652	-0.4183
15	-1.5081	-1.5603
16	1.0550	1.0851
17	1.3226	1.2946
18	1.2621	1.2827
19	0.6457	0.5706
20	1.2428	1.2796

Korrelationen zwischen den neuen und den alten Variablen

Linke Seite

(1) 0.6011 (2) -0.0242 (3) -0.8197

Korrelation zwischen den neuen und den alten Variablen

Rechte Seite

(1) 0.6018 (2) 0.5564 (3) 0.5755 (4) 0.3361 (5) -0.7224 (6)
0.0122 (7) 0.2374 (8) 0.0800 (9) -0.8134 (10) -0.5630

Linke Varianzproportion die von der rechten Teilmenge erklärt werden kann = 0.3446

Rechte Varianzproportion die von der linken Seite erklärt werden kann = 0.3448

In diesem Fall herrscht ein beinahe symmetrisches Verhältnis, was sehr ungewöhnlich ist. Dies beruht auf der sehr hohen Korrelation zwischen den Teilmengen.

Kanonische Korrespondenzanalyse

Eine neuere Methode, die kanonische Korrespondenzanalyse (Akronym = CANOCO), wurde von C.J. ter Braak (1986) für die Analyse synekologischer Daten eingeführt. Diese Methode basiert auf einer Synthese von Korrespondenzanalyse und **kanonischer Korrelation**.

Beispiel der Methode

Nematoden in 4 Nahrungsgruppen (Manger & Schouten, 1989) in drei Wäldern (G,D,E), Holland.

CCA Kanonische Achsen: 2

Spez: Rechenmasswerte- Arten								N	NAME
AX1	AX2	AX3	AX4	Gewicht	N2				
	EIG	.0120	.005	.0689	.0319				
1	bact eat	.0222	.012	-.1290	-.1121	7529.00	62.65		
2	plan tpar	.2029	-.1354	.6959	-.0279	1286.00	42.93	3	fung ieat
		-.1551	-.0164	-.0417	.2953	3270.00	54.28		
4	rest grou	.1892	.3132	.5058	-.2048	420.00	34.53		
N	NAME	AX1	AX2	AX3	AX4	Gewicht	N2		
	EIG	.0120	.0053	.0689	.0319				
G00 1	-.4162	-2.8716	.5870	.3268	225.00	2.48			
G10 1	-1.4064	.9295	-.4682	-.2150	232.00	2.04			
G20 1	-.9251	.1855	-.4452	-.5253	250.00	1.96			
G30 1	-.9490	-3.5923	.7720	.9606	151.00	2.63			
G00 2	-1.6006	-1.434	-.0838	.4440	247.00	2.25			
G13 2	-.1117	1.444	-.1540	.8246	229.00	2.52			
G23 2	-.5604	-.300	-.4874	.8773	245.00	2.30			
G33 2	.9978	-1.720	-.4436	-.2006	118.00	1.99			
G00 3	-1.5933	.741	-.1791	.2465	241.00	2.25			
G19 3	-.7550	.461	-.4208	.6037	250.00	2.12			
G29 3	-1.2671	1.53	-.5877	.7688	267.00	2.11			
G39 3	-2.3681	1.74	-.9210	1.2368	95.00	2.05			
G00 4	-2.5609	.33	-.2505	.9210	218.00	2.29			
G19 4	-1.5400	1.17	-.4775	1.1093	162.00	2.22			
G29 4	-.4953	.178	-.3047	.5433	290.00	2.15			
G39 4	-.7782	3.4799	-.2902	.6385	126.00	2.28			
G00 5	-1.5516	-.77	-.3180	.1273	225.00	2.12			
G10 5	-.3044	.9754	-.3550	-.9317	179.00	1.94			
G11 5	1.5349	-1.0269	.3813	-1.4524	45.00	2.02			
G30 5	-3.0377	-1.3176	-.3462	1.2476	253.00	2.23			
G00 6	-3.8413	-1.6252	.0017	2.2457	233.00	2.38			
G13 6	-.0870	.3382	-1.0467	-.1117	282.00	1.85			
G23 6	.3687	3.4168	-.4963	.0098	134.00	2.22			
G33 6	-1.9120	-2.6097	-.1339	2.3797	78.00	2.58			
E00 1	-1.8800	1.4126	-.4770	.1154	183.00	2.11			
E19 1	3.6255	3.1051	1.7594	-.4798	201.00	2.99			
E29 1	1.3381	1.4006	.4960	-.0355	152.00	2.42			
E39 1	3.7149	.5681	1.6334	-.5797	172.00	2.73			
E00 2	-1.6200	.2036	-.2026	.2635	196.00	2.22			
E13 2	3.4846	-2.1148	.4452	-1.1143	220.00	2.03			
E23 2	4.2900	-1.8695	.9355	-1.2074	268.00	2.19			
E33 2	4.8357	-5.5213	1.9533	-.3743	136.00	2.63			
E00 3	-2.3324	-.8210	.1614	1.2391	314.00	2.47			
E10 3	1.2874	-2.5097	1.8723	.4045	77.00	3.11			
E20 3	3.7353	-2.7717	2.5514	-.7114	213.00	3.10			
E30 3	4.5783	-4.2109	2.9302	-.8841	104.00	3.05			
E00 4	-2.9296	.3457	-.5720	.8538	167.00	2.14			
E10 4	3.0545	-2.0368	2.1343	-.6743	111.00	3.00			
E20 4	-2.0591	1.7290	.1478	.9124	278.00	2.51			
E30 4	2.4455	-.3745	1.3771	-1.0967	49.00	2.61			
E00 5	-4.2468	-1.4808	-.3002	2.2222	174.00	2.26			
E13 5	3.1939	-4.0408	1.5973	.4279	248.00	2.92			
E23 5	4.9629	-1.6882	2.1711	-.3931	128.00	3.03			
E33 5	3.6500	-6.7091	2.6593	1.3364	178.00	3.20			
E00 6	-.2763	-3.1543	.6225	.2705	164.00	2.47			
E19 6	.4715	-.7481	.8046	1.0435	161.00	2.76			
E29 6	4.7626	9.0473	1.6967	-1.6547	208.00	2.72			
E39 6	2.6755	-3.2762	.8302	-.4991	75.00	2.18			
D00 1	-.9842	-1.2871	-.1414	-.0923	145.00	2.13			
D13 1	-.2723	.1113	-1.0470	.0382	132.00	1.88			
D23 1	-.8398	-.0560	-.8751	.6614	203.00	2.09			
D33 1	.2434	.6550	-1.3938	-.7499	101.00	1.59			
D00 2	-1.9074	.5740	-.8830	-.2674	114.00	1.85			
D19 2	.1660	1.8883	-.7391	-.4975	204.00	1.78			
D29 2	-.8502	.3463	-1.0256	.0285	142.00	1.75			
D39 2	.9133	1.5597	-.6564	-.9624	32.00	1.69			
D00 3	.4150	1.2639	-.1019	-1.2169	104.00	1.98			
D10 3	-1.9611	.8308	-.2038	.4955	175.00	2.28			
D20 3	1.2022	.8208	-.4832	-2.2079	160.00	1.58			

D30	3	.3352	-.5463	.0664	-.9006	97.00	2.06
D00	4	-1.4384	.8880	-.6796	-.4170	108.00	1.91
D10	4	.5471	2.8569	-.7939	-2.1290	220.00	1.56
D20	4	.3728	2.2142	-1.0261	-2.2206	192.00	1.45
D30	4	1.5677	-.9065	.2669	-1.6057	150.00	1.94
D00	5	-1.9915	1.5048	-.9597	-.3243	162.00	1.83
D13	5	.0384	1.0691	-1.1801	-.3806	221.00	1.76
D23	5	-.5103	.1393	-1.3418	-.1003	156.00	1.74
D33	5	.5714	.1325	-1.3119	-.8893	195.00	1.57
D00	6	-.0006	2.4215	-1.0099	-1.9281	121.00	1.52
D19	6	.5302	1.6968	-.6403	-.6597	204.00	1.77
D29	6	-.3561	1.7314	-1.0433	-.4227	128.00	1.68
D39	6	.9932	1.8952	-1.3932	-1.8315	87.00	1.29

BipE: Biplotwerte: Milieuvvariablen				N	NAME
AX1	AX2				
	R(SPEZ,MIL)	.4560	.4369		
1	L0	-.9639	-.2664		
2	L3	.7896	-.6136		
3	L9	.3174	.9483		
4	Kalk	.6249	.7807		
CenE: Baryzentrum der Milieuvvariablen					
N	NAME	AX1	AX2		
	R(SPEz,mil)	.4560	.4369		
1	L0	-.9601	-.2654		
2	L3	1.3264	-1.0307		
3	L9	.5705	1.7044		
4	lime	.7742	.9672		

Stichprobenwerte: lineare Vereinigungen der Milieuvvariablen

N	NAME	AX1	AX2	WEIGHT	% FIT
EIG		.0120	.0053	.0689	.0319
1	G00 1	-.9601	-.2654	225.00	61.79
2	G10 1	-.9601	-.2654	232.00	94.12
3	G20 1	-.9601	-.2654	250.00	97.43
4	G30 1	-.9601	-.2654	151.00	80.15
5	G00 2	-.9601	-.2654	247.00	86.88
6	G13 2	1.3264	-1.0307	229.00	-72.90
7	G23 2	1.3264	-1.0307	245.00	-11.23
8	G33 2	1.3264	-1.0307	118.00	79.15
9	G00 3	-.9601	-.2654	241.00	80.70
10	G19 3	.5705	1.704	250.00	20.47
11	G29 3	.5705	1.704	267.00	99.36
12	G39 3	.5705	1.7044	95.00	99.98
13	G00 4	-.9601	-.2654	218.00	92.07
14	G19 4	.5705	1.7044	162.00	99.42
15	G29 4	.5705	1.7044	290.00	-124.44
16	G39 4	.5705	1.7044	126.00	68.03
17	G00 5	-.9601	-.2654	225.00	90.95
18	G10 5	-.9601	-.2654	179.00	98.96
19	G20 5	-.9601	-.2654	45.00	71.27
20	G30 5	-.9601	-.2654	253.00	98.79
21	G00 6	-.9601	-.2654	233.00	99.80
22	G13 6	1.3264	-1.0307	282.00	98.89
23	G23 6	1.3264	-1.0307	134.00	-55.07
24	G33 6	1.3264	-1.0307	78.00	100.00
25	E00 1	-.9601	-.2654	183.00	80.33
26	E19 1	.5705	1.7044	201.00	86.37
27	E29 1	.5705	1.7044	152.00	99.65
28	E39 1	.5705	1.7044	172.00	100.00
29	E00 2	-.9601	-.2654	196.00	95.67
30	E13 2	1.3264	-1.0307	200.00	92.17
31	E23 2	1.3264	-1.0307	268.00	98.26
32	E33 2	1.3264	-1.0307	136.00	89.65
33	E00 3	-.9601	-.2654	314.00	98.34
34	E10 3	-.9601	-.2654	77.00	99.94
35	E20 3	-.9601	-.2654	213.00	99.49
36	E30 3	-.9601	-.2654	104.00	96.32

37	E00 4	-.9601	-.2654	167.00	97.18
38	E10 4	-.9601	-.2654	111.00	99.91
39	E20 4	-.9601	-.2654	278.00	54.16
40	E30 4	-.9601	-.2654	49.00	98.87
41	E00 5	-.9601	-.2654	174.00	99.97
42	E13 5	1.3264	-1.0307	248.00	97.27
43	E23 5	1.3264	-1.0307	128.00	98.24
44	E33 5	1.3264	-1.0307	178.00	96.07
45	E00 6	-.9601	-.2654	164.00	54.15
46	E19 6	.5705	1.7044	161.00	90.81
47	E29 6	.5705	1.7044	208.00	56.21
48	E39 6	.5705	1.7044	75.00	8.93
49	D00 1	-.9601	-.2654	145.00	63.33
50	D13 1	1.3264	-1.0307	132.00	99.61
51	D23 1	1.3264	-1.0307	203.00	85.28
52	D33 1	1.3264	-1.0307	101.00	99.77
53	D00 2	-.9601	-.2654	114.00	99.67
54	D19 2	.5705	1.7044	204.00	92.48
55	D29 2	.5705	1.7044	142.00	57.87
56	D39 2	.5705	1.7044	32.00	85.22
57	D00 3	-.9601	-.2654	104.00	92.46
58	D10 3	-.9601	-.2654	175.00	80.63
59	D20 3	-.9601	-.2654	160.00	95.42
60	D30 3	-.9601	-.2654	97.00	63.64
61	D00 4	-.9601	-.2654	108.00	99.59
62	D10 4	-.9601	-.2654	220.00	97.42
63	D20 4	-.9601	-.2654	192.00	99.95
64	D30 4	-.9601	-.2654	150.00	73.21
65	D00 5	-.9601	-.2654	162.00	97.66
66	D13 5	1.3264	-1.0307	221.00	95.54
67	D23 5	1.3264	-1.0307	156.00	99.67
68	D33 5	1.3264	-1.0307	195.00	97.35
69	D00 6	-.9601	-.2654	121.00	99.65
70	D19 6	.5705	1.7044	204.00	88.59
71	D29 6	.5705	1.7044	128.00	89.19
72	D39 6	.5705	1.7044	87.00	87.17

Quellen: Canoco.prn
Canoco.exe

Methoden für zwei oder mehrere Stichproben

Die bekannteste Methode dieser Klasse ist die **Diskriminanzanalyse** (Fischer, 1936). Als Diskriminanzanalyse wird eine Gruppe multivariatstatistischer Verfahren bezeichnet, die es gestatten:

(1) Ein Individuum auf Grund einer Reihe von beobachteten Merkmalen einer bestimmten Gruppe gleichartiger Individuen zuzuordnen.

(2) Die s.g. mehrdimensionale Varianzanalyse: die Signifikanzprüfung der Ergebnisse einer Diskriminanzanalyse.

(3) Die Deutung der Diskriminanzkoeffizienten. Diese Seite der Methodik läßt sich nicht ohne weiteres ausführen, weil die Stabilität der Eigenvektoren bei >2 Dimensionen Schwierigkeiten bereiten könnten.

Zur Historik des Verfahrens

Der erste, und bis heute am häufigsten angewendete diskriminanalytische Ansatz wurde bereit von Fisher (1936) angegeben. Der Amerikaner Wald hat die theoretische Seite sichergestellt (im Rahmen der Entscheidungstheorie). Andere wichtige Beiträge wurden von C. R. Rao und M. S. Bartlett gemacht. Die ursprüngliche Form der Diskriminanzanalyse war linear. Eine spätere besonders bei Identifikationsproblemen wertvolle Entwicklung ist der quadratische Ansatz (quadratische Diskriminanzfunktion).

Die mehrdimensionale Normalverteilung

Die Theorie der Multivariatanalyse basiert auf bestimmten Gesetzmäßigkeiten der mehrdimensionalen Normalverteilung.

Eine eindimensionale bzw skalare Zufallsgröße y ist normalverteilt, wenn für ihre Wahrscheinlichkeitsdichte

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad (12)$$

(43)

wo bei μ der Mittelwert und σ die Standardabweichung sind. Die Generalisierung erfolgt ohne weiteres.

Die Schätzung für Σ ist S und für μ schreiben wir die entsprechende Stichprobenschätzung \bar{y} . Daher hat man die Gleichung (13) für S :

$$S = \frac{1}{N-1} \sum_{j=1}^N (y_j - \bar{y})(y_j - \bar{y})'$$

(44)

Die Matrix S ist eine positive semidefinite, meist sogar eine positive definite symmetrische $p \times p$ -Matrix. In ihrer Hauptdiagonale stehen die Streuungen bzw Quadrate der Standardabweichungen (d.h. die Varianzen) der p einzelnen Merkmale.

Prüfung von Hypothesen hinsichtlich des Mittelwerts

1. Prüfung der Hypothese daß $\mu = 0$.

Die Verallgemeinerung der Teststatistik Students wird hier eingesetzt:

$$\tilde{F} = \frac{f_2 - p + 1}{f_1 p} \text{Spur}(HG^{-1}) \quad (14)$$

(45)

Die Freiheitsgrade sind p und $f_2 - p + 1$. Hier $f_2 = N - 1$.

2. Die Hypothese $\mu_1 = \mu_2$

$$\tilde{F} = \frac{(N_1 + N_2 - p - 1)}{p(N_1 + N_2 - 2)} \frac{N_1 N_2}{N_1 + N_2} (\bar{Y}_1 - \bar{Y}_2)' S^{-1} (\bar{Y}_1 - \bar{Y}_2)$$

Die Teststatistik ist der sogenannte T^2 -test von Hotelling (1932).

(46)

Dieses Problem steht im Vordergrund des Interesses bei vielen Untersuchungen. Man geht von zwei Stichproben aus, von denen man die Mittelwertvektoren und die **klasseninterne Kovarianzmatrix** benötigt.

Für die Freiheitsgrade haben wir p und $N_1 + N_2 - p - 1$.

Vergleich der Mittelwertvektoren zweier mehrdimensionaler normalverteilter Grundgesamtheiten

Der Malalanobis-Abstand

Der generalisierte statistische Abstand von P. C. Mahalanobis wird definiert:

$$D^2 = (\bar{Y}_1 - \bar{Y}_2)' S^{-1} (\bar{Y}_1 - \bar{Y}_2) \quad (16)$$

(47)

wurde schon 1928 eingeführt. Diese Größe charakterisiert inwieweit die beiden gegebenen Stichproben der Hypothese $\mu_1 = \mu_2$ widersprechen bzw wie groß der *statistische Abstand* zwischen den

beiden Gesamtheiten ist.

Diskriminanzfunktion

Wie wollen eine Linearkombination der gegebenen p Merkmale Y_1, \dots, Y_p

$$v = d_1 y_1 + \dots + d_p y_p$$

wobei der Vector \mathbf{d} durch die folgende Gleichung festgelegt ist:

$$\mathbf{d} = \mathbf{S}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2). \quad (48)$$

Dieser Vektor enthält die gesuchten Komponenten der Diskriminanzfunktion. Für jedes Individuum erhält man einen Wert von v , ein neues Merkmal. Diese kann man als die Diskriminanzrechenmaßwerte bezeichnen.

Wie haben also

$$v = \mathbf{d}'\mathbf{y}. \quad (49)$$

Dieses Merkmal besitzt in den beiden Gesamtheiten die Mittelwerte:

$$\bar{v}_1 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \mathbf{S}^{-1} \bar{\mathbf{y}}_1 \quad (50)$$

und

$$\bar{v}_2 = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \mathbf{S}^{-1} \bar{\mathbf{y}}_2 \quad (51)$$

Das multivariate Trennmaß β von v ist die T^2 -Statistik von Hotelling. Hierunten folgt ein Beispiel.

Beispiel der Diskriminanzanalyse

Vergleich zwischen Weibchen und Männchen der Krabbenart *Leptograpsus variegatus* aus Australien

Korrelationsmatrix der Weibchen

1.000	.964	.989	.990	.980
.963	1.000	.978	.979	.964
.989	.978	1.000	.997	.991
.990	.979	.997	1.000	.988
.980	.964	.991	.988	1.000

Korrelationsmatrix der Männchen

1.000	.968	.995	.995	.993
.968	1.000	.977	.979	.969
.995	.977	1.000	.999	.994
.995	.998	.999	1.000	.995
.993	.969	.994	.995	1.000

Summe der beiden Korrelationsmatrizen

	1	2	3	4	5
1	1.000	.966	.992	.992	.987
2	.965	1.000	.977	.979	.967
3	.992	.977	1.000	.998	.993
4	.992	.988	.998	1.000	.991
5	.987	.967	.993	.991	1.000

Die Mittelwertvektoren

(1)	17.594	14.832	34.618	39.936	15.642
(2)	14.842	11.718	32.014	36.810	13.350

Differenzvektor (d.h. Vektor 1 - Vektor 2)

2.752	3.114	2.604	3.126	2.292
-------	-------	-------	-------	-------

Der Diskriminanzvektor (verschiedene Standardisierungen)

Var.	Stand.1	Stand.2	Stand.3
1	-4.706	-4.706	-.501
2	-1.547	-1.547	-.165
3	-105.359	-105.359	-11.223
4	142.017	142.017	15.128
5	-27.790	-27.790	-2.960

$$D^2 = 88.1278$$

$$D = 9.3876$$

$$F = 422.6537 \text{ mit } 5 \text{ und } 94 \text{ Freiheitsgrade}$$

Die Wahrscheinlichkeit, daß die Stichproben derselben
Gesamtheit entstammen = 0.0

Fehlbestimmungswahrscheinlichkeit = 0.0

$$D^2 \text{ (ohne Bias)} = 82.53$$

$$D = 9.08$$

Quellen: Altdis.exe
Altdis.dat

Aussonderung entbehrllicher Merkmale

Bei fast allen multivariaten Untersuchungen interessiert man sich für das Vorkommen von überflüssigen Variablen. Man möchte die Anzahl der benutzten Originalmerkmale verringern, jedoch das Trennmaß beibehalten - das Leistungsvermögen der Funktion soll etwa gleich bleiben.

Die Unentbehrlichkeit U_i von y_i ist definiert als die Größe, um die das Trennmaß absinkt, wenn aus der Menge aller Merkmale das Merkmal y_i eliminiert wird.

$$\text{Also } U_i = T_p^2 - T_{(p-1)}^2$$

Wenn man bestrebt ist, eine möglichst hohes Trennmaß β zu behalten (bei Aussonderung), so muß man das Merkmal mit dem kleinsten U_i eliminieren.

Viele Standardprogramme enthalten die Möglichkeit eine

schrittweise Merkmalsreduktion durchführen zu können. Bei jedem Schritt wird das Merkmal eliminiert, das durch ein Ausscheiden den geringsten Abfall des multivariaten Trennmaßes verursacht.

Mehr als drei Grundgesamtheiten

Die statistische Analyse geht von mehreren Gesamtheiten (Klassen) bestimmter Individuen aus. Man spricht dann von mehrdimensionaler Varianzanalyse für k Klassen, kanonischer Variablen, Diskriminanzanalyse mehrerer Gruppen, usw. Es wird vorausgesetzt, daß aus jeder Klasse eine normalverteilte Stichprobe von p -dimensionalen Beobachtungsvektoren gegeben sei.

Für k Klassen (Gruppen) und p Variablen haben wir das Verhältnis:

$$Y_{ij} = \mathbf{c}'\mathbf{x}_{ij} \quad (53)$$

Der Rechenmaßwert der kanonischen Variablen x_{ij} ist die i -te Beobachtung der j -ten Gruppe.

Den ersten kanonischen Vektor (also generalisierter Eigenvektor) erhält man freilich durch die Bedingung, daß folgender Quotient

$$y = \mathbf{c}'\mathbf{B}\mathbf{c}/\mathbf{c}'\mathbf{W}\mathbf{c} \quad (54)$$

möglichst groß ausfallen soll. \mathbf{B} ist die Matrix der Summen der Abweichungsquadrate zwischen Gruppen. \mathbf{W} ist die Matrix der klassinternen Summen der Abweichungsquadrate. Ganz kurz könnten wir sagen "Zwischenmatrix" für \mathbf{B} und "Innenmatrix" für \mathbf{W} .

Dann hat man die folgende Beziehungen für die Lösung der Eigenaufgabe.

$$(\mathbf{B} - f\mathbf{W})\mathbf{c} = \mathbf{0} \quad (55)$$

und

$$\mathbf{c}'\mathbf{W}\mathbf{c} = N_w \quad (56)$$

Hier ist N_w die Anzahl der Freiheitsgrade.

Die Werte von f sind die gesuchten Eigenwerte des generalisierten Eigenproblem

$$|\mathbf{B} - f\mathbf{W}| = 0 \quad (57)$$

und die Vektoren \mathbf{c} sind die generalisierten Eigenvektoren (häufig von anglophonen Statistikern als "kanonische Vektoren" bezeichnet).

Sind die Kovarianzmatrizen ungleich, kann es befugt sein, eine *quadratische Diskriminanzanalyse* auszuführen.

BEISPIEL: Kanonische Variableanalyse einer Pflanzenart

Kanonische Variablen:

Picea Glauca

Gruppen = 5

Variablen = 7

Die Innengruppen Matrix W

1	2	3	4	5	6	7
1	42.0659	13.1388	10.1983	13.7615	12.6611	13.6273
0.6570						
2	13.1388	25.5823	5.4549	1.2344	4.7091	5.3109
0.1247						
3	10.1983	5.4549	15.6746	3.0514	11.9165	4.7232
0.2893						
4	13.7615	1.2344	3.0514	9.9476	6.2813	8.6655
0.2632						
5	12.6611	4.7091	11.9165	6.2813	17.9862	10.4178
0.3249						
6	13.6273	5.3109	4.7232	8.6655	10.4178	19.9568
0.5031						
7	0.6570	0.1247	0.2893	0.2632	0.3249	0.5031
0.1649						

Die Zwischengruppenmatrix B

1	2	3	4	5	6	7
1	36.9172	27.1652	21.2005	11.9269	15.7719	35.4796
0.8127						
2	27.1652	23.6193	15.1515	7.4301	9.9294	27.8209
0.3380						
3	21.2005	15.1515	12.8021	7.2461	9.5979	21.5006
0.4852						
4	11.9269	7.4301	7.2461	4.9393	6.1372	12.5604
0.3548						
5	15.7719	9.9294	9.5979	6.1372	7.8718	15.8395
0.4600						
6	35.4796	27.8209	21.5006	12.5604	15.8395	40.9247
0.6296						
7	0.8127	0.3380	0.4852	0.3548	0.4600	0.6296
0.0368						

Die Matrix B + W (= die Matrix der Gesamtvariation)

1	2	3	4	5	6	
1	78.9831	40.3040	31.3987	25.6884	28.4330	49.1070
1.4697						
2	40.3040	49.2016	20.6064	8.6645	14.6385	33.1318
0.4627						
3	31.3987	20.6064	28.4767	10.2975	21.5143	26.2237
0.7745						
4	25.6884	8.6645	10.2975	14.8869	12.4185	21.2259
0.6180						
5	28.4330	14.6385	21.5143	12.4185	25.8579	26.2573
0.7849						
6	49.1070	33.1318	26.2237	21.2259	26.2573	60.8815
1.1328						
7	1.4697	0.4627	0.7745	0.6180	0.7849	1.1328
0.2017						

Die Gesamtzahl = N = 113

Tabelle der Mittelwerte

1	2	3	4	5	6	7
1	11.4272	7.7580	6.3340	3.3960	5.6692	4.1856
0.3184						

2	10.8262	7.7408	5.9600	3.2208	5.3269	4.2677
0.2746						
3	12.3944	8.9088	6.8632	3.5664	5.9016	5.4108
0.3076						
4	12.5296	8.6496	6.9284	3.8580	6.1716	5.5716
0.3380						
5	12.0656	8.3380	6.8544	3.7192	6.0412	5.4616
0.3208						

Matrix der Mahalanobisabstände

	1	2	3	4	5
1	0.0000	3.4352	19.1103	14.6937	13.7935
2	1.8534	0.0000	16.5943	15.1336	12.2830
3	4.3715	4.0736	0.0000	2.2633	2.5978
4	3.8332	3.8902	1.5044	0.0000	0.8819
5	3.7140	3.5047	1.6118	0.9391	0.0000

Die generalisierte Eigenanalyse

EIGENWERT 1	3.335593
EIGENWERT 2	0.387639
EIGENWERT 3	0.231244
EIGENWERT 4	0.040796

EIGENVEKTOREN VON $W^{-1}B$

	1	2	3	4
1	-0.0305	-0.0084	0.1458	0.0347
2	-0.0353	-0.0452	0.0719	-0.1342
3	-0.1676	0.0412	-0.0153	0.2611
4	0.0921	0.1912	-0.1610	-0.2078
5	0.1735	0.0930	0.0205	-0.1014
6	-0.2442	-0.0868	-0.1406	0.0294
7	0.4381	1.4509	0.4253	-0.2404

Matrix der kanonischen Mittelwertvektoren

2.60309	0.42454	0.37437	0.10342
2.25859	-1.08831	-0.58292	-0.23030
-1.62367	-0.67810	0.54126	0.07031
-1.18852	0.61324	0.01112	-0.28365
-0.96536	0.20625	-0.62364	0.22967

F-Quotient bezüglich der generalisierten Abstände: Wahrscheinlichkeiten in der unteren Dreiecksmatrix

	1	2	3	4	5
1	0.0000	3.9639	32.2296	24.7811	23.2628
2	0.0008	0.0000	19.1486	17.4630	14.1737
3	0.0000	0.0000	0.0000	3.8170	4.3812
4	0.0000	0.0000	0.0011	0.0000	1.4873
5	0.0000	0.0000	0.0003	0.1796	0.0000

Signifikanzprüfung der Eigenwerte

LAMBDA 1	2.0424728
CHI ² =	216.5021210
F.G. =	28
Wahrscheinlichkeit =	0.0000E+00
LAMBDA 2	0.5756144
CHI ² =	61.0151253
FG =	18
Wahrscheinlichkeit =	0.1824E-05
LAMBDA 3	0.2480110
CHI ² =	26.2891636
FG =	10
Wahrscheinlichkeit =	0.3476E-02
LAMBDA 4	0.0399858
CHI ² =	4.2384996
F.G. =	4
Wahrscheinlichkeit =	0.3751

Test für die Homogenität der Mittelwerte

DET T =	0.770965052E+01
WILKS LAMBDA =	0.129708
CHI ₂ =	204.2473 mit 28 Freiheitsgraden

Wahrscheinlichkeit = 0.000000
 HOMOGENITÄT der Kovarianzmatrizen
 CHI² = 162.960327
 Freiheitsgrade = 112
 Wahrscheinlichkeit = 0.001208
 Schluss: Kovarianzmatrizen nicht gleich.

Die Homogenität von Kovarianzmatrizen

Die Prüfung der Homogenität von zwei (oder mehreren) Kovarianzmatrizen besteht aus zwei Teilen. Zuerst untersucht man das Verhältnis:

$$2I(H_1:H_2(*)) = N_1 \ln \left(\frac{|S|}{|S_1|} \right) + N_2 \ln \left(\frac{|S|}{|S_2|} \right)$$

(58)

wo **S** die Innenkovarianzmatrix darstellt. Diese Statistik ist als χ_2 verteilt mit Freiheitsgraden $p(p+1)/2$, p ist die Anzahl der Variablen.

Die **Orientierung** der Hauptachsen der beiden Ellipsoide wird durch den folgenden χ^2 -Ausdruck geprüft:

$$n(d_i S_1^{-1} b_i + \frac{1}{d_i} b_i S_1 b_i - 2)$$

(59)

Die Vektoren **b_i** sind die Eigenvektoren der zweiten Matrix **S₂**.

Nächststehend gebe ich ein Beispiel.

Vergleich zweier Vorkommen von Wisentknochen. Analyse der Homogenität zweier Kovarianzmatrizen

Die Metatarsalen von *Bos*: Ortschaften = Barrington (N=18) und Wretton (N=15) (Großbritannien).

Stichprobe 1	
Mittelwerte	Standardabweichungen
278.27777	11.57485
64.77778	5.95627
64.16666	5.87367
74.38889	6.54522
43.50000	3.53553
34.77778	3.28196
40.00000	5.69830
51.05556	4.63328

Stichprobe 2	
264.26666	9.16879
59.13334	4.30725
67.80000	4.09180
58.46667	3.48193
41.13334	2.55976
30.20000	1.78085
34.60000	3.37639
45.26667	4.30061

Stichprobe 1				
Variable	Schiefheit	t	Kurtosis	t
1	-0.3632	-0.6774	-0.9741	-0.9387
2	0.3857	0.7192	0.0354	0.0341
3	0.8020	1.4955	0.1676	0.1615
4	0.6079	1.1335	0.3212	0.3095
5	0.2866	0.5344	-0.6905	-0.6653
6	-0.7757	-1.4464	0.5926	0.5710
7	0.1441	0.2688	-0.8516	-0.8206
8	0.0138	0.0257	-0.6476	-0.6240

Stichprobe 2				
1	-0.1024	-0.1765	-1.0341	-0.9225
2	0.5210	0.8980	-1.2457	-1.1113
3	0.4204	0.7246	-0.6960	-0.6209
4	0.0690	0.1190	-1.0387	-0.9267
5	0.0296	0.0511	-1.3923	-1.2422
6	0.0033	0.0058	-1.2008	-1.0712
7	0.5618	0.9683	-1.1562	-1.0315
8	0.5085	0.8765	0.3773	0.3366

Kovarianzmatrix für Barrington

	1	2	3	4	5	6	7	8
1	133.977	51.53	51.00	52.65	25.91	23.24	44.05	36.68
2		35.47	32.03	36.79	18.58	14.94	30.35	17.89
3			34.50	33.40	17.38	13.09	26.46	21.46
4				42.83	20.02	15.26	33.41	20.15
5					12.50	8.11	14.47	10.85
6						10.77	14.47	9.60
7							32.47	16.70
8								21.46

Kovarianzmatrix für Wretton

	1	2	3	4	5	6	7	8
1	84.06	17.17	13.20	12.36	7.60	5.01	8.40	19.13
2		18.55	15.81	13.36	8.98	6.11	12.70	7.96
3			16.74	12.81	8.38	5.04	12.05	7.70
4				12.12	7.43	4.61	9.98	7.65
5					6.55	3.90	5.62	4.89
6						3.17	3.15	3.08
7							11.40	5.75
8								18.49

Die

Eigenanalysen

(1) Ergebnisse- Barrington

Eigenwerte

260.20419	36.18302	10.98748	7.09341	4.15210	3.42540	1.31145	0.64615
-----------	----------	----------	---------	---------	---------	---------	---------

Eigenvektoren

	1	2	3	4	5	6	7	8
1	-0.6631	0.7295	0.1554	0.02060	0.0027	0.0412	0.0053	-0.0406
2	0.1788	0.2361	-0.1951	-0.2284	0.2011	0.7564		
3	-0.3323	-0.2136	-0.3933	0.3730	0.0881	-0.6045	-0.2116	0.3632
4	-0.3647	-0.4053	0.1772	0.1550	0.2900	0.5641	-0.4930	-0.0416
5	-0.180	-0.1960	-0.1078	0.3129	-0.4191	0.3922	0.5885	-0.3787
6	-0.1542	-0.1254	0.0077	-0.4341	-0.7640	-0.0547	-0.4212	-0.0881
7	-0.3038	-0.3341	0.3260	-0.5915	0.2975	-0.2255	0.3690	-0.2534
8	-0.2239	-0.0215	-0.7998	-0.3798	0.1483	0.2338	0.1078	0.2776

(2) Ergebnisse- Wretton

Eigenwerte

109.01131 42.36156 11.76245 3.74905 1.85670 1.30552 0.76978 0.28840

Eigenvektoren

	1	2	3	4	5	6	7	8
1	0.8007	0.5713	-0.1697	0.0185	0.0315	-0.0177	-0.0289	-0.0333
2	0.3053	-0.4056	-0.1972	-0.0854	-0.6434	0.2451	0.2047	0.4241
3	0.2660	-0.4250	-0.1174	0.1261	0.6423	0.5316	0.1504	-0.0326
4	0.2364	-0.3434	-0.0105	-0.0772	0.2405	-0.7726	0.3934	0.1023
5	0.1519	-0.2339	-0.0073	-0.5769	0.1491	-0.0916	-0.7261	0.1769
6								0.0975
7	-0.1436	-0.0099	-0.5044	-0.2049	0.0900	0.2229	-0.7845	
8	0.1931	-0.3592	-0.0954	0.6167	-0.2136	-0.1902	-0.4501	0.3995
9	0.2626	-0.0776	0.9534	0.0523	-0.0681	0.0856	0.0117	0.0320

Die Logarithmen der Determinanten

Barrington	Wretton	Summe
15.995	11.604	15.532

$B^2 = 47.0955887$
 $\beta^2 = 9.8647327$
 Freiheitsgrade = 36

Richtungen der Ellipsoidenachsen

VEKTOR	CHI ²	FG
1	24.289	7
2	16.865	7
3	32.494	7
4	93.287	7
5	16.061	7

(NB: Die Orientierungsunterschiede bei den Achsen 1, 3 und 4 sind bedeutend.)

Schätzung des Diskriminanzvektors

Variable

1	0.03052
2	-0.67540
3	2.79741
4	-4.36531
5	2.16172
6	-0.85565
7	1.66634
8	-0.40601

Generalisierte Abstände im Fall der Heterogenität der Kovarianzmatrizen

D	D ²
8.409	70.718

Signifikanzprüfung

$T^2 = 565.059$
 $F = 54.68$
 $FG1 = 8$ $FG2 = 24$

Bemerkungen: Die beiden Kovarianzmatrizen sind ungleich. Der Unterschied zwischen den Mittelwertsvektoren ist statistisch signifikant.

Die Stabilität der generalisierten Eigenvektoren

Eine sehr bedeutende Frage gilt der Stabilität der Eigenvektoren einer Multivariatanalyse. Die wichtigsten Schritte bei der statistischen Lösung des Problems - in bezug auf kanonische Variablen- gebe ich hier unten. Die von N. A. Campbell entwickelte Methode (1978, 1980) kann auf folgende Weise zusammengefaßt werden: Sie basiert auf der in der Regressionsanalyse vorkommenden Methode der s. g. *Schrumpfung*. Die einzelnen Stufen der Berechnungen, algorithmisch dargestellt, sind:

1. Man fängt mit der Matrix der gruppinternen Summe der Abweichungsquadrate \mathbf{W} und der entsprechenden Matrix \mathbf{B} der Abweichungsquadrate zwischen Gruppen an.

2. \mathbf{W} wird dann "standardisiert", um die entsprechende Korrelationsmatrix \mathbf{W}^* zu bilden. Man wendet das gleiche Verfahren auf \mathbf{B} an. Folglich

$$\mathbf{W}^* = \mathbf{S}^{-1}\mathbf{W}\mathbf{S}^{-1} \quad (60)$$

und

$$\mathbf{B}^* = \mathbf{S}^{-1}\mathbf{B}\mathbf{S}^{-1} \quad (61)$$

Hier bezeichnet \mathbf{S} die Diagonalmatrix, deren Elemente $\sqrt{w_{ii}}$ sind.

3. Berechne die Eigenwerte e_i und Eigenvektoren \mathbf{u}_i von \mathbf{W}^* .

$$\mathbf{W}^* = \mathbf{U}\mathbf{E}\mathbf{U}' \quad (62)$$

4. Die Schrumpfung erfolgt normalerweise dadurch, daß man den Eigenwerten die Schrumpfungskonstante k_i hinzufügt.

5. Man definiert dann eine Matrix

$$\mathbf{U}^* = \mathbf{U}(\mathbf{E} + \mathbf{K})^{-\frac{1}{2}} \quad (63)$$

6. Bilde dann die Zwischengruppsmatrix \mathbf{G} im Raum der Hauptkomponenten innerhalb der Gruppen:

$$\mathbf{G} = \mathbf{U}^*\mathbf{B}^*\mathbf{U}^* \quad (64)$$

Hier, bezeichnet g_i das i -te Diagonalelement der Matrix \mathbf{G} . Die Eigenwerte und Eigenvektoren von $\mathbf{G}_{(0,\dots,0)}$ sind die normalen Eigenwerte und Eigenvektoren der gewöhnlichen kanonischen Analyse (generalisiertes Eigenwertproblem), also,

$$\mathbf{c}^U = \mathbf{U}_{(0,\dots,0)}^* \mathbf{a}^U \quad (65)$$

Die geschrumpften Schätzungen erhält man durch die Eigenvektoren \mathbf{a}^S von $\mathbf{G}_{(k_1,\dots,k_P)}$.

Campbell und Reyment (1978) machten eine Untersuchung über die Stabilität der Eigenvektoren bei fossilen Foraminiferen. Im allgemeinen kann man sagen, daß auch bei existierender

Instabilität die Einflüsse derselben oft ziemlich gering sind.

Die Faktoranalyse als multivariatstatistische Methode

Es gibt zwei faktoranalytische Modelle. Die eine kann man, in Ermangelung eine bessere Bezeichnung, als **die wahre oder echte Faktoranalyse** bezeichnen: nämlich die klassische Faktoranalyse, welche in der Psychometrie verwendet wird.

Das zweite Modell ist eigentlich eine Variante der Hauptkomponentenanalyse- mit vielen der Berechnungsweisen der echten Faktoranalyse, Drehung der Achsen usw, ohne jedoch die Grundstruktur des wahren Faktor-Modells zu besitzen.

Das Grundmodell ist

$$\mathbf{X}_{N,p} = \mathbf{F}_{N,k} \mathbf{A}'_{k,p} + \mathbf{E}_{N,p} \quad (66)$$

wobei die Matrizen folgende Bedeutungen besitzen:

\mathbf{X} ist die Datenmatrix, \mathbf{F} ist die Matrix der faktoriellen Rechenmaßwerte, \mathbf{A} ist die Matrix der zu bestimmenden Faktorelemente, und \mathbf{E} ist die Matrix der Residuen. Dazu haben wir k , das die Anzahl der erwünschten Faktoren angibt ($k \leq p$).

Das Hauptkomponentenmodell der Faktoranalyse

HKA ist das in den Naturwissenschaften am häufigsten vorkommenden Modell der angewandten Faktoranalyse. Das Ziel ist, die Summen der Quadrate von [67] zu minimieren:

$$\mathbf{E} = \mathbf{Y} - \mathbf{FA}'$$

[67]

Diese Lösung erhält man durch die bekannte Zerlegung:

$$\hat{\mathbf{F}}\hat{\mathbf{A}}' = \gamma_1 \mathbf{v}_1 \mathbf{u}_1' + \dots + \gamma_k \mathbf{v}_k \mathbf{u}_k' \quad (20)$$

(68)

Diese Gleichung hat eine beliebige Anzahl von Lösungen, weil wir nur das Produkt \mathbf{FA} bestimmen können, aber nicht die einzelnen Matrizen. Es gibt zwei ungefährliche Lösungsmöglichkeiten:

(i) Für \mathbf{Y} die zentrierte Datenmatrix:

$$\hat{\mathbf{A}} = \mathbf{V}_k \mathbf{\Lambda}_k^{-1/2}; \mathbf{V}'\mathbf{V} = \mathbf{I}$$

$$\hat{F} = \hat{Y} A \Lambda_k^{-1}$$

(69)

(70)

(ii) Zweite Möglichkeit

$$\hat{F} = V \Gamma_k$$

(71)

wo Γ ist wie oben definiert.

Das echte Faktormodell kommt, wie schon notiert, in der Naturwissenschaft kaum vor (siehe Reyment, Jöreskog und Klován, 1993).

BEISPIEL

Analyse der Daten Klováns (1966).

*Angewandter Abschnitt*Fünfter Teil: *Bereich: QUANTITATIVE GENETIK*

1. Die Erweiterung der Methoden der Haustierkunde nach D. S. Falconer.
2. Umfang der multivariaten Quantitativgenetik.
3. Univariate Methoden und die Modelle R. Landes
Verallgemeinerung der univariaten Verfahren:
Nachteile, Vorteile und Vorbehalte
4. Geographische Stellungnahmen
Beispiele: Die Entwicklung der Säugetiere.
5. Polymorphismus.
6. Zufallsvariablen: und die Deutung von Entwicklungsreihen.
Selbstbetrug und die Trugschlüsse der s.g. Evolutionsbiologen (die

Gegenüberstellung von Evolutionsvorstellungen, beispielsweise der sog. phylogenetische Gradualismus gegenüber verzweigten Stillstandsepochen). Das Modell von Feller-Bookstein für Entwicklungsreihen gibt eine ernüchternde Einstellung zur ganzen Fragestellung.

Einleitendes

Die quantitative Genetik als selbständige Wissenschaft beruht in der Hauptsache auf den Arbeiten des schottischen Genetikers D. S. Falconer. Die heutige paläontologisch anwendbare Theorie, die die besonders bei Spezialisten der Haustierkunde üblichen Modelle in geeigneter Weise adaptiert hat entstammt aber den Arbeiten des Amerikaners R. Lande, (Lande, 1976, 1979, 1988). Folgende Modelle werden aufgestellt:

1. Eindimensionale Modelle

- (a) Selektion: Abwesenheit genetischer Drift
- (b) Genetische Drift; Abwesenheit von Selektionseffekten
- (c) Selektion und Drift wirken zusammen

2. Mehrdimensionale Modelle

- (a) Multivariate Darstellung von [1a] oben
- (b) Multivariate Darstellung von [1b] oben

1. Selektion und die Entwicklung des Phänotyps

Dieses Modell (1a) berücksichtigt die Entwicklung des Phänotyps bei der totalen Abwesenheit der Effekte der genetischen Drift. Die Geschwindigkeit der morphologischen Veränderung wird mittels der von Haldane (1949) eingeführten Evolutionsrate gemessen. Diese Rate verwendet als Maßstab Standardabweichungseinheiten. Zu beachten ist die Tatsache, daß β diese Annahme wegen der Abgrenzung des Modells notwendig ist. Es sei hier darauf hingewiesen, daß alle Versuche, morphologisch gegründete Evolutionsraten zu berechnen auf mathematische Schwierigkeiten stoßen, wie es Bookstein (1987, 1988) bewiesen hat. Siehe oben!

Die Anzahl der durch Selektion ausgemerzten Individuen wird mit Hilfe folgender Formel berechnet:

(72)

$$\left| \frac{z}{\sigma} \right| = \frac{h^2 t}{2\pi^{1/2}} e^{\left(\frac{-b^2}{2} \right)}$$

z ist die beobachtete phänotypische Veränderung und σ die Standardabweichung desselben. Daraus erhält man

$$b = \pm \left[-2 \ln \left(\frac{(2\pi)^{\frac{1}{2}} \left(\frac{|z|}{\sigma} \right)}{h^2 t} \right) \right]^{\frac{1}{2}}$$

(73)

Die Signifikanz von b findet man in einer Tabelle der standardisierten Normalintegrale. Der aus der Tabelle ermittelte Wert gibt die Ausmerzungsanzahl/10⁶ Individuen / Generation an.

Eine multivariante Verallgemeinerung beschafft man durch den s.g. Selektionsindex, I .

$$I = (\nabla \ln \bar{w})' z; = \Delta z' G^{-1} z$$

(74)

Der Index I stellt also eine lineare Zusammensetzung der Ausgangsvariablen dar.

Die Mittelwertvektoren der Stichproben werden für die Population bei den Generationen $\underline{0}$ (die erste Generation der Reihe) und \underline{t} (die letzte Generation der Reihe) beobachtet. Es wird angenommen, daß $\mathbf{G} \approx h^2 \mathbf{P}$, wo \mathbf{G} die genetische Kovarianzmatrix und \mathbf{P} die phänotypische Kovarianzmatrix sind. Hier stellt h eine durchschnittliche Schätzung der Heritabilität sämtlicher Variablen dar (Manly, 1985, 359).

Daraus erhält man:

$$\mathbf{a} = (\mathbf{z} \cdot (t) - \mathbf{z} \cdot (0))' \mathbf{P}^{-1} \quad (75)$$

BEISPIELE

2. Ziellose Genetische Drift

Dieses zweite Modell (von S. Wright 1928 eingeführt) nach Lande (1976) umfaßt den zwar unrealistischen Fall von genetischer Drift in kleinen Populationen, ohne irgendwelchen Selektionseffekt. Dieses Gegenstück hat sich als notwendig erwiesen, um die logische Konstruktion der mathematischen Modelle sichern zu können.

Man schätzt die effektive Populationsgröße N^* um eine beobachtete phänotypische Veränderungen durch genetische Drift allein erklären zu können. Die (approximative) Schätzung erfolgt durch Berechnung von

$$N^* = \frac{(1.96)^2 h^2 t}{|z| \sigma^2}$$

(77)

Wenn $N > N^*$, zieht man den Schluß, daß die beobachtete morphologische Veränderung größer als durch Drift allein möglich ist.

Die effektive Populationsgröße im multivariaten Fall

Die einschlägige Formel lautet wie folgt:

$$N^* = \frac{t_{\chi^2_{0,0.05}}}{(\bar{z}(t) - \bar{z}(0))' G^{-1} (\bar{z}(t) - \bar{z}(0))}$$

(78)

Auch hier kann die Annäherung $\mathbf{G} \approx \mathbf{P}$ bei günstigen Fällen benutzt werden, was von großer praktischer Bedeutung ist. Sie werden sehen, daß die statistische Form des Nenners der Gleichung (77) einen generalisierten statistischen Abstand darstellt, (vgl. S. 000). Dieser Abstand mißt den Formenwandel der Evolutionsreihe über das in Frage stehende Zeitintervall. Dieselbe Beobachtung gilt für den Selektionsindex, wo derselbe Abstand zur Anwendung kommt. Man könnte somit behaupten, daß β der ganze Entwicklungsverlauf, dem Modell nach, durch zwei Stichproben zu beurteilen sei!

BEISPIELE

Gültigkeit des regionalen Phänotyps

Eine wichtige Frage der quantitativen Genetik bezieht sich auf die regionale Gültigkeit des multivariaten Phänotyps. Die herkömmlichen Methoden der Multivariatstatistik werden verwendet, um die Hypothese der Einstimmigkeit zu prüfen. Dazu braucht man den Vektor \underline{a} des vorstehenden Abschnittes. Die Prüfung läßt sich am leichtesten durch die Berechnung des multivariaten Winkels von Mahalanobis ausführen.

Der Winkel zwischen zwei Vektoren \mathbf{x}_r und \mathbf{x}_s erhält man mit Hilfe der folgenden Formel, eine Adaption der bekannten Formel der Vektorrechnung:

$$\cos\theta = \left(\frac{x_r'}{S^{-1}x_s} D(x_r', 0) \cdot D(x_s', 0) \right)^{\frac{1}{2}}$$

(78)

wo wir definieren:

$$D(x_r - x_s)' = [(x_r - x_s)' S^{-1} (x_r - x_s)]^{\frac{1}{2}}$$

(79)

Sechster Teil: **GRÖSSE, ZUWACHS UND FORM**

1. Historische Einleitung

Die Darstellung D. W. Thompsons: Formveränderung als eine Verbiegung oder Deformation.

2. Allgemeiner Ueberblick der Fragestellungen.

(i) die Eigenlösung: die Aufspaltung in wenigstens zwei Teile: d.h. Größe- und Gestaltvektoren.

(ii) Umrissmethoden (Form als Schattenbild betrachtet, und durch beliebig gewählte Koordinatenpaare dargestellt); Beispiel: ein taxonomisches Scherzrätsel bei Ammoniten.

(iii) Analyse mittels Formtensoren und das Spalten der Variation in zwei Teilen, d.h.:

Ein regelmä ßiger - gleichförmiger - Teil (die affine Transformation), und

Ein unregelmä ßiger - ungleichförmiger - Teil (die nicht-affine Transformation):

(Siehe Abschnitt unter **Differentialgeometrie**).

3. "Prokrustes-Anpäßung". Die zwangsläufige Anpäßung einer Form an eine andere. Diese Lösung beachtet nicht, daß jede Veränderung der biologischen Form aus einem affinen und einem nicht-affinen Teil besteht.

4. Die generalisierten Spline-Hauptkomponenten J.-F. Durands. Das Suchen nach einer optimalen Form.

5. Die geometrische Morphometrie

Bezugspunkte (Anhaltspunkte): durch Koordinatenpaare angegebene Variablen (wovon die geometrische Behandlungsweise herstammt).

a. Analyse des Anhaltspunktes im Verhältnis zu einer geometrischen Grundlinie (Mittels Booksteinvariablen: nach K. Mardia).

b. Theorie der Verbiegung einer dünnen Metallschicht als mathematische Einschiebungsmethode (Théorie des plaques minces)

nach Duchon).

c. Analyse in bezug auf Hauptverbiegungen.

d. Analyse in bezug auf relative Verbiegungen (Analoge der Hauptkomponentenanalyse).

e. Berechnung der Achsenlängen der affinen Ellipsoiden (biorthogonale Gitter).

Beispiele:-

Publizierte Beispiele sind:- Die amerikanischen Kinder (Ann Arbor, Michigan): regelmäßig ausgeführte Röntgenaufnahmen des Schädels unter Zuwachs.

Syndrom von Apert- Zuwachsdeformation des Vorderteils des Schädels.

Die dänischen Ratten Vilmanns.

Die Mäuschen von Montpellier. Beleuchtung des Dünoblechmodells und die Spline-Transformation.

Krustacéen: Ostrakoden und Krabben: Bedeutung der affinen Transformation für Schalenträger, besonders Krustaceen, und Polymorphismus.

Beziehungen zur Strukturgeologie und die statistisch orientierten tensoranalytischen Darlegungen D. S. Watsons. Schon Ernst Cloos hatte vor 50 Jahren, hatte gute Einsichten in die mechanischen Grundbegriffe.

Grösse und Form

Die algebraische Lösung

Der Franzose Teissier versuchte im Jahre 1938 eine Lösung der Zuwachsfrage auf eine Formel zu bringen. Widmen wir uns dieser Lösung einige Minuten. Teissier wollte eine multidimensionelle Darstellung der Formenvariation bei Krabben finden. Die damals ziemlich neue Methode der Faktoranalyse schien ihm eine bequeme Möglichkeit anzubieten. Mit der faktoriellen Darstellung konnte er die Verhältnisse zwischen Variablen bestimmen. Betrachte die faktorielle Zerlegung:

$$y = a_1 X_1 + a_2 X_2 + \dots + a_p X_p \quad (80)$$

Die **x** sind die Variablen, während die **a** die Faktor - Koeffizienten- sind. Die Variablen sind fast immer **ABSTANDSMERKMALE**. Die Koeffizienten (Masse) stellen die Beiträge jeder Messvariable zur Gesamtvariation dar. Teissier konnte nur einen Vektor benutzen (Schwierigkeit verursacht durch Berechnungsschwierigkeiten).

In der Tat hat er eigentlich keine Faktoranalyse geschaffen, sondern eine primitive Hauptkomponentenanalyse.

Der zweite Akt kommt jetzt.

Die Lösung Jolicoeurs

Pierre Jolicoeur (Kanada) hat die Ideen von Teissier weiterentwickelt. Er hat

a. einen Größenvektor anerkannt dessen sämtliche Elemente positiv seien;

b. und einen Formvektor, mit positiven und negativen Elementen angenommen.

Die Deutung Jolicoeurs ist nicht ohne Logik, es mangelt ihr jedoch an schlußrichtigen Folgen. Das bezieht sich auf folgende Tatsachen: Betrachte die Rolle des zweiten Eigenvektors!

$$y_2 = a_1x_1 - \dots + \dots - \dots + a_px_p \quad (81)$$

und so weiter:

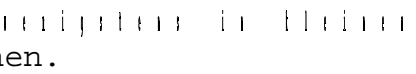
Es gibt aber noch einen dritten Vektor und einen p-ten Vektor. Also eine beliebige Anzahl von formspezifizierenden Vektoren. Diese Deutung wird dadurch logisch unhaltbar. Die Gegenüberstellung von positiven Vorzeichen und negativen Vorzeichen veranlasste die Deutung dieses Vektors als eine Art von Maßstab der biologischen Formveränderlichkeit. Die Methode ist jedoch nicht ganz unverwendbar, da man sie benutzen kann, um *redundante Variablen* zu finden, d. h. die Variablen welche wenig Information enthalten. Die automatisch eingesetzte algebraische Lösung muss durch eine spezifische Lösung ersetzt werden. Eine solche Lösung, basiert auf Differentialgeometrie, ist vor 10 Jahren von Fred L. Bookstein abgeleitet worden.

Die Geometrische Morphometrie

In diesem Abschnitt behandle ich die Analyse von Zuwachs, Größe und Form (Gestalt werde ich von jetzt an sagen). Wichtige Begriffe der Gestaltanalyse sind Umriß und **Landmarke** (= Referenzpunkte an mehr oder wenige homologen Stellen des Organismus).

Eine der klassischen Illustrationen von D'Arcy Wentworth Thompson (1917)- die Fische - findet man in beinahe jeder allometrischen Arbeit. Diese Figur befindet sich auch in fast allen Lehrbüchern über Zuwachs und Form. Es hat sich herausgestellt, daß Thompson seine Figuren nur mit Hilfe seiner Augen konstruiert hat- aber man hat jahrelang in seinen Papieren nach der Konstruktionsformel gesucht **ehe man die Wahrheit begriff.**

Grundbegriffe

Zu beachten ist, daß diese Theorie  Es gibt zwei Transformationen.

1. Die affine Transformation

Bei einer Parallelverschiebung verbleiben die Linien eines Viereckes parallel zu einander. Siehe das Beispiel: Viereck zu Parallelogram. Die Formen sind projektiv-äquivalent.

2. Die nicht-affine Transformation

wobei eine nicht-affine Umformung erreicht wird. Eine unregelmässig entwickelte Veränderung einer Gestalt.

Die Formveränderung zeigt sich dadurch, daß die kleinen Vierecke des Gitters deformiert werden. Die räumliche Verteilung der kleinen Vierecke des Gitters er bietet eine gute Möglichkeit für die Darstellung des Zuwachsmusters.

Beispiele aus der Literatur

1. Die Schädel von Thompson (1917) und auch Sneath (1967).

Hier wird die Form einer dünnen Blechschablone beim Vergleich der einen Form mit einer anderen dargestellt, eine Art von "Formieren", um die Anpassung der einen Gestalt der anderen auszudrücken. Zu beachten ist die Zerlegung des affinen Teiles in Eigenvektoren, welche die Verbiegungen in verschiedenen Teilen des figurativen Bleches darstellen.

Ontogenetische Entwicklung des ventralen Rippchens bei der oligozänen Ostrakoden-Art *Neobuntonia airella* Australiens und die morphologische Variation bei der Ostrakodenart *Hemicytherura fulva*, Eozän, Australien.

Die Variablen

Wie sehen die geometrischen Variablen aus?

FIGUR: Die drei Landmarkskategorien mittels *Mutilus* dargestellt:

1. Landmarken- Referenzpunkte
2. Blums Referenzpunkte
3. normale Distanzvariablen

Man kann, wenn man will, die einzelnen Landmarken und ihre Lagen mittels der Methode von formbestimmenden Dreiecken untersuchen, wobei die Begriffe der Tensoranalyse benutzt werden. Eine Standlinie wird gewählt. Dann folgt man den Schwankungen der übrigen Landmarken im Verhältnis zu dieser Referenzlinie.

Zu Beachten: Die Deformationen der Gevierte des Maschenwerkes.

Eine sehr verwendbare Darstellungsweise.

Nachtrag:- Zur Analyse der Gehäusegestalt

Die Methode der s.g. Eigenformen (Lohmann, 1983):

Eine Methode welche der Umriss (= Kreislinie) eines Fossils mit auf gleichen Abständen gelegenen Koordinaten versieht und diese als Eingangswerte einer Eigenanalyse benutzt. Sie unterscheidet sich von der geometrischen Darstellungsweise, dadurch, daß die Koordinaten beliebig gewählt worden sind. Sie sind nicht mit biologisch auswertbaren Landmarken, Strukturen, usw verknüpft.

Beispiel - Echinocythereis

Nachteil: Dieses Verfahren ist nur als Ordinierungsmethode gebrauchbar. Es kann nicht mathematisch weiterentwickelt werden. Eine etwa ähnliche Methode ist die Anwendung der Umrißkoordinaten in einer s. g. Fourieranalyse (Rohlf, 1986).

BOOKSTEINVARIABLEN

In diesem Abschnitt betrachten wir die von Mardia (1990) als Booksteinvariablen bezeichnete Verwendung von Landmarken. Bei der Analyse und Deutung kleiner Gestaltvariationen für N Landmarken, hat es sich herausgestellt, daß ein sehr brauchbares Verfahren darin besteht, die Landmarken zu drei zu analysieren. Für N Landmarken braucht man $N-2$ Dreiecke: zwei der Landmarken haben die Funktion einer Standlinie (siehe oben!).

Betrachten wir nun die drei Landmarken A, B und C, wobei A und B die Endpunkte einer Standlinie darstellen, kann man die Gestaltkoordinaten der digitalisierten Punkte A, B, und C auf folgende Weise repräsentieren:

$$\varepsilon = \frac{(x_B - x_A)(x_C - x_A) + (y_B - y_A)(y_C - y_A)}{(x_B - x_A)^2 + (y_B - y_A)^2} \quad (82)$$

und

$$\nu = \frac{(x_B - x_A)(y_C - y_A) + (y_B - y_A)(x_C - x_A)}{(x_B - x_A)^2 + (y_B - y_A)^2} \quad (83)$$

Wenn man die Größenvariation auch umfassen möchte, braucht man

nur die Quadratwurzel der Mittelwerte der quadrierten Abstände der Landmarken von ihrem Baryzentrum einzuschließen.

BEISPIEL: Analyse der Mäuse
Mus domesticus domesticus
Mus domesticus spetus

Die Hauptachsen einer Verformung

Durch Verschiebung wird ein in die Verschiebungsebene eingezeichneter Kreis in eine Ellipse (die **Verformungsellipse**) umgewandelt. Hierbei erfolgen in zwei zueinander senkrechten Richtungen der Verschiebungsebene nur Längenänderungen (die Hauptachsen der Ellipse). Diese Achsen heißen die **Hauptspannungen**.

Die Summe der Logarithmen der Längen der Verschiebungsachsen $\log d_1 + \log d_2$ wird **Größenkomponente** genannt [d_1 ist die lange Hauptachse der Ellipse und d_2 ist ihre kurze Hauptachse].

Die Differenz dieser beiden Quantitäten $\log d_1 - \log d_2$ wird die **Gestaltkomponente (= Formkomponente)** genannt.

Berechnung der Achsenlängen der Dehnungsellipse

Die Berechnungen für die affine Transformation lassen sich am besten algorithmisch darstellen.

1. Beginne mit der affinen Transformationsmatrix

$$T = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}$$

(84)

NB t_{12} und t_{21} sind verschieden. Die Orientierungen oder Dehnungsrichtungen der Hauptspannungen ergeben sich durch Lösung der quadratischen Gleichung

$$D\alpha^2 + (L_1^2 - L_2^2)\alpha - D = 0$$

(85)

WO

$$\begin{aligned} L_1^2 &= t_{11}^2 + t_{12}^2 \\ L_2^2 &= t_{21}^2 + t_{22}^2 \\ D &= t_{11}t_{21} + t_{12}t_{22} \end{aligned}$$

Da die Gleichung quadratisch ist, hat man somit zwei Werte, α_1 und α_2 , welche die entsprechenden Winkel ergeben: $\Theta_i = \tan^{-1}\alpha_i$.

Die Spannungsquotienten erhält man auch durch Lösung einer quadratischen Gleichung, nämlich:

$$\mu^2 - (L_1^2 + L_2^2)\mu + A^2 = 0 \quad (86)$$

wobei $A^2 = L_1^2 L_2^2 - D$. Durch die quadratische Lösung erhält man wieder zwei Werte: die Quadratwurzeln von μ geben das erwünschte Ergebnis.

|||||

Vorformung als morphometrisches Merkmal

Hier führe ich einige Definitionen an.

1. Jede Verformung eines Organismen besteht aus zwei Teilen
 - (a) Einem uniformen Teil: die **affine** Umformung.
 - (b) Einem nichtuniformen Teil: die **nicht affine** Umformung.

Jede Stichprobe von Landmarken (Referenzpunkten) umfaßt die Variation in beiden diesen Teilen, welche um einen Mittelwert herum geschieht (die Gestaltung der Mittelwerte).

2. Der uniforme Teil und der nichtuniforme Teil jeder Verformung stellen ergänzende Untermengen der vollständigen Mannigfaltigkeit der Gestaltvariation dar.

3. Der uniforme Teil kann als Variation eines typischen Dreieckes betrachtet werden, wobei jede Landmarke die gleiche Behandlung erhält. Der uniforme Teil der Verformung ist durch die anisotropischen Verhältnisse kalibriert (d.h. der Quotient der Hauptspannungen d_1 und d_2).

4. Die beiden Teile einer Verformung sind unermesslich (*math.* mit dem gleichen Maß nicht meßbar), weil der uniforme Teil eine Anisotropie darstellt, während der nicht uniforme Teil in mathematischer Hinsicht eine **Energie** repräsentiert. Hinzufügen sei, daß sich auch die dritte Komponente einer Verformung, also die **Größe**, mit diesen beiden Komponenten nicht vereinen läßt.

Die Theorie der Eigenverformung

Die Theorie der Eigenverformung basiert auf den Eigenschaften der Dünnsblechinterpolation (Dünnsblechspline), einer allgemeinen **Interpolationsfunktion**. Jede Verschiebung einer bestimmten Menge Landmarken ist mit einer bildlichen **Biegungsenergie** verbunden, d.h. die Nettokraft ist notwendig, um ein unendliches, unendlich dünnes Blech, gelegen über diesen Landmarken, zu biegen, damit die Höhe jeder Landmarke erstens der X-Koordinate und, zweitens der Y-Koordinate der entsprechenden Landmarke einer zweiten Menge

homologer Landmarken gleicht.

Eine affine Verschiebung geschieht allein durch Kippung der Platte, wobei die Biegekraft null beträgt.

Eine nicht-affine Verschiebung kann als die finite Summe der Verformungskomponenten dargestellt werden.

Die Einzelheiten der Methode

Man gebraucht eine Funktion welche die Fläche

$$z(x,y) = -U(r) = -r^2 \log r^2 \quad [87]$$

beschreibt.

Hier ist r^2 der Abstand $(x^2 + y^2)^{\frac{1}{2}}$ des dem Cartesianischen Koordinatenanfangspunkt. Die Funktion U wird bestimmt wie folgt:

$$\Delta^2 U = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)^2 U \propto \delta_{0,0}$$

(88)

Diese Funktion ist überall null, abgesehen vom Koordinatenanfangspunkt. Sie beschreibt die Verformung eines infinit dünnen Bleches, am Anfang ganz platt, aber nun zwangsläufig nach bestimmten Abständen, über und unter der bestimmten Menge von Landmarken, angepaßt. Die morphometrische Anwendung der Funktion U ergibt eine glatte Transformation, die den Lagen der Landmarken genau gleicht. Die Biegekraftmatrix wird auf folgende Weise konstruiert:

(1) Wir definieren die folgende Matrix

$$P_k = \begin{pmatrix} 0 & U_{(x_{12})} & \dots & U_{(x_{1k})} \\ U_{(x_{12})} & 0 & \dots & U_{(x_{k2})} \\ \dots & \dots & \dots & \\ U_{(x_{k1})} & U_{(x_{k2})} & \dots & 0 \end{pmatrix}$$

(89)

Die Ordnung dieser Matrix ist $k \times k$. Betrachten wir nun k Punkte

$\mathbf{z}_1 = (x_1, y_1)$, $\mathbf{z}_2 = (x_2, y_2)$ gelegen in der Euklidischen Ebene.

(2) Definiere dann eine Matrix \mathbf{Q}

$$Q = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \dots & \dots & \dots \\ 1 & x_k & y_k \end{pmatrix}$$

(90)

und auch (3) die Matrix L

$$L = \begin{pmatrix} P_k & Q \\ Q' & 0 \end{pmatrix}$$

(91)

Q ist $k \times 3$ und 0 3×3 , eine Nullmatrix.

(4) Die Biegungskraftmatrix ist L^{-1}_k , d.h. die obere linke Untermatrix von L^{-1} in Stufe 3 oben.

(5) Die Eigenvektoren dieser Matrix machen die Koeffizienten einer zur Basalebene befestigten Dünnspläne aus. Sie sind die Koeffizienten der Funktion U und sind an den k Landmarken gelegen.

Einige Bemerkungen zur Methodik:

(1) Die Eigenverformungen können benutzt werden, um die nichttrivialen Eigenwerte der Biegungskraftmatrix zu deuten (Beachte: es gibt stets drei triviale Eigenwerte weil 0 in der Matrix L oben von der Ordnung 3×3 ist).

Der nichttriviale Eigenwert jedes Eigenvektors kann als eine spezifische Biegungskraft gedeutet werden - d.h. die partielle Verformung des entsprechenden Eigenvektors, wenn jeder Koeffizient als die Höhe über der entsprechender Landmarke betrachtet wird.

2. Die nichttrivialen Teile jeder Verformung können auf diese Weise zerlegt werden, also als die Summe der **partiellen Verformungen** der Dünnspläne.

BEISPIEL

Triassische Ichthyosaurien (Reyment, 1991)

Die Ichthyosaurien waren eine Gruppe mariner Saurier, welche von der Untertrias bis zur Oberkreide vorkommt. Die frühtriassischen Arten unterscheiden sich hinsichtlich der

Schädelmorphologie von späteren Formen. Die Gattung *Grippia* aus dem Spath Svalbards wird was sowohl Schädel als auch Extremitäten als Urform betrachtet. Es könnte daher interessant sein, die Schädelform dieser Gattung mit der einer jüngeren Gattung, *Cymbospondylus*, aus der Mitteltrias zu vergleichen.

Die Methode der Dünnsplinespläne er bietet ein verwendbares Vorgehen, weil sie eine auf evolutionären Erscheinungen gegründete Analyse erlaubt.

Die Analyse:

Vierzehn Landmarken und Pseudolandmarken wurden gewählt. Die Analyse wird in zwei Teilen gemacht: (1) die affine Verformung und, (2) die nichtaffine Verformung.

Die Affine Verformung

Folgende Matrix enthält die notwendige Information, um die affine Transformation zu berechnen:

$$\begin{pmatrix} 1.1742 & 0.5744 \\ 0.1666 & 1.1329 \end{pmatrix}$$

Setzt man diese Werte in die diesbezügliche quadratische Gleichung, ergibt sich die Hauptspannungsrichtungen.

1. *Die Hauptspannungsrichtungen:*

Die diesbezügliche quadratische Gleichung lautet wie folgt:

$$\begin{aligned} L_1^2 &= 1.1742^2 + 0.5744^2 \\ L_2^2 &= 0.1666^2 + 1.1329^2 \\ D &= (1.1742)(0.1666) + (0.5744)(1.1329) \end{aligned}$$

Die Lösungen sind dann $\alpha = 0.78186$, -0.84637 welche den Winkel 38.39° und -51.61° . 2 entsprechen.

Die Spannungsquotiente: Diese Werte erhält man aus der quadratische Gleichung unten:

$$\mu^2 - (L_1^2 + L_2^2)\mu + A^2 = 0$$

wobei $A^2 = L_1^2 L_2^2 - D$.

Die erwünschten Quotiente sind:
 $\lambda = \mu = 1.543$ und 0.800 . Die längere dieser Achsen stellt die erste Hauptachse der Spannungselipse dar, die kürzere ist die zweite Hauptachse.

Der nichtaffine Teil

Die drei ersten Eigenwerte und Eigenvektoren der Biegunskraftmatrix (= die Hauptverformung) sind wie folgt.

Eigenwerte	Eigenvektoren
1. 1.2568	(0.00459, -0.00631, 0.01046, -0.19193, 0.25524, -0.74749, 0.57649, 0.01407, 0.00600, -0.01206, 0.00264, 0.00220, 0.00655, 0.07955) ^T
2. 0.63462	(0.00092, 0.01615, -0.30056, 0.56485, -0.65621, -0.12938, 0.28249, -0.02478, 0.02219, -0.01239, 0.00432, 0.00272, -0.01846, 0.24814) ^T
3. 0.33199	(-0.25942, 0.06562, 0.02915, -0.10171, -0.01082, 0.05140, 0.02599, -0.46633, 0.70512, -0.26656, -0.01445, 0.30628, -0.15858, 0.09430) ^T

Die Projektionen der Spline f_x und f_y (die X- und Y-Komponenten der berechneten Deformation) auf die vier ersten Hauptverformungen sind:

Nr.	Gewichte der Hauptverform.		Kraft jeder Teilverformung
1	-0.08046	-0.23414	0.077035
2	-0.21632	-0.30475	0.088636
3	2.21687	0.20874	1.645053
4	0.41752	0.08382	0.053132

Die Biegunskraftmatrix hat 11 nichttriviale Eigenwerte und drei triviale Eigenwerte. Die letzteren entsprechen dem Muster der Landmarkenverschiebung des affinen Teiles der Verformung.

Die Teilverformungen: Der Beitrag jeder Teilverformung zum Ganzen wird auf folgende Weise berechnet: Im Falle des ersten Eigenwertes haben wir:

$$1.256832 \times (-0.08046^2 + (-0.23414)^2) = 0.077035.$$

Das heisst: erster Eigenwert mal zweites und drittes Gewicht der Hauptverformung.

Relative Verformungen

Im Vorgehenden haben wir die Zerlegung von einzelnen Veränderungen einer Gestaltung von zwei Mengen durchschnittlicher Landmarken untersucht: also die Mittelwerte der Landmarken unserer beiden Stichproben (beispielsweise Stichproben je zweier Arten).

Diese Methodik läßt sich ohne weiteres zur Analogie der Hauptkomponentenanalyse erweitern - um die Beschreibung der innerhalb einer einzigen Stichprobe vorkommenden Variation zu erlauben.

Jedes Exemplar einer Stichprobe läßt sich als seine Anpaßung

3:	0.08687	-0.30189	0.38761	-0.19671	0.03121	-0.00054	0.01140
-0.01796							
4:	0.00814	0.07213	-0.19671	0.20474	-0.09068	0.03585	-0.01616
-0.01730							
5:	0.01752	-0.00369	0.03121	-0.09068	0.07475	-0.04457	0.00823
0.00724							
6:	-0.01381	0.01249	-0.00054	0.03585	-0.04457	0.12901	-0.17461
0.05618							
7:	0.06088	-0.00941	0.01140	-0.01616	0.00823	-0.17461	0.38434
-0.26467							
8:	-0.22761	0.06727	-0.01796	-0.01730	0.00724	0.05618	-0.26467
0.39684							

Hauptverformungen

8	8						
1:	-0.41962	-0.07046	-0.48199	0.25960	0.38650	-0.08124	0.54541
0.25651							
2:	0.50079	-0.36672	0.22923	0.41362	-0.22795	-0.37961	0.44134
0.01548							
3:	-0.44997	0.49042	0.23631	-0.07175	-0.39729	-0.42675	0.29937
-0.25779							
4:	0.18139	-0.25668	-0.49131	-0.52611	0.01279	-0.25947	0.13173
-0.54607							
5:	-0.05392	0.04896	0.22173	0.42597	0.45703	0.17671	0.00092
-0.72386							
6:	0.13485	0.18986	-0.36087	0.21090	-0.53494	0.62337	0.26035
-0.16888							
7:	-0.34110	-0.52518	0.44443	-0.35279	-0.07105	0.39370	0.35344
-0.02989							
8:	0.44757	0.48980	0.20247	-0.35943	0.37491	0.15823	0.45615
0.12621							

Eigenwerte

0.914306467 0.676595135 0.362703067 0.153584693 0.118567865 (NB Drei sind ja trivial)

Durch Eigenvektoren Rangierte Individuen

N= 21 Landmarken= 8 (Koordinatenpaare)

Individuen	Koordinaten										
1,1:	0.0	-0.6809	-0.2431	0.7610	2.4931	3.1450	2.1409	1.0	1,2:	0.0	1.0622
1,2:	1.8812	2.4599	2.3688	0.2555	0.1768	0.0	2,1	0.0	-0.4798	-0.0424	1.0173
2,1:	2.6705	3.0039	2.0250	1.0	2,2	0.0	1.0482	1.7245	2.2794	2.0250	0.0250
2,2:	-0.0039	0.0	3,1:	0.0	-0.5412	-0.3235	0.8941	2.5765	3.0569	2.0804	1.
3,1:	0.0	0.964	1.7941	2.2902	2.1608	0.0725	0.0451	0.			
4,1:	0.0	-0.5636	-0.3008	0.7616	2.4679	3.0529	2.0517	1.			
4,2:	0.0	0.8918	1.6659	2.2253	2.1837	0.1694	0.1207	0.			
5,1:	0.0	-0.5386	-0.1732	0.8715	2.6111	3.1255	2.0819	1.0			
5,2:	0.0	0.9794	1.7242	2.1833	2.0136	0.0365	0.0896	0.			
6,1:	0.0	-0.6409	-0.3023	0.8259	2.6854	3.1959	2.1122	1.			
6,2:	0.0	1.1428	1.9082	2.2797	2.1792	0.1245	0.0750	0.			
7,1:	0.0	-0.4680	-0.1740	0.8686	2.5183	2.9550	1.9343	1.0			
7,2:	0.0	0.9367	1.6225	2.2047	2.0840	0.0521	0.1024	0.0			
8,1:	0.0	-0.5372	-0.2845	0.8801	2.6700	3.0296	2.0534	1.0			
8,2:	0.0	1.0310	1.8686	2.3913	2.1978	0.0238	-0.0058	0.0			
9,1:	0.0	-0.5028	-0.1771	0.7874	2.5277	2.8794	1.9273	1.0			
9,2:	0.0	0.9627	1.6383	2.1659	2.0354	0.0590	0.0305	0.0			
10,1:	0.0	-0.5632	-0.2101	0.8148	2.3478	3.0083	1.9329	1.0			
10,2:	0.0	0.9543	1.6932	2.2042	2.1294	0.1703	0.1252	0.0			
11,1:	0.0	-0.3490	-0.0421	1.0545	2.6881	3.1361	2.0569	1.0			
11,2:	0.0	1.0099	1.8292	2.2946	2.1312	0.1114	0.0693	0.0			
12,1:	0.0	-0.5134	-0.0575	0.9976	2.6357	2.8778	1.9756	1.0			
12,2:	0.0	1.0770	1.7836	2.1504	1.9046	0.0183	0.0037	0.0			
13,1:	0.0	-0.5353	-0.1490	0.8078	2.5412	3.0157	2.1627	1.0			
13,2:	0.0	1.0412	1.7294	2.1686	2.0353	0.2039	0.1157	0.0			
14,1:	0.0	-0.5025	-0.2277	0.8688	2.6757	2.8787	1.9530	1.0			
14,2:	0.0	0.9752	1.7228	2.1881	2.0074	0.0371	0.0297	0.0			
15,1:	0.0	-0.6468	-0.3461	0.6991	2.5890	3.0102	2.0330	1.0			
15,2:	0.0	0.8912	1.8274	2.3881	2.3268	0.3132	0.1827	0.0			
16,1:	0.0	-0.4793	-0.0511	1.0323	2.747	3.0537	2.0821	1.0			
16,2:	0.0	1.0614	1.8014	2.2523	2.0323	0.0188	0.0408	0.0			

17,1:0.0	-0.5717	-0.1496	0.8484	2.4713	2.9754	2.0123	1.0
17,2:0.0	0.9611	1.6045	2.0820	1.9344	0.2295	0.1352	0.0
18,1:0.0	-0.4027	-0.0353	0.8736	2.5492	2.9157	1.9339	1.0
18,2:0.0	0.9691	1.6031	2.0779	1.9290	0.0519	0.0059	0.0
19,1:0.0	-0.5198	-0.1634	0.9901	2.7228	3.0916	2.1361	1.0
19,2:0.0	1.0520	1.8663	2.4010	2.2277	0.1658	0.1114	0.0
20,1:0.0	-0.5105	-0.2704	0.7523	2.4695	2.8904	1.9557	1.0
20,2:0.0	1.0237	1.6870	2.1743	2.0944	0.1446	0.0486	0.0
21,1:0.0	-0.3308	0.0402	1.0553	2.7985	3.0850	2.0681	1.0
21,2:0.0	1.0728	1.7746	2.2114	1.9563	-0.0856	-0.0344	0.0

Gewichte für Individuen (Zeilen) & Verformungen (Spalten)

1,1:	-0.01889	0.00432	-0.00342	-0.01180	-0.01280
1,2:	-0.00478	-0.00583	-0.00196	0.00233	-0.00539
2,1:	-0.00409	0.00466	-0.00202	-0.00730	-0.00352
2,2:	0.01264	-0.00389	-0.00739	-0.00440	0.00382
3,1:	0.01047	-0.01876	-0.00973	-0.00485	0.00714
3,2:	-0.00673	0.00745	-0.00155	-0.00323	0.00919
4,1:	0.00384	-0.00515	-0.00320	-0.00009	-0.00401
4,2:	-0.00772	-0.00121	-0.00405	0.00042	0.01356
5,1:	-0.00340	0.00142	-0.00302	0.00188	-0.01067
5,2:	-0.00514	-0.00338	0.00707	-0.00730	0.00306
6,1:	-0.00477	-0.00139	-0.00892	0.00771	0.00420
6,2:	0.00019	0.00492	0.01406	0.01406	-0.01428
7,1:	0.01133	0.00472	-0.00459	0.00349	-0.00118
7,2:	-0.00029	-0.01212	0.00081	-0.00574	0.01908
8,1:	0.00662	-0.01115	-0.00196	0.00602	0.01645
8,2:	-0.00088	0.00865	-0.00454	-0.00490	0.00989
9,1:	0.00282	0.00949	0.00793	0.00765	0.01046
9,2:	0.00560	-0.00233	-0.00279	0.00215	0.01176
10,1:	0.00603	0.01203	-0.01803	-0.01166	-0.01668
10,2:	-0.00408	-0.00317	0.00050	0.00270	0.00243
11,1:	0.01295	-0.00583	-0.00591	0.00551	-0.02413
11,2:	-0.00538	0.00586	-0.00009	-0.00303	-0.00394
12,1:	-0.00669	0.00693	0.00109	-0.01306	0.01142
12,2:	0.00554	0.00493	0.00637	0.00001	-0.01445
13,1:	-0.01636	-0.00724	0.01833	-0.00866	-0.00851
13,2:	0.00325	-0.00278	0.00240	0.00759	-0.01828
14,1:	0.00737	-0.00190	0.00633	0.01088	0.02951
14,2:	-0.00071	0.00334	0.00011	-0.00414	0.00365
15,1:	-0.00456	0.00183	0.00441	0.01103	0.02124
15,2:	-0.01927	0.00738	-0.01158	-0.00571	-0.00343
16,1:	-0.00628	-0.00080	0.00130	-0.00376	-0.00175
16,2:	0.00187	-0.00049	0.00533	-0.00434	-0.00282
17,1:	-0.00785	0.00817	-0.00266	-0.01336	-0.00647
17,2:	0.00382	-0.00714	-0.00376	0.00256	-0.01706
18,1:	0.00356	0.01282	0.00984	0.00774	-0.00876
18,2:	0.00914	-0.00029	-0.00182	0.00506	0.00483
19,1:	-0.00397	-0.01209	-0.00095	-0.00554	0.00213
19,2:	-0.00288	-0.00103	-0.00302	-0.00379	-0.00634
20,1:	0.00863	-0.00133	0.00559	0.00536	0.01130
20,2:	0.00829	-0.00037	-0.00025	0.01397	0.00014
21,1:	0.00326	-0.00075	0.00960	0.01281	-0.01536
21,2:	0.00754	0.00149	0.00615	-0.00429	0.00459

Eigenwerte

0.06868	0.05669	0.04407	0.03738	0.03252	0.02886	0.02131	0.02016
0.01396	0.01188						

Die Projektionen der Individuen (Zeilen) auf gewichteten Relativverformungen (Spalten)

1:	-0.02338	0.00431	-0.00574	0.00737	-0.00733	0.00372	-0.00106
-0.00705	0.00377	0.00032					
2:	-0.00685	-0.00289	0.00540	0.01116	-0.00347	-0.00861	0.00084
0.00458	0.00543	0.00503					
3:	0.01526	-0.00762	-0.02002	-0.00310	0.00240	-0.01052	0.00113
-0.00101	0.00118	-0.00272					
4:	0.00521	-0.01313	-0.00600	0.00122	-0.00469	-0.00089	0.00271

```

-0.00742  0.00095  -0.00250
 5:  -0.00727  -0.00905  -0.00240  -0.00179  -0.00433  0.00550  -0.00974
-0.00052  -0.00152  -0.00034
 6:  -0.00365  0.01254  -0.00241  -0.01094  0.01910  0.00748  -0.00530
-0.00629  0.00292  0.00348
 7:  0.00890  -0.02090  0.00529  0.00984  -0.00059  -0.00105  -0.00482
-0.00372  -0.00690  0.00173
 8:  0.02427  -0.00049  -0.00671  -0.00166  -0.00103  -0.00546  -0.00000
0.00183  0.00495  0.00096
 9:  0.01371  -0.00095  0.01623  0.00709  -0.00128  0.00121  0.00067
-0.00253  0.00144  -0.00128
10:  -0.01710  -0.01755  -0.00736  0.01174  0.01106  0.00512  0.00319
0.00066  -0.00023  -0.00300
11:  -0.00978  -0.01957  -0.00757  -0.01672  0.00203  0.00443  0.00615
0.00755  -0.00170  0.00063
12:  -0.00609  0.02063  -0.00082  0.00738  0.00765  -0.00253  -0.00430
0.00815  0.00012  -0.00566
13:  -0.02110  0.02088  0.00422  -0.01151  -0.00990  -0.00667  0.00272
-0.00454  -0.00314  -0.00283
14:  0.03077  0.01137  0.00376  0.00160  0.00153  0.00081  -0.00307
0.00297  -0.00392  0.00044
15:  0.02155  0.01444  -0.00821  0.00237  -0.01069  0.01820  0.00517
0.00157  0.00038  0.00027
16:  -0.00597  0.00381  -0.00083  -0.00027  -0.00256  -0.00207  -0.00723
0.00171  0.00026  -0.00094
17:  -0.02182  0.01032  -0.00092  0.00868  0.00226  0.00008  0.00490
0.00386  -0.00389  0.00392
18:  -0.00320  -0.00688  0.02089  -0.00145  -0.00018  0.00328  0.00504
0.00144  0.00278  -0.00211
19:  -0.00142  0.00660  -0.01160  -0.00259  -0.00530  -0.00608  0.00037
0.00029  -0.00238  0.00367
20:  0.01118  0.00603  0.01091  -0.00337  0.01053  -0.00586  0.00770
-0.00520  -0.00207  0.00057
21:  -0.00320  -0.01189  0.01389  -0.01504  -0.00520  -0.00006  -0.00509
0.00369  0.00156  0.00035
  Quellen: Program tpsrw.exe
           Daten: Rwexmp2.dta

```

Beispiel: Die relative Verformung der Unterarten von *Mus domesticus*.

Bereich OZEANOGRAPHIE

1. Überblick des Problemfeldes.
2. Foraminiferenbeispiel J. Imbries: planktische Arten und ozeanologische Faktoren. Analysenmethode kanonische Korrespondenzanalyse als wirksames Hilfsmittel. Wird im folgenden Abschnitt berücksichtigt.
3. Daten aus Tiefseeuntersuchungen. Die von dem ODP-Program publizierten Analysen.
4. Korrelation zwischen Tiefseebohrkernen.

Bereich Ökologie

1. Einleitendes
2. Einfachere Verfahren darunter
 - kanonische Korrelation
 - Multiregression
 - Korrespondenzanalyse und die Darstellung der Ordination nach Variablen und Individuen auf der gleichen Graphik.
3. Kanonische Korrespondenzanalyse (nunmehr kanonische Gemeinschaftsanalyse) von Carl-Johan J. ter Braak.
 - Darstellung der Analysenmethode.
 - Erläuterungsbeispiel: Die Ökologie der nordholländischen Dünen (Beitrag zur Ökologie des Strandgebietes).

Statistische Ökologie

Die statistische Ökologie bildet eine natürliche Verzweigung der angewandten Multivariatanalyse. Seit Jahren hat sie ein eigenes Leben, mit Konferenzen, Seminarien und einer eigenen Literatur. Man fängt am besten mit einer allgemeinen Betrachtung des Bausteins der statistischen Ökologie an, d.h. der Frage eines geeigneten Koeffizienten. Als einfaches Beispiel können wir folgenden Koeffizienten anführen:

$$\left[\frac{a_{ij}}{(a_{ii} + a_{jj})} \right]$$

Hier bezeichnet a_{ij} wie oft die Art i zusammen mit der Art j auftritt. Ähnlichkeiten und Abstände besitzen ein gegenseitiges reziprokes Verhältnis zueinander. Folglich können Ähnlichkeiten innerhalb der Reichweite 0-1 können auf folgende Weise transformiert werden:

$$d_{ij} = 1 - s_{ij}$$

oder

$$d_{ij} = -\log s_{ij}$$

Die kanonische Korrespondenzanalyse

Diese ziemlich neu entwickelte Methode vereinigt die Prinzipien der *kanonischen Korrelation* mit denen der *Korrespondenzanalyse*. Die Methode wurde vom Holländer C. J. ter Braak, auf Grund früherer Fortschritte des Engländers M. Hill, entwickelt. Das

erhaltene Ordinationsbild vereinigt die drei Klassen (1) Orte, (2) Artfrequenzen, und (3) Milieufaktoren.

Die Berechnungen lassen sich am leichtesten algorithmisch darstellen:

1. Beginne mit beliebigen aber ungleichen Ortsrechenmaßwerten (= Ortswerte)
2. Berechne die Artrechenmaßwerte mittels gewichteter Durchschnitte der Ortswerte.
3. Berechne neue L-werte mittels gewichteter Durchschnitte der arteigenen Werte.
4. Berechne die Regressionskoeffizienten mittels gewichteter Multiregression der Ortswerte auf die Milieuwerte.
5. Berechne nun neue Ortswerte.
6. Zentriere und standardisiere die Ortswerte:
Also,

$$\sum_i y_{i+} x_i = 0; \sum_i y_{i+} x_i^2 = 1$$

7. Bei erhaltener Konvergenz aufhören.

Die damit erhaltenen Regressionskoeffizienten heißen die *kanonischen Koeffizienten*. Den Multikorrelationskoeffizient nennt man die *Art-Milieu-Korrelation*. Diese Korrelation stellt eine Art des Wirkungsgrads der Milieuvvariablen bei Ausscheidung der der Gesellschaft angehörenden Varianz dar.

Das Ordinationsdiagramm von C.C.A.

Der Einfluß der Milieuvvariablen wird durch Pfeile angegeben. Arten und Orte werden durch Pünktchen markiert.

BEISPIEL: Die friesischen Dünen

Das hier untersuchte Material entstammt einer größeren Untersuchung der Botanik der friesischen Inseln. Man hat Arten, Orte und ökologische Faktoren in einem Gebiet, in dem die Einflüsse von Ackerbau mit natürlichen Verhältnissen verglichen werden können studiert.

Dünenökologie: Arten (BATTERINK AND G. WIJFFELS, 1983)

Artenhinweise:		Eigenwerte und Eigenvektoren				N	NAME	AX1
AX2	AX3	AX4	Gewicht	N2				
EIG	.4612	.2981	.1601	.1337				
1 Ach mil	-.8402	.3816	.0275	-.3342	16.00	6.10		
2 Agr sto	.7704	-.5000	-.1143	-.0800	48.00	9.14		

3	Air pra	.7395	1.7874	-1.0769	.5319	5.00	1.92
4	Alo gen	.3541	-.9700	-.3470	.1389	36.00	6.61
5	Ant odo	-.3858	.7781	-.0414	.2262	21.00	5.73
6	Bel per	-.7122	.1691	-.2580	-.7176	13.00	5.83
7	Bro hor	-.8324	.0178	-.1609	-.8580	15.00	4.59
8	Che alb	.9308	-1.6474	-.6932	.0272	1.00	1.00
9	Cir arv	-.3976	-.8452	-.9877	-.6664	2.00	1.00
10	Ele pal	1.4970	-.0905	.5745	-.3829	25.00	4.56
11	Ely rep	-.6377	-.3819	-.3719	-.0036	26.00	5.83
12	Emp nig	1.2056	1.5490	-1.4515	.5584	2.00	1.00
13	Hyp rad	.5605	1.4863	-.7026	.2168	9.00	2.45
14	Jun art	.9606	-.0437	.0466	.3284	18.00	4.91
15	Jun buf	.2067	-.8045	-.1238	1.1062	13.00	3.76
16	Leo aut	.0001	.4308	.0323	-.0671	54.00	15.19
17	Lol per	-.6238	-.0423	-.0652	-.2815	58.00	10.51
18	Pla lan	-.8016	.5702	.5883	.1855	26.00	6.38
19	Poa pra	-.4592	-.0398	-.1293	-.1565	48.00	12.80
20	Poa tri	-.2636	-.5322	-.1299	-.0345	63.00	11.37
21	Pot pal	2.0288	.3988	2.1389	-.8602	4.00	2.00
22	Ran fla	1.3738	.0706	.1166	-.1576	14.00	5.44
23	Rum ace	-.8239	-.2073	.7901	.9617	18.00	4.15
24	Sag pro	.2991	-.4324	-.4472	.2875	20.00	6.06
25	Sal rep	.7849	1.7592	-1.0928	.5204	11.00	2.81
26	Tri pra	-1.1121	-.0586	1.0146	.9325	9.00	2.45
27	Tri rep	-.0227	.1103	.3014	-.1253	47.00	12.34
28	Vic lat	-.5373	1.0154	.3189	-.6805	4.00	2.67
29	Bra rut	.1354	.2529	.1384	.2249	49.00	12.70
30	Cal cus	1.656	.4507	.3862	-.2538	10.00	2.94

Also, dreissig Arten sind bestimmt worden.

Rechmasswerte für die Stichproben

N	NAME	AX1	AX2	AX3	AX4	WEIGHT	N2
	EIG	.4612	.2981	.1601	.1337		
1	1	-1.2192	-.4968	-.9350	-1.2524	18.00	3.77
2	2	-.8644	-.2504	-.5356	-1.7028	42.00	9.09
3	3	-.3149	-1.0096	-.9001	-.6378	40.00	8.25
4	4	-.2373	-.9306	-1.2712	-1.0467	45.00	10.07
5	5	-1.1465	.2404	.9145	.6239	43.00	11.63
6	6	-1.0310	.3684	1.7991	1.4954	48.00	10.02
7	7	-1.0342	.1512	.9626	.6486	40.00	10.81
8	8	.6975	-.7170	-.1078	-.2760	40.00	10.96
9	9	-.0533	-1.0048	-.4894	1.2413	42.00	11.31
10	10	-.9630	.6002	.2358	-1.4576	43.00	10.33
11	11	-.5075	.9320	.1462	-.5152	32.00	7.53
12	12	.3537	-1.4422	-.5238	1.8253	35.00	7.61
13	13	.4720	-1.6186	-.9074	.6554	33.00	6.76
14	14	2.0206	.2660	2.5433	-1.8041	24.00	6.00
15	15	1.9582	.0494	2.1693	-.8602	23.00	6.70
16	16	1.9316	-.6907	.8016	-.5279	33.00	6.37
17	17	-.3882	2.7700	-1.0653	.9045	15.00	6.08
28	18	-.3107	1.4947	-.1467	-.0740	27.00	7.22
29	19	.6647	2.8731	-2.6645	1.7243	31.00	7.94
30	20	2.0014	1.0029	-.2635	.3276	31.00	7.57
20	EXTRASAM	-1.4219	.0180	.2932	-.8283	.00	2.88
21	PAS 17	-.3882	2.7700	-1.0653	.9045	.00	6.08

Widerstandfähigkeit der Arten

N	NAME	AX1	AX2	AX3	AX4	RMSTOL	N2
	FR FITTED	.2180	.1409	.0757	.0632		
1	Ach mil	.3702	.7193	.8210	1.0546	78.11	6.10
2	Agr sto	.8635	.9474	1.1364	.8557	95.75	9.14
3	Air pra	.5708	.2919	.4588	.0326	39.45	1.92
4	Alo gen	.6917	.4639	.4415	1.0285	69.76	6.61
5	Ant odo	.8836	.8873	1.0289	.8303	91.05	5.73
6	Bel per	.3414	1.0117	.7499	.7062	74.18	5.83
7	Bro hor	.2606	.4668	.7869	.6916	58.81	4.59
8	Che alb	.0000	.0000	.0000	.0001	.00	1.00
9	Cir arv	.0000	.0000	.0000	.0001	.01	1.00

10	Ele pal	.4370	.9867	1.3081	.5009	88.41	4.56
11	Ely rep	.3918	.3515	.8198	1.2607	79.66	5.83
12	Emp nig	.0000	.0000	.0000	.0001	.00	1.00
13	Hyp rad	.7461	.4963	.9748	.6042	72.77	2.45
14	Jun art	.6774	.9385	1.1527	1.0693	97.62	4.91
15	Jun buf	.6017	.7421	.4596	.8448	67.78	3.76
16	Leo aut	.9908	1.0750	1.0323	.9680	101.74	15.19
17	Lol per	.4953	.6830	.8787	.9595	77.53	10.51
18	Pla lan	.3864	.8583	.6885	.8612	72.49	6.38
19	Poa pra	.5972	.8921	.7989	1.0136	83.94	12.80
20	Poa tri	.8206	.6712	.7638	1.0556	83.99	11.37
21	Pot pal	.1704	.3932	.0303	.0698	21.76	2.00
22	Ran fla	.4565	1.1967	1.4299	.5744	100.18	5.44
23	Rum ace	.5838	.4438	.5260	.7711	59.35	4.15
24	Sag pro	.6453	1.0323	.7200	.9883	86.27	6.06
25	Sal rep	.6734	.3145	.6685	.1267	50.38	2.81
26	Tri pra	.1513	.2371	.1620	.7055	38.83	2.45
27	Tri rep	1.1423	.8675	1.0401	1.1513	105.65	12.34
28	Vic lat	.1611	.7314	.7061	.5893	59.30	2.67
29	Bra rut	1.0076	1.1408	1.0389	.9475	103.61	12.70
30	Cal cus	.4499	1.0854	1.5275	.5928	100.81	2.94

BipE: Biplotwerte für die 10 ökologischen Variablen

N	NAME	AX1	AX2	AX3	AX4
	R(SPEZ,ÖK)	.9580	.9018	.8554	.8888
1	A1	.5629	-.1732	.5894	-.1094
2	Moisture	.9221	-.1702	-.1402	.1696
3	Manure	-.3092	-.7646	-.1972	-.1803
4	Hayfield	-.0756	.6046	-.2523	.2821
5	Haypastu	-.1719	-.5535	-.1320	-.0864
6	Pasture	.2795	-.0312	.4279	-.2110
7	SF	.1484	-.6956	-.4210	-.0864
8	BF	-.3645	.1750	-.0298	-.5845
9	HF	-.3611	-.1161	.4394	.5224
10	NM	.5704	.7381	.0008	.0426

CenE: Baryzentren der ökologischen Variablen

N	NAME	AX1	AX2	AX3	AX4
	R(SPEZ,ÖK)	.9580	.9018	.8554	.8888
1	A1	.2236	-.0688	.2342	-.0434
2	Moisture	.5698	-.1051	-.0866	.1048
3	Manure	-.2215	-.5478	-.1413	-.1292
4	Hayfield	-.1057	.8448	-.3526	.3942
5	Haypastu	-.2043	-.6577	-.1569	-.1027
6	Pasture	.4884	-.0545	.7477	-.3686
7	SF	.2278	-1.0681	-.6464	-.1327
8	BF	-.8030	.3856	-.0656	-1.2879
9	HF	-.5375	-.1728	.6541	.7777
10	NM	1.0726	1.3880	.0015	.0801

SamE: Rechenmaßwerte für die Stichproben: Lineare Zusammensetzungen der ökologischen Variablen

N	NAME	AX1	AX2	AX3	AX4	WEIGHT	% FIT
	EIG	.4612	.2981	.1601	.1337		
1	1	-.8862	-.4334	-1.2765	-.2324	18.00	29.38
2	2	-1.0430	.1049	-.2283	-1.6488	42.00	62.37
3	3	-.3909	-.8594	-.9477	-.6970	40.00	54.69
4	4	-.3976	-.8452	-.9877	-.6666	45.00	41.33
5	5	-1.1917	-.0737	1.2901	.0691	43.00	57.55
6	6	-1.1935	-.2242	.9844	1.5619	48.00	68.13
7	7	-.8289	.3706	.8147	.2226	40.00	66.19
8	8	.8449	-.6797	.2872	-.2363	40.00	43.45
9	9	-.1571	-.2504	-.1782	2.1014	42.00	22.60
10	10	-.7701	.4475	-.6865	-1.2146	43.00	54.72
11	11	-.5323	.6709	.9822	-.9126	32.00	14.81
12	12	.5454	-1.3139	-.1116	1.3621	35.00	56.93

13	13	.9308	-1.6474	-.6932	.0271	33.00	43.68
14	14	2.1991	.7920	2.1086	-.7906	24.00	67.78
15	15	1.8584	.0055	2.1692	-.9302	23.00	72.96
16	16	1.3985	-1.1310	.0075	-.4113	33.00	60.32
17	17	.0404	2.1449	-.5150	.4919	15.00	35.84
28	18	-.3146	2.2724	-.0025	.3176	27.00	24.73
29	19	1.2056	1.5490	-1.4515	.5585	31.00	53.83
30	20	1.1923	1.5775	-1.5317	.6193	31.00	39.33

Quellen:

Program Canoco.exe Daten: Dune.spe Dune env

Bereich TAXONOMIE

1. Einleitung
 2. Geeignete Beispiele:
 3. Taxonomische Verwendbarkeit des generalisierten Abstandes von P. C. Mahalanobis.
 4. Kanonische Variablen und die Ordination von mehreren Gruppen
 5. Taxonomische Vergleiche mittels Eigenvektoren (Verfahren von T. W. Anderson).
 6. Der Winkel von Mahalanobis als taxonomisches Hilfsmittel. Paläogeographische Anwendung dieses Winkels.
 7. Wahl von informationsreichen Variablen.
 - (i) mittels Hauptkomponentenanalyse
 - (ii) Stufenweise Regression als Ausscheidungsverfahren (Wegnehmen überflüssiger Variablen).
 8. Zuwachsfreie Diskriminanten und Abstände
-

Bereich ANTHROPOMETRIE

Kurze Einführung in die einschlägige statistische Methodik. Die Kasten Indiens (klassische, bahnbrechende Darstellung des generalisierten Abstands in rassenwissenschaftlicher Hinsicht, die Untersuchung wurde schon 1928-1930 von P. C. Mahalanobis durchgeführt).

Andere Beispiele aus der Biologie der Menschheit

Analyse von Blutgruppen als Hilfsmittel bei der historischen Forschung. Beispiel: die Geschichte der Iberischen Halbinsel. Sprachwissenschaftliche Anwendung: Aufteilung der Zigeuner in Stämme auf Grund sprachlicher Eigenheiten (Sinti, Kalderasha, Vlach, Anglo, Finnisch, Lovari, Jenische, usw). Diskussion der Paläohistorik von Spanien und Portugal.

Bereich PETROLOGIE

1. Analyse von Tabellen (Bestandteile als Oxide).
2. Beispiele von Analysen mit dazu gehöriger Graphik:
Ordination (=Einordnung) der Punkte.

Wie gut stimmen die mathematischen Eigenschaften einer derartigen Analyse mit den nur gedachten (d.h. geologisch mehr oder weniger willkürlich formulierten Ideen) überein? Läßt sich eine auf statistischen Gründen gebaute Klassifizierung der Gesteinsarten durchführen?

Einzelne Fragestellungen:

1. Die Analyse geschlossener Mengen

1. Simplexraum: dessen statistische Eigenart.
2. Beispiele geschlossener Datenbestände.
Problem der Konstantsumme eines Variablevektors
Beispiel: Organische Chemie des norwegischen Erdöls
3. Vergleich zwischen "normalen" und Simplexanalysen.

Die einschlägigen Verfahren für die Behandlung eingeschränkter Daten in chemischen Zusammenhängen finden sich in Teil X, auf S. 000.

Diese wichtige Frage wird nur ganz kurz berührt, um so zu sagen die Szene vorzubereiten. Was ist eine geschlossene Menge? Besser ausgedrückt wäre eine Menge von Variablen, wobei die Summe der Beobachtungen stets konstant ist. Der richtige mathematische Ausdruck dafür ist *Bestandteilvariablen*. Betrachte zum Beispiel Blutgruppsbestimmungen, vorschlagsweise, die Allelen **A**, **B**, **O**. Kennt man, für eine Population, die Frequenz von **O** und **A**, erhält man ohne weiteres die von **B**, durch eine ganz einfache Subtraktion. Trotzdem wir hier mit drei Variablen laborieren, ist die Lage durch zwei beliebig gewählte Variablen bestimmt. Man hat nur zwei **Freiheitsgrade**. Dazu hat man den Raum geändert. Statt des cartesianischen Raumes haben wir nun einen Simplex-Raum- das heißt einen gekrümmten Raum im Sinne Riemanns. Ein Simplex-Raum stellt einen Unterraum, oder eine Untermenge, eines Cartesianischen Raumes dar.

Hauptkomponentenanalyse von Bestandteildaten

Jeder Vektor mit nichtnegativen Komponentenn x_1, \dots, x_D , die Proportionen darstellen, nennt man geschlossen weil

$$x_1 + \dots + x_D = 1.$$

Falls man den Einheitssummezustand nicht in einer Analyse inkludiert, läuft man Gefahr, eine fehlerhafte Behandlung der Daten vorzunehmen. Beispiele geschlossener (eingeschränkter) Daten sind:

- (a) Die geochemische Zusammensetzung von Gesteinen
- (b) Blutgruppenfrequenzen, sowie auch andere polymorphe Bestimmungen.
- (c) Ökologisch bedingte Speziesvorkommen
- (d) Pollenfrequenzen
- (e) Die Zusammensetzung sedimentärer Gesteine
- (f) "Standardisierte" Vermessungen.

Als Hauptreferenz ist das Lehrbuch von Aitchison (1986) zu nennen. Der Baustein der multivariaten Bestandteilanalyse ist die Kovarianzmatrix der Logquotienten der Eingangvariablen. Wir definieren diese Kovarianzmatrix auf folgende Weise:

$$\Sigma = [\text{cov}(\log(x_i/x_D), \log(x_j/x_D)) : i, j = 1, \dots, d]$$

(92)

Die Matrix Σ ist eine Kovarianzmatrix des d -dimensionalen Zufallsvektors \mathbf{y} , definiert wie folgt:

$$y_i = \log(x_i/x_D) \quad : \quad i = 1, \dots, d \quad (93)$$

Beachte: Vorausgesetzt, daß die Logquotiententransformation eins zu eins ist, hat man das Verhältnis:

$$\mathbf{x} \in S^d \rightarrow \mathbf{y} = \log(\mathbf{x}_{-D}/x_D) \in \mathbb{R}^d$$

(94)

was sagt, daß der Vektor \mathbf{y} sowohl in der Mannigfaltigkeit \mathbb{R}^d als auch in der Simplexmangfaltigkeit S^d gilt. Der obenangeführten Kovarianzmatrix fehlt doch die begehrte Eigenschaft der Symmetrie. Die zentrierte Form dieser Matrix wird bei den meisten multivariaten Analysen verwendet, u.a. H.K.A.:

Die zentrierte Matrix gebraucht den geometrischen Mittelwert

$$g(\mathbf{x}) = (x_1, \dots, x_D)^{1/D}. \quad (95)$$

Somit erhält man:

$$\Gamma = \text{cov}[\log x_i / g(x), \log x_j / g(x)]; i, j = 1, \dots, D$$

(96)

Die eingeschränkte Hauptkomponentenanalyse basiert auf der vorgehenden Matrix. Man verwendet die übliche Methode für die Lösung des Eigenproblems.

In ähnlicher Weise kann man auch eine "normale" Hauptkoordinatenanalyse ausführen.

Beispiel: *Hauptkomponentenanalyse eingeschränkter Daten*

Verfahren von J. Aitchison (1986) und früherer Publikationen:

Die zentrierte Kovarianzmatrix wird hier benutzt. Die hier analysierte Datamatrix entstammt der Gradualavhandlung von O. Kvalheim (1987) und bezieht sich auf Phenanthrene des Nordseeöls.

(1) Die zentrierte log-Quotient- Kovarianzmatrix
Anzahl der chemischen Analysen = 36 Variablen = 11
Ausgangsmatrix (Datenmatrix): chemische Bestandteile

15.8	8.20	9.00	14.70	12.80	2.20	2.70	15.80	7.70	6.40	4.60
16.1	8.20	9.10	14.20	12.70	2.20	2.80	15.9	7.80	6.40	4.70
16.3	9.20	9.70	14.70	11.00	3.00	4.20	15.00	7.40	5.70	4.00
16.7	9.40	9.60	13.10	11.30	3.00	3.40	15.90	7.20	6.30	4.10
16.2	9.70	8.90	14.40	10.20	2.50	3.10	16.60	7.90	6.20	4.30
16.1	9.20	9.30	14.80	11.20	2.90	3.00	15.50	7.30	6.20	4.40
16.1	10.0	9.40	16.30	9.90	2.60	4.00	15.60	7.70	5.70	3.70
18.4	7.60	8.90	14.30	11.60	2.50	3.20	14.90	8.00	6.00	4.50
16.9	8.70	9.30	14.80	10.80	2.60	3.40	15.80	7.70	6.00	3.80
16.5	8.50	9.10	14.60	10.80	2.90	3.40	16.10	7.80	6.20	3.90
16.3	9.20	9.20	15.20	11.20	2.90	3.50	15.30	7.10	6.10	4.10
16.3	9.20	9.30	11.80	12.10	3.20	3.60	15.40	6.80	6.50	5.80
16.1	9.50	8.50	13.30	10.20	2.80	3.60	17.80	7.60	6.40	4.20
16.5	10.1	8.80	16.90	10.70	2.70	3.70	14.40	6.60	5.70	3.90
15.1	9.60	9.10	14.90	7.70	3.60	4.80	17.90	7.40	5.90	3.90
15.5	9.20	9.60	13.90	10.50	3.20	3.90	16.30	7.60	6.20	4.00
15.6	9.40	9.10	14.60	10.70	3.10	4.20	15.20	7.50	6.00	4.60
18.3	7.90	8.90	14.20	12.10	2.40	2.60	15.00	7.80	6.10	4.70
17.9	11.7	10.3	10.90	9.50	4.30	4.70	15.50	6.20	5.10	3.60
16.0	11.4	9.30	15.20	8.40	2.90	4.20	17.30	6.90	5.40	3.00
16.9	8.20	9.00	15.50	11.60	2.50	3.20	15.40	7.20	6.20	4.30
14.8	9.30	8.70	14.40	10.30	2.60	4.30	17.10	7.40	6.40	4.60
15.0	7.90	8.90	13.80	10.10	2.90	3.70	17.90	8.40	6.50	4.90
13.5	8.30	9.30	14.10	10.60	3.00	4.00	17.20	8.30	6.70	4.90
14.9	10.0	11.2	11.20	9.40	4.70	5.10	15.30	8.30	6.20	3.10
14.1	9.90	10.1	12.30	9.30	3.90	4.60	16.70	8.80	6.00	4.20
18.4	9.30	10.6	12.20	10.00	4.50	4.50	14.20	7.30	5.70	3.50
15.1	8.60	9.10	14.30	11.00	2.80	4.60	15.90	7.40	6.50	4.80
13.5	8.10	8.60	14.30	10.60	2.90	3.90	17.90	8.20	6.80	5.10
15.2	8.60	9.20	14.10	11.20	2.70	4.50	16.20	7.30	6.40	4.60
15.9	8.20	9.00	13.80	10.30	3.00	3.90	17.00	8.20	6.10	4.70
14.7	8.70	9.20	13.50	10.40	2.90	4.10	17.70	8.00	6.10	4.60
14.4	9.00	9.30	13.80	10.40	3.00	4.40	17.10	8.10	6.00	4.70
14.0	8.20	8.90	15.30	12.50	1.60	3.20	16.50	8.10	6.60	5.10

Hauptkomponentenanalyse nach dem eingeschränkten Modell

Eigenwerte
 0.08534 0.01962 0.01557 0.00752 0.00473 0.00419 0.00224 0.00153
 0.00068 0.00042

Spur = 0.141851
 Prozent der Spur
 60.1618 13.8347 10.9770 5.3034 3.3379 2.9543 1.5822 1.0771 0.4776 0.2939

Eigenvektoren der eingeschränkten Modelle

	1	2	3	4	5	6	7	8	9	10
1	-0.0519	0.5569	-0.1053	0.1100	-0.0048	0.2904				
2	0.2017	0.3025	0.2920	0.2088	0.4366	-0.5267				
3	0.0971	0.1531	-0.1324	-0.0925	-0.2416	-0.2722	0.0646			
4	-0.1668	0.6283	0.5394	0.0101	-0.0186	0.5544	0.4305			
5	-0.3568	0.1986	-0.2858	0.2358	-0.4084	-0.2609				
6	0.6197	0.0487	-0.4719	-0.1321	0.1719	0.3431				
7	0.4297	-0.4467	0.3230	0.4294	-0.4234	0.0182	-0.2232			
8	-0.0490	-0.2131	0.2158	-0.3060	0.4060	-0.1360				
9	-0.0962	-0.1844	0.0189	-0.6121	-0.2386	-0.0051	-0.1224			
10	-0.1530	-0.1476	-0.0554	-0.2410	-0.0652	-0.1642	0.1979			
11	-0.3959	-0.4473	-0.3507	0.3897	0.3860	0.1589				
	-0.0124	-0.2513	-0.0090	0.1981						

Rechenmaßwerte der Komponentenanalyse (für die graphische Darstellung der Ergebnisse)
 Individuum

1	-1.3780294	1.1719798	0.7323057	-0.4564434
2	-1.3611479	1.1467894	0.7178232	-0.4434303
3	-0.8336870	1.1212375	0.8358368	-0.3092515
4	-0.9292102	1.1855459	0.6992718	-0.4022458
5	-1.0952513	1.1503369	0.8545842	-0.4783346
6	-1.0616595	1.2018545	0.7173594	-0.4272105
7	-0.9022624	1.1492015	1.0559773	-0.3724207
8	-1.1905232	1.1627513	0.7023666	-0.4204137
9	-1.0193614	1.2000016	0.8566613	-0.4617005
10	-0.9714111	1.1563215	0.7896054	-0.4931802
11	-0.9698719	1.1848725	0.8043306	-0.3610202
12	-1.0048386	0.9906879	0.4791968	-0.2029763
13	-0.9450093	1.0561059	0.8270385	-0.4451731
14	-0.9457516	1.2535159	0.9662361	-0.2523599
15	-0.5369147	0.9311422	0.9763777	-0.4258447
16	-0.8131932	1.0735532	0.7843198	-0.4277312
17	-0.8665537	1.0130168	0.7951139	-0.2847652
18	-1.3279079	1.2506783	0.6353120	-0.4603948
19	-0.3134600	1.2289882	0.6784143	-0.2071709
20	-0.6045370	1.2392936	1.1508596	-0.4212957
21	-1.1678403	1.1824413	0.7957492	-0.3853465
22	-0.9666126	0.9123229	0.9189645	-0.3077898
23	-1.0193629	0.8711713	0.7381925	-0.4971848
24	-0.9691920	0.8206731	0.7546113	-0.4496884
25	-0.2284086	1.0645405	0.7017588	-0.5165164
26	-0.5515583	0.8914410	0.7288698	-0.4828749
27	-0.4119919	1.2160328	0.6045597	-0.3771288
28	-0.9416186	0.8861627	0.8207328	-0.2563747
29	-1.0359122	0.7871574	0.7669698	-0.4465570
30	-0.9562517	0.9204651	0.8352952	-0.2674277
31	-0.9460956	0.9420874	0.7434066	-0.4316769

32	-0.9166512	0.8992158	0.7990943	-0.4075891
33	-0.8672080	0.8715658	0.8148362	-0.3623993
34	-1.5510257	0.9447053	0.9529592	-0.3703415
35	-1.3245773	1.0848211	0.7617692	-0.2510157
36	-1.2903020	1.0166105	0.8543843	-0.3014592

Übliche Komponentenanalyse

Eigenwerte								
3.11712	2.49441	1.43144	0.54193	0.34753	0.18776	0.08302	0.05821	0.04177
0.01457	0.00099							

Eigenvektoren										
1	2	3	4	5	6	7	8	9	10	11
6	7	8	9	10						
1	-0.0918	0.7214	-0.0954	0.5878	-0.0403	0.0851	-0.0206	0.1033	-0.0298	0.0819
2	0.3906	0.1496	-0.3368	-0.2704	0.5915	-0.2919	0.2809	0.1120	-0.1262	0.0205
3	0.2351	0.1369	0.1492	-0.2432	-0.2337	-0.2622	-0.2645	-0.2296	0.5112	0.5062
4	-0.5018	-0.1544	-0.7381	-0.1509	-0.1952	0.0260	-0.0457	-0.1282	0.0991	-0.0074
5	-0.5292	0.1362	0.4302	-0.2862	0.2820	-0.1914	-0.2513	0.2060	0.0889	-0.3275
6	0.3225	0.1024	0.0710	-0.0910	-0.2356	0.1703	0.0530	-0.5030	0.0014	-0.6503
7	0.2911	-0.0708	-0.0515	-0.2321	-0.1801	0.5320	-0.3131	0.5917	-0.0980	0.0248
8	0.1437	-0.5530	-0.0066	0.5887	0.2795	-0.1021	-0.2822	0.0131	0.2250	-0.1009
9	-0.0055	-0.1958	0.1604	0.0933	-0.5057	-0.4127	0.5181	0.3651	-0.0846	-0.0417
10	-0.0694	-0.1289	0.1597	-0.0038	-0.0280	-0.0684	-0.2262	-0.3170	-0.7767	0.3285
11	-0.1940	-0.1407	0.2578	-0.0113	0.2317	0.5514	0.5343	-0.1674	0.1831	0.2934

Beachte die Unterschiede zwischen den beiden Analysen.

Sonstige Angaben: Programm Aitchnew.exe
Daten Fenantrn.dat

Beispiel 2: Analyse von Morph-Frequenzen

Hier sind zu beobachten:

1. Der große Einfluß der Einschränkung auf die Korrelationen. Dieses Verhältnis ist ganz unerwartet .
2. Der Einfluß auf die Vektorelemente der Multivariatanalyse, d..h. die Eigenvektoren.

Die Daten bestehen aus den Proportionen von drei Morphen bei der Ostrakodengattung *Echinocythereis* (=Skulpturtypen) und die Frequenzen von Va Cr B

Beobachtungsvektoren = 10 Variablen = 6

Die Datenmatrix					
1	2	3	Va	Cr	N
0.73	0.21	0.06	0.25	0.50	0.25
0.33	0.66	0.01	0.30	0.40	0.30

0.16	0.79	0.05	0.35	0.45	0.20
0.26	0.58	0.16	0.35	0.40	0.25
0.13	0.74	0.13	0.50	0.40	0.25
0.39	0.57	0.05	0.40	0.40	0.20
0.21	0.71	0.07	0.32	0.48	0.20
0.08	0.88	0.04	0.31	0.39	0.30
0.20	0.72	0.05	0.34	0.40	0.28
0.15	0.72	0.13	0.48	0.42	0.25

Liste der logarithmischen Proportionen

1	-0.3147	-1.5606	-2.8134	-1.3863	-0.6931	-1.3863
2	-1.1087	-0.4155	-4.6052	-1.2040	-0.9163	-1.2040
3	-1.8326	-0.2357	-2.9957	-1.0498	-0.7985	-1.6094
4	-1.3471	-0.5447	-1.8326	-1.0498	-0.9163	-1.3863
5	-2.0402	-0.3011	-2.0402	-0.6931	-0.9163	-1.3863
6	-0.9416	-0.5621	-2.9957	-0.9163	-0.9163	-1.6094
7	-1.5606	-0.3425	-2.6593	-1.1394	-0.7340	-1.6094
8	-2.5257	-0.1278	-3.2189	-1.1712	-0.9416	-1.2040
9	-1.6094	-0.3285	-2.9957	-1.0788	-0.9163	-1.2730
10	-1.8971	-0.3285	-2.0402	-0.7340	-0.8675	-1.3863

Die zentrierte Kovarianzmatrix

	1	2	3	4	5	6
1	0.4010	-0.1950	-0.1810	-0.0663	0.0387	0.0026
2	-0.1950	0.1923	-0.0834	0.0505	0.0016	0.0341
3	-0.1810	-0.0834	0.4544	-0.0049	-0.0726	-0.1125
4	-0.0663	0.0505	-0.0049	0.0285	-0.0088	0.0011
5	0.0387	0.0016	-0.0726	-0.0088	0.0237	0.0174
6	0.0026	0.0341	-0.1125	0.0011	0.0174	0.0574

Korrelationsmatrix in der Simplexmangfaltigkeit

	1	2	3	4	5	6
1	1.0000	-0.7021	-0.4240	-0.6203	0.3964	0.0169
2	-0.7021	1.0000	-0.2823	0.6818	0.0231	0.3246
3	-0.4240	-0.2823	1.0000	-0.0431	-0.6984	-0.6968
4	-0.6203	0.6818	-0.0431	1.0000	-0.3392	0.0265
5	0.3964	0.0231	-0.6984	-0.3392	1.0000	0.4709
6	0.0169	0.3246	-0.6968	0.0265	0.4709	1.0000

Die übliche Korrelationsmatrix

	1	2	3	4	5	6
1	1.0000	-0.9670	-0.2142	-0.5147	0.5546	-0.0951
2	-0.9670	1.0000	-0.0355	0.3837	-0.5490	0.1308
3	-0.2142	-0.0355	1.0000	0.6058	-0.1008	-0.1797
4	-0.5147	0.3837	0.6058	1.0000	-0.4270	-0.2047
5	0.5546	-0.5490	-0.1008	-0.4270	1.0000	-0.4830
6	-0.0951	0.1308	-0.1797	-0.2047	-0.4830	1.0000

Hauptkomponentenanalyse: eingeschränkte Modelle Aitchisons

Eigenwerte					
2.57164	2.28954	0.52440	0.44295	0.17147	Spur = 6
Prozent der Spur					
42.8607	38.1589	8.7400	7.3824	2.8579	
Eigenvektoren					

	1	2	3	4	5
1	-0.5238	0.2480	-0.5209	0.1118	-0.1838
2	0.2770	-0.5540	0.0725	-0.2786	-0.6061
3	0.4140	0.4482	0.4081	-0.0238	0.2607
4	0.4151	-0.3702	-0.6038	0.0285	0.5483
5	-0.4823	-0.2539	0.2290	-0.6611	0.4392
6	-0.2695	-0.4787	0.3739	0.6866	0.1932

Rechenmaßwerte der Komponenten

Individuum					
1	-1.5300846	-0.4271523	-0.9933909	-0.2520616	
2	-1.4413645	-2.3676627	-1.3300535	-0.3689002	
3	-0.2033293	-1.6691204	-0.4944095	-0.8689899	
4	-0.0241736	-0.7017457	-0.2286035	-0.4923371	
5	0.4599569	-1.2818577	-0.1520645	-0.6294839	

6	-0.6317711	-1.1939517	-1.0857801	-0.5835272
7	-0.2908966	-1.2986230	-0.4343110	-0.8512119
8	-0.0122498	-2.2205089	-0.0290689	-0.6168938
9	-0.3827583	-1.6315744	-0.4988265	-0.5028226
10	0.3404323	-1.2086610	-0.1919186	-0.6364447

Hauptkomponentenanalyse der üblichen Korrelationsmatrix

	Eigenwerte				
2.85152	1.59855	0.95251	0.40117	0.19501	
Eigenvektoren					
	1	2	3	4	5
1	0.5493	-0.0416	0.3092	-0.3350	0.0350
2	-0.5124	0.1851	-0.4384	0.1773	0.0328
3	-0.2036	-0.5848	0.4675	0.5198	-0.3069
4	-0.4173	-0.4506	0.1585	-0.4315	0.6415
5	0.4543	-0.2649	-0.3638	0.5304	0.5563
6	-0.1170	0.5906	0.5798	0.3444	0.4272

Beachte die sehr großen Unterschiede zwischen den beiden Analysen.

Schluss des Beispielles:

Quellen: Aitchnew.exe

Echprop.dat

Fraktalgeometrie

Ein neues Arbeitsfeld er bietet die Fraktalgeometrie Benoit Mandelbrots. Diese Theorie hat sich unter anderem in bezug auf die Datologie und Bautechnik der Bildschirme mit hohem Leistungsvermögen, die Astronomie, die Biologie, die Ozeanographie, usw als besonders brauchbar herausgestellt. In morphometrischen Zusammenhängen intressiert man sich besonders für die Fraktaldimension.

Die Fraktaldimension gibt ein quantitatives Maß β der Unregelmässigkeit oder Ziellosigkeit einer Kurve- so zu sagen das ziellose Gekritzelt der schartig ausgebildeten Kurve. Man spricht von zufälligen Fraktalen.

Hier ist r die Schrittlänge, sehr wichtig für die Berechnungen und N stellt die Anzahl der selbstähnlichen Teile der Kurve dar. Der Begriff $|||||$ besitzt eine Schlüsselrolle in der Fraktalgeometrie, wobei gemeint wird: das Vorkommen von N gleichwertigen Nachbildungen der Gesamtkurve. Die Kurve besteht also aus dieser Anzahl gleicher Bausteine. Im Falle selbstähnlicher Figuren kann jeder Teil aus dem Ganzen durch eine Ähnlichkeitsabbildung erzeugt werden.

Betrachte eine Gerade. Diese Linie kann in N gleichen Teilen geteilt werden, wobei

$$r = 1/N$$

oder

$$Nr = 1.$$

Folglich ist der skalierende Faktor nicht mehr als ein kleines Nachbild des Ganzen. Jeder Teil ist somit eine exakte Abbildung der anderen $N-1$ Teile. Die Fraktaldimension einer Kurve kann 1

übersteigen.

Dies gilt ebenfalls für zwei Dimensionen, d.h. für eine Ebene: Die Euklidische Dimension einer Ebene ist 2.

$$Nr^2 = 1.$$

Jede rechteckige Form ist nicht mehr als ein kleines Nachbild des Ganzen, ebenso wie auch bei einer Geraden. Jeder Teil kann aus dem Ganzen durch eine Ähnlichkeitsabbildung erhalten werden.

Für eine beliebige Anzahl von Dimensionen haben wir

$$Nr^D = 1.$$

Für alle D-dimensionalen Parallelepipede gilt diese Gleichung.

Die **fraktale Dimension** ergibt sich durch den folgenden Quotienten:

$$D = \frac{\log N}{\log \frac{1}{r}}$$

(97)

In praktischen Zusammenhängen ist, wie schon oben bemerkt, die Wahl der "Schrittlänge" oder Skala von großer Bedeutung; es gibt keine einwandfreie Methode, um die korrekte Schrittlänge wählen zu können. Eine Reihe von Versuchen wird im allgemeinen notwendig.

Beispiel

Die Fraktaldimension des Umrißes eines Ahornsblatts:

Ein schwieriges Problem der quantitativen Taxonomie ist die Quantifizierung eines unregelmässigen Gegenstandes. Wie beschreibt man zum Beispiel statistisch den Umriß eines Blattes. Die Fraktalgeometrie er bietet eine richtige Möglichkeit, eine annehmbare Lösung zu finden. Bisher ist das Problem keineswegs befriedigend erforscht worden, aber die ersten Schritte in dieser Richtung sind schon gemacht worden.

Ich habe oben erwähnt, daß die Wahl der "Schrittlänge" von allergrößter Bedeutung ist. Ich werde jetzt die Ergebnisse einer Reihe von Versuchen angehend die Fraktaldimension des Ahornblattes mitteilen. Beachte: Die Koordinatenpaare sollen sehr dicht nebeneinander gestellt werden.

Mit Umrißlänge standardisiert und bei wiederholten Versuchen mit ziellos gewähltem Ausgangspunkt und mit linearer Schrittlänge erhielt ich die folgenden Schätzungen der Fraktaldimension.

1,026	1,025	1,030
1,031	1,030	1,017

Sämtliche sind eine Ahnung größer als 1 (die Euklidische Dimension einer Kurve).

Eine neue Versuchsreihe mit exponentieller Schrittlänge gab die folgenden Ergebnisse

1,193	1,163	1,166
1,173	1,251	1,191

Sämtliche Werte sind also etwas größer als die mit linearer Schrittlänge berechneten.

Der Schlußsatz hier scheint ziemlich klar, nämlich daß die Wahl der Schrittlänge außerordentlich wichtig erscheint.

Schrifttum

ABE, K. 1983. Population structure of Keijella bisanensis (Okubo) (Ostracoda, Crustacea). Journal of the Faculty of Sciences, University of Tokyo, 20: 443-448.

ABE, K., REYMENT, R.A., BOOKSTEIN, F.L., HONIGSTEIN, A., ALMOGI-LABIN, A., ROSENFELD, A., HERMELIN, O. 1988. Microevolution in two species of ostracods from the Santonian (Cretaceous) of Israel. Historical Biology, 1: 303-322.

AITCHISON, J. 1986. The Statistical Analysis of Compositional Data. Chapman and Hall, xv + 416 pp.

ANDERSON, T. W. 1963. Asymptotic theory for principal components analysis. Annals of Mathematical Statistics, 34: 122-148.

ANDERSON, T. W. 1984. An Introduction to Multivariate Statistical Analysis, Second Edition, Wiley and Sons, New York: xvi + 675 pp.

ANDERSON, T. W. and BAHADUR, R. R. 1962. Two sample comparisons of dispersion matrices for alternatives of immediate specificity. Annals of Mathematical Statistics, 33: 420-431-

BAILEY, D. W. 1956. A comparison of genetic and environmental components of morphogenesis in mice. Growth, 20: 63-74.

BARNETT, V. and LEWIS, T. 1978. Outliers in Statistical Data, Wiley and Sons, New York.

BARNETT, V. 1976. The ordering of multivariate data. Journal of the Royal Statistical Society, 139: 318-344.

BECKER, W. A. 1987. Manual of Procedures in Quantitative Genetics. Academic Enterprises, Pullman, WA, 198 pp.

BELL, M. A. and LEGENDRE, P. 1987. Multicharacter chronological clustering in a sequence of fossil sticklebacks. Systematic Zoology, 36: 52-61.

BELL, M.A., BAUMGARTNER, J.V., OLSON, E.C. 1985. Patterns of temporal change in single morphological characters of a Miocene stickleback fish. Paleobiology, 11: 258-271.

BELLMAN, R. 1960. Introduction to Matrix Analysis. McGraw-Hill, New York. xx + 328 pp.

- BENSON, R. H. 1972. The Bradleya problem, with descriptions of two new psychospheric ostracode genera, Agrenocythere and Poseidonamicus (Ostracoda, Crustacea). Smithsonian Contributions in Paleobiology, 12: 138 pp.
- BENZECRI, J. P. 1973. L'Analyse des Données:2, l'Analyse des Correspondances. Dunod, Paris. vi + 619 pp.
- BIRKS, H. J. B. and GORDON, A. D. 1985. Numerical Methods in Quaternary Pollen Analysis. Academic Press, xiii + 317 pp.
- BLUM, H. 1973. Biological shape and visual science. Journal of Theoretical Biology, 38: 205-287.
- BODERGAT, A. M. 1983. Les Ostracodes, temoins de leur environnement: approche chimique et écologique en milieu lagunaire et océanique. Documents du Laboratoire de Géologie Lyon, 88, 1-246.
- BOOKSTEIN, L. F. 1978. The Measurement of Biological Shape and Shape Change. Lecture Notes in Biomathematics, Springer Verlag, Berlin-Heidelberg, 24:vii+191 pp.
- BOOKSTEIN, F. L. 1983. The geometry of craniofacial growth invariants. American Journal of Orthodontics, 83: 221-234.
- BOOKSTEIN, F. L. 1984. A statistical method for biological shape comparisons. Journal of Theoretical Biology, 107: 475-520.
- BOOKSTEIN, F. L. 1986. Size and shape spaces for landmark data in two dimensions. Statistical Science, 1: 181-242.
- BOOKSTEIN, F. L. 1987. Random walk and the existence of evolutionary rates. Paleobiology, 13: 446-464.
- BOOKSTEIN, F.L. 1988. Random walk and the biometrics of morphological characters. Evolutionary Biology, 23: 369-398.
- BOOKSTEIN, L. F. 1989a. Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Machine. Intell., 11: 567-585.
- BOOKSTEIN, F.L. 1989b. Comment to "A survey of the statistical theory of shape by D. G. Kendall". Statistical Science, 4: 99-105.
- BOOKSTEIN, L. F. 1989c. Four metrics for image variation. Proc. Intl Confer. on Information Processing in Medical Imaging, New York, Alan R. Liss. Inc.
- BOOKSTEIN, F. L. 1991. Morphometric Tools for Landmark Data., Cambridge University Press, New York.
- BOOKSTEIN, F. L. and SAMPSON, P. D. 1987. Statistical models for geometric components of shape change. Proceedings of the Section on Statistical Graphics. 1987 Meeting of the American Statistical Association, 18-27.
- BOOKSTEIN, F.L., CHERNOFF, B., ELDER, R., HUMPHRIES, J., SMITH G., STRAUSS, R. 1985. Morphometrics in Evolutionary Biology. Special Publication No. 15, Academy of Natural Sciences, Philadelphia, xvii + 277 pp.
- BULMER, M. G. 1980. The Mathematical Theory of Quantitative Genetics. Oxford University Press, x + 256 pp.
- BURNABY, T. P. 1966. Growth invariant discriminant functions and generalized distances. Biometrics, 22: 96-110.
- CAMPBELL, N. A. 1978. The influence function as an aid in outlier detection in discriminant analysis. Applied Statistics, 27: 251-258.
- CAMPBELL, N. A. 1979. Canonical Variate Analysis: Some Practical Aspects. Ph.D. Thesis, Imperial College, University of London. 179 S.
- CAMPBELL, N. A. 1980a. Shrunken estimators in discriminant and canonical variate

- analysis. Applied Statistics, 29: 5-14.
- CAMPBELL, N. A. 1980b. Robust procedures in multivariate analysis: I, Robust covariance estimation. Applied Statistics, 29: 231-237.
- CAMPBELL, N. A. 1982. Robust procedures in multivariate analysis. II. Robust canonical variate analysis. Applied Statistics, 31: 1-8.
- CAMPBELL, N. A. 1984. Canonical variate analysis with unequal covariance matrices: generalizations of the usual solution. Mathematical Geology, 16: 109-124.
- CAMPBELL, N. A. and ATCHLEY, W. R. 1981. The geometry of canonical variate analysis. Systematic Zoology, 30: 268-280.
- CAMPBELL, N. A. and MAHON, R. J. 1974. A multivariate study of variation into species of rock crab of the genus Leptograpsus. Australian Journal of Zoology, 22: 417-425.
- CAMPBELL, N. A. and REYMENT, R. A. 1978. Discriminant analysis of a Cretaceous foraminifer using shrunken estimators. Mathematical Geology, 10: 347-359.
- CAMPBELL, R. A. and REYMENT, R. A. 1980. Robust multivariate procedures applied to the interpretation of atypical individuals of a Cretaceous foraminifer. Cretaceous Research, 1: 207-221.
- CHARLESWORTH, B. 1984. Some quantitative methods for studying evolutionary patterns in single characters. Paleobiology, 10: 308-313.
- CHARLESWORTH, B., LANDE, R., SLATKIN, M. 1982. A Neo-Darwinian commentary on macroevolution. Evolution, 36: 474-498.
- CHEVERUD, J. M. 1984. Quantitative genetics and developmental constraints on evolution by selection. Journal of Theoretical Biology 110: 155-171.
- CHEVERUD, J. M. 1989. A comparative analysis of morphological variation patterns in the papionins. Evolution, 43: 1737-1747.
- CHEVERUD, J. M. 1988. A comparison of genetic and phenotypic correlations. Evolution, 42: 958-968.
- COX D.R. and LEWIS, P.A.W. 1966. The Statistical Analysis of Series of Events. Methuen, London, 285 pp.
- CRESPI, B. and BOOKSTEIN, F.L. 1989. A path-analytic model for the measurement of selection on morphology. Evolution, 43: 18-28.
- DARROCH, J. N. and MOSIMANN, J. E. 1983. Canonical and principal components of shape. Biometrika, 72: 241-252.
- DEGENS, E. T. 1968. Geochemie der Sedimente. Ferdinand Enke Verlag, Stuttgart, viii + 282 pp.
- DIGBY, P.G.N. and KEMPTON, R. A. 1988. Multivariate Analysis of Ecological Communities. Chapman and Hall, London. viii + 206 pp.
- EASTMENT, H. T. and KRZANOWSKI, W. J. 1982. Cross-validators choice of the number of components from a principal component analysis. Technometrics, 24: 73-77.
- EBEL, K. 1983. Berechnungen zur Schwebefähigkeit von Ammoniten. Neues Jahrb. Geol. Paläont. Mh (1983:10): 614-640.
- ECKART, C. and YOUNG, G. 1936. The approximation of one matrix by another of lower rank. Psychometrika, 1:211-218.
- EVERITT, B. 1978. Graphical Techniques for Multivariate Data, Heinemann, London, 117 pp.
- FALCONER, D. S. 1981. Introduction to Quantitative Genetics. Second Edition,

Longman Inc., New York, ix + 340 pp.

FISHER, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics, London, 7: 179-188.

FISHER, R. A. 1940. The precision of discriminant functions. Annals of Eugenics London, 10: 422-429.

FLURY, B. 1988. Common Principal Component Analysis. Wiley & Son, New York, xiii + 258 pp.

FLURY, B. and RIEDWYL, H. 1988. Multivariate Statistics. Chapman and Hall, xi + 296 pp.

GABRIEL, K. R. 1968. The biplot graphical display of matrices with application to principal components analysis. Biometrika, 58: 453-467.

GALTON, F. 1887. Natural Inheritance. Macmillan, London.

GANTMACHER, F. R. 1965. Matrizenrechnung. Vol. 1, xi + 324 pp; vol. 2, viii + 244 pp., VEB deutscher Verlag der Wissenschaften, Berlin.

GENSTAT 5 COMMITTEE. 1988. Statistics Department, Rothamsted Experimental Station; GENSTAT 5 Reference Manual. Oxford Science Publications, Oxford, 749 pp.

GNANADESIKAN, R. 1977. Methods for Statistical Data Analysis of Multivariate Observations. Wiley and Son, New York. x + 311 pp.

GOLDSTEIN, M. and DILLON, W. R. 1978. Discrete Discriminant Analysis. Wiley and Sons, New York, 186 pp.

GOLUB, G.H., HEATH, N., and WAHBA, G. 1979. Generalized cross-validation as a method of choosing a good ridge parameter. Technometrics, 21: 215-222.

GORDON, A. D. 1973. A sequence-comparison statistic and algorithm. Biometrika, 60: 197-200.

GORDON, A. D. 1981. Classification. Monographs on Applied Probability and Statistics. Chapman and Hall, London. xii + 193 pp.

GORDON, A. D. 1982. On measuring and modelling the relationship between two stratigraphically-recorded variables. In Quantitative Stratigraphic Correlation: pp. 241-248. Eds. J. M. Cubitt and R. A. Reyment, Wiley and Sons, Chiswick.

GORDON A. D. and REYMENT, R. A. 1979. Slotting of borehole sequences. Mathematical Geology, 11: 309-327.

GOWER, J. C. 1966a. Some distance properties of latent roots and vectors used in multivariate analysis. Biometrika, 53: 325-338.

GOWER, J. C. 1966b. A Q-technique for the calculation of canonical variates. Biometrika, 53: 588-589.

GOWER, J. C. 1968. Adding a point to vector diagrams in multivariate analysis. Biometrika, 55: 582-585.

GOWER, J. C. 1971a. A general coefficient of similarity and some of its properties. Biometrics, 27: 857-874.

GOWER, J. C. 1976. Growth-free canonical variates and generalized inverses. Bulletin of the Geological Institutions of the University of Uppsala, 7: 1-10.

GOWER, J. C. 1977. The analysis of asymmetry and orthogonality. In Recent Developments in Statistics, J. Barra, Ed., North Holland Publishing Company, Amsterdam.

GREENACRE, M. J. 1984. Theory and applications of Correspondence Analysis. Academic Press, London, xi + 364 pp.

HALDANE, J. G. S. 1949. Suggestions as to quantitative measurement of rates of

evolution. Evolution, 3: 51-56.

HAMPEL, F.R., RONCHETTI, E. M., ROUSSEEUW, P.J., STATEL, W. A. 1986. Robust Statistics., Wiley and Sons, New York, xxi + 502 pp.

HAND, D. J. 1981. Discrimination and Classification. Wiley and Sons, x + 218 pp.

HARTL, F. L. 1980. Principles of Population Genetics, Sinauer Associates Inc., Mass. xvi + 488 pp.

HEGMAN, N.J.P. and DEFRIES, J.C. 1970. Are genetic and environmental correlations correlated? Nature, 226: 284-285.

HILL, M.O. 1974. Correspondence analysis: a neglected multivariate method. Journal of the Royal Statistical Society, ser. c, 23: 340-354.

HIRSCHFELD, H.O. 1935. A connection between correlation and contingency. Proceedings of the Cambridge Philosophical Societ, (Math. Proc.), 31: 520-524.

HOAGLIN, D.C., MOSTELLER, F., and TUKEY, J. W. (Eds.). 1987. Exploring Data Tables, Trends and Shapes. Wiley and Son, xxii + 527 pp.

HORST, P. 1963. Matrix Algebra. Holt, Rhinehart and Winston Inc., New York, xxi + 517 pp.

HOTELLING, H. 1931. The generalization of Student's ratio. Annals of Mathematical Statistics, 2: 360-378.

HOTELLING, H. 1936. Relations between two sets of variates. Biometrika, 28: 321-377.

HOUSEHOLDER, A. S. 1964. The Theory of Matrices in Numerical Analysis. Blaisdell Publ. Co., New York, xi + 257 pp.

JOLICOEUR, P. 1963. The degree of robustness in Martes americana. Growth, 27: 1-27.

JOLICOEUR, P. and MOSIMANN, J. E. 1960. Size and shape variation in the painted turtle. A principal component analysis. Growth, 24: 339-354.

JOLLIFFE, I. T. 1986. Principal Component Analysis. Springer Series in Statistics, Springer-Verlag, New York, xiii + 271.

JÖRESKOG, K.G., KLOVAN, J. E., REYMENT, R. A. 1976. Geological Factor Analysis, Elsevier, Amsterdam, 178 pp.

KIRKPATRICK, M. 1982. Quantum evolution and punctuated equilibria in continuous genetic characters. American Naturalist, 199: 833-848.

KOHN, L.A.P. and ATCHLEY, W.R. 1988. How similar are genetic correlation structures? Data from mice and rates. Evolution, 42: 467-481.

KRISHNAIAH, P. R. (Ed.). 1977. Multivariate Analysis - IV. North Holland Publ. Co., Amsterdam, xiii + 549 pp.

KRZANOWSKI, W. J. 1979. Between-groups comparison of principal components. Journal of American Statistical Association, 74: 703-707.

KRZANOWSKI, W. J. 1983. Cross-validatory choice in principal component analysis; some sampling results. Journal of statistical Computer Science, 18: 294-314.

KRZANOWSKI, W. J. 1984. Principal component analysis in the presence of group structure. Applied Statistics, 33: 164-168.

KRZANOWSKI, W. J. 1987a. Cross-validation in principal component analysis. Biometrics, 43: 575-584.

KRZANOWSKI, W. J. 1987b. Selection of variables to preserve multivariate data structure using principal components. Journal of the Royal Statistical Society, Ser. C, 36: 22-33.

- LANDE, R. 1976. Natural selection and random genetic drift in phenotype evolution. Evolution, 30: 314-334.
- LANDE, R. 1977. Statistical tests for natural selection on quantitative characters. Evolution, 31: 442-444.
- LANDE, R. 1978. Evolutionary mechanisms of limb loss in tetrapods. Evolution, 32: 73-92.
- LANDE, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution, 33: 402-416.
- LANDE, R. 1980a. The genetic covariances between characters maintained by pleiotropic mutations. Genetics, 94: 203-215.
- LANDE, R. and ARNOLD, S. J. 1983. The measurement of selection on correlated characters. Evolution, 37: 1210-1226.
- LOHMANN, G.P. 1983. Eigenshape analysis of microfossils. A general morphometric procedure for describing changes in shape. Mathematical Geology, 15: 659-672.
- LOHMANN, G.P. and SCHWEITZER, P. N. 1989. On eigenshape analysis. Morphometrics in Systematic Biology Workshop, 16-28 May, 1988, Ann Arbor. Eds. F.J. Rohlf and F.L. Bookstein.
- LOVE, W. A. and STEWART, D. K. 1968. Interpreting Canonical Correlations: Theory and Practice. Palo Alto, American Institutes for Research, 66 pp.
- MANDELBROT, B. B. 1987. *Die fraktale Geometrie der Natur*. Birkhäuser Verlag, Basel, 491 S.
- MANLY, B.F.J. 1985. The Statistics of Natural Selection, Chapman and Hall, London, xvi + 484 pp.
- MARDIA, K.V. 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika, 57:519-530.
- MARDIA, K. V. 1977. Mahanalobis distances and angles. In P. R. Krishnaiah (Ed.), Multivariate Analysis - IV, North Holland Publ. Co., Amsterdam, pp. 495-511.
- MARDIA, K. V. 1989. Shape analysis of triangles through directional techniques. J. roy Statist. Soc., B 51: 449-458.
- MARDIA, K.V., KENT, J.T., BIBBY, J.M. 1979. Multivariate Analysis, Academic Press, xv + 521 pp.
- MARONNA, R.A. 1976. Robust M-estimators of multivariate location and scatter. Annals of Statistics, 4: 51-67.
- MARQUARDT, D.W. and SNEE, R.D. 1975. Ridge regression in practice. American Statistician, 29: 3-19.
- MAXWELL, E. A. 1958. Coordinate Geometry with Vectors and Tensors. Oxford University Press, ix + 194 pp.
- MOSIMANN, J.E. 1970. Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. Journal of the American Statistical Association, 65: 930-945.
- MOSIMANN, J. E. and JAMES, F.C. 1979. New statistical methods for allometry with applications to Florida red-winged blackbirds. Evolution, 33: 444-459.
- MOURANT, A. 1983. Blood Relations. Oxford University Press, vi + 146 pp.
- MURDOCH, D.C. 1957. Linear Algebra for Undergraduates. Wiley and Sons, New York, vii + 239 pp.
- ORLOCI, L. 1975. Multivariate Analysis in Vegetation Research, Junk B,V, The Hague, ix + 276 pp.

- PEARSON, K. 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2 (ser. 6): 559-572.
- RAO, C. R. 1964. The use and interpretation of principal components analysis in applied research. Sankhya, 12: 229-246.
- RAUP, D.M. and GOULD, S.J. 1974. Stochastic simulation and evolution of morphology towards a nomothetic paleontology. Systematic Zoology, 23: 305-322.
- REYMENT, R.A. 1969a. A multivariate paleontological growth problem. Biometrics, 22: 1-8.
- REYMENT, R.A. 1969b. Covariance structure and morphometric analysis- a contribution to paleogenetics. Mathematical Geology, 1: 185-197.
- REYMENT, R.A. 1970. Spectral breakdown of morphometric chronoclines. Mathematical Geology, 2: 365-376.
- REYMENT, R.A. 1971a. Introduction to Quantitative Paleoecology. Elsevier, Amsterdam, xii + 226 pp.
- REYMENT, R.A. 1971b. Multivariate normality in morphometric analysis. Mathematical Geology, 3: 357-368.
- REYMENT, R.A. 1972. Models for studying the occurrence of lead and zinc in a deltaic environment. In Mathematical Models of Sedimentary Processes Ed. T. W. Merriam, Plenum, New York, 233-245.
- REYMENT, R.A. 1976. Chemical components of the environment and Late Campanian microfossil frequencies. Geologiska Föreningens Förhandlingar, 98: 322-328.
- REYMENT, R.A. 1982b. Phenotypic evolution in a Cretaceous foraminifer. Evolution, 36: 1182-1199.
- REYMENT, R.A. 1982c. Speciation in a Late Cretaceous lineage of Veenia (Ostracoda). Journal of Micropalaeontology, 1: 37-44.
- REYMENT, R.A. 1982c. Analysis of trans-specific evolution in Cretaceous ostracods. Paleobiology, 8: 292-305.
- REYMENT, R.A. 1982d. Morphological variation in time of a Paleocene species of Cytherella. In Fossil and Recent Ostracoda, Ed. R. H. Bate, E. Robinson, L. M. Sheppard, London, 165-168.
- REYMENT, R.A. 1983. Phenotypic evolution in microfossils. Evolutionary Biology, 16: 209-254.
- REYMENT, R.A. 1985a. Multivariate morphometrics and analysis of shape. Mathematical Geology, 17: 591-609.
- REYMENT, R.A. 1985b. Phenotypic evolution in a lineage of the Eocene ostracod Echinocythereis. Paleobiology, 11: 74-194.
- REYMENT, R.A. 1987. Multivariate analysis in geoscience: fads, fallacies and future. Chemometrics and Intelligent Laboratory Systems, 2: 79-91.
- REYMENT, R.A. and BANFIELD, C.F. 1976. Growth-free canonical variates applied to fossil foraminifers. Bulletin of the Geological Institutions of the University of Uppsala, NS 7: 11-21.
- REYMENT, R.A. and BANFIELD, C.F. 1981. Analysis of asymmetric relationships in geological data In Modern Advances in Geomathematics, Eds. R. Craig and M. Labovitz, Pion Ltd, London, 236-253.
- REYMENT, R.A. and KENNEDY, J. M. 1990. Paper dealing with Knemiceras. Evolutionary Biology
- REYMENT, R.A., BOOKSTEIN, F.L., MCKENZIE, K.G., MAJORAN, S. 1988. Ecophenotypic variation in Mutilus pumilus (Ostracoda) from Australia studied by canonical

- variate analysis and tensor biometrics. Journal of Micropalaeontology, 7: 11-20.
- REYMENT, R.A. and STURESSON, U. P. A. 1987. Correlation of chemical and physical environmental fluctuations in a Late Cretaceous borehole sequence - a multivariate study. Sedimentary Geology, 53: 311-325.
- REYMENT, R.A. and VAN VALEN, L. 1969. Buntonia olokundudui sp nov. (Ostracoda, Crustacea): a study of meristic variation in Paleocene and Recent ostracods. Bulletin of the Geological Institutions of the University of Uppsala, NS 1: 83-94.
- REYMENT, R.A., BLACKITH, R.E., CAMPBELL, N.A. 1984. Multivariate Morphometrics, Second Edition, Academic Press, vii + 233 pp.
- ROHLF, F.J. 1986. Relationships among eigenshape analysis, Fourier Analysis and analysis of coordinates. Mathematical Geology, 18: 845-854.
- ROHLF, F.J. and ARCHIE, J.W. 1984. The comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). Systematic Zoology, 33: 302-317.
- ROHLF, F.J. and BOOKSTEIN, F.L. 1987. A comment on shearing as a method for size correction. Systematic Zoology, 36: 356-367.
- ROHLF, F. J. and BOOKSTEIN, F. L. 1990. Michigan Morphometrics Workshop. University of Michigan Press, Ann Arbor.
- ROHLF, F.J. AND FERSON, S. 1983. Image analysis. In J. Felsenstein (Ed.). Numerical Taxonomy, NATO ISI Series 6, Ecological Science No. 1, Springer Verlag, New York, pp. 583-599.
- RÖHRS, M. 1959. Neue Ergebnisse und Probleme der Allometrieforschung. Zeitschrift für wissenschaftliche Zoologie, 162: 1-95.
- ROUGHGARDEN, J. 1979. Theory of Population Genetics and evolutionary Ecology: an Introduction. MacMillan Publishing Co., New York, 634 pp.
- ROY, S. N., GNANADESIKAN, R. and SRIVASTAVA, J. N. 1971. Analysis and Design of Certain Multiresponse Experiments. Pergamon Press, Oxford.
- SCHEINER, S. M. and LYMAN, R. F. 1989. The genetics of phenotype plasticity. J. evol. Biol., 2: 95-107-
- SCHMALHAUSEN, I.I. 1949. Factors of Evolution: the Theory of Stabilizing Selection, Blakiston, Philadelphia.
- SEARLE, S.R. 1961. Phenotypic, genetic and environmental correlations. Biometrics, 17: 474-480.
- SEARLE, S.R. 1966. Matrix Algebra for the Biological Sciences, Wiley and Sons, New York, xii + 296 pp.
- SEBER, G. A. F. 1984. Multivariate Observations, Wiley and Sons, New York, xx + 686 pp.
- SIEGEL, A.F. and BENSON, R.H. 1982. A robust comparison of biological shapes. Biometrics, 38: 341-350.
- SPRENT, P. 1972. The mathematics of size and shape. Biometrics, 28: 23-37.
- TELNAES, W. BJÖRSETH, A., CHRISTY, A.A., KVALHEIM, O. M. 1987. Interpretation of multivariate data: relationship between phenethrenes in crude oils. Chemometrics and Intelligent Laboratory Systems, 2: 149-153.
- TER BRAAK, C.F.J. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology, 67: 1167-1179.
- TER BRAAK, C. F. J. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio, 69: 69-77.

- THOMPSON, D.W. 1917. On Growth and Form, Cambridge University of Press.
- TORGERSON, W.S. 1958. Theory and Methods of Scaling. Wiley and Sons, New York, 460 pp.
- TROWELL, H. 1975. Diabetes mellitus and obesity. Chapter 16 in D.P. Burkett & H. C. Trowell (Eds.) Refined Carbohydrate Foods and Disease. Academic Press, London.
- TURELLI, M. 1988. Phenotypic evolution, constant covariances, and the maintenance of additive variance. Evolution, 42: 1342-1347.
- TURELLI, M., GILLESPIE, J. H. and LANDE, R. 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. Evolution, 42: 1085-1089.
- VIA, S. and LANDE, R. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution, 39: 505-522.
- WATSON, G.S. 1970. Orientation statistics in the Earth Sciences. Bulletin of the Geological Institutions of the University of Uppsala, NS 2: 73-89.
- WATSON, G. S. 1981. The interaction of statistics and geology - finite deformations. In D. F. Merriam Ed. Down to Earth Statistics: Solutions looking for Geological Problems, Syracuse University, Geology Contribution 8, 17-27.
- WATSON, G. S. 1989. Comment on "D. G. Kendall "A survey of the statistical theory of shape" Statistical Science, 4: 113-115.
- WOLD, S. 1978. Cross-validatory estimation of the number of components in factor and principal component models. Technometrics, 20: 397-405.
- WRIGHT, S. 1931. Evolution in Mendelian populations. Genetics, 16: 97-159.
- WRIGHT, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the Sixth International Congress on Genetics, 1: 356-366.
- WRIGHT, S. 1968. Evolution and the Genetics of Populations. Vol. 1. Genetic and Biometric Foundations. University of Chicago Press, Chicago, vii + 469 pp.
- YOUNG, G. and HOUSEHOLDER, A.S. 1938. Discussion of a set of points in terms of mutual distances. Psychometrika, 3: 19-22.
- ZAHN, C.R. and ROSKIES, R. Z. 1972. Fourier description for plane closed curves. IEEE Transactions on Computers, c-21: 269-281.
- ZURMÜHL, R. Matrizen Springer Verlag, Berlin xii + 452S.